Article

Language Processing Units Are Not Equivalent to Sentences: Evidence from Writing Tasks in Typical and Dyslexic Children

Georgeta Cislaru [1,*], Quentin Feltgen [2], Elie Khoury [3], Richard Delorme [3,4] and Maria Pia Bucci [5]

1 MoDyCo, UMR 7114 CNRS Université Paris Nanterre, 9200 Nanterre, France
2 Department of Linguistics, Ghent University, 9000 Ghent, Belgium; quentin.feltgen@gmail.com
3 Child and Adolescent Psychiatry Department, Robert Debré Hospital, 75019 Paris, France; elie.khoury@aphp.fr (E.K.); richard.delorme@aphp.fr (R.D.)
4 InovAND, Paris University, 75000 Paris, France
5 ICAR, UMR 5191 CNRS Université Lyon 2, 69000 Lyon, France; mariapia.bucci@gmail.com
* Correspondence: gcislaru@parisnanterre.fr

Abstract: Despite recent research on the building blocks of language processing, the nature of the units involved in the production of written texts remains elusive: intonation units, which are evidenced by empirical results across a growing body of work, are not suitable for writing, where the sentence remains the common reference. Drawing on the analysis of the writing product and process, our study explores how children with and without dyslexia handle sentences. The children were asked to write a short story and the writing process was recorded using keystroke logging software (Inputlog 7 & 8). We measured the number of pauses, the nature of the language sequences segmented by pauses, and the revision operations performed throughout the process. We analyzed sentences both in product and process. Our results showed that both the written product and the writing process reflect the establishment of a syntactic schema during language processing in typical children, in line with the first functional step in processing. This was not clearly evidenced in the case of dyslexic children, due to their limited production: beyond spelling, syntactic elaboration was also affected. In contrast, it appeared that the units of language processing cannot be equated with sentences in writing: the information flow is produced through usually smaller bursts that each carry part of the meaning or correspond to a specific operation of text crafting and revision.

Keywords: writing; typical children; dyslexia; sentence processing; keystroke logging

1. Introduction

Written production involves several levels of conceptualization and segmentation that children must learn to master (Fayol 2013; Tolchinsky and Teberosky 1998): they must gain use of letters or characters, syllables, words, sentences, etc. However, while letters or words can be retrieved from stably available resources in alphabetic languages, the production of sentences requires an effort of conceptualization and the mobilization of in situ cognitive resources. Sentences appear, therefore, as the first "scale" at which meaning elaboration actively takes place with the aim of producing a text.

The present paper studies sentence crafting when writing in French-speaking children, both typical and dyslexic. Based on data from the writing process and its product, we challenge the view according to which sentences are the units of language processing. Our hypothesis is that written production requires the handling of infra-sentential sequences, which contrasts with the necessity to produce sentences when writing.

The article starts with a brief overview of the literature on the concept of the sentence and its place in the processing of written and spoken language compared to infra-sentential informational units (Section 2). An overview of children's sentence production in writing is drawn up in Section 3. Section 4 introduces materials and methods employed. The results of the study are presented in Section 5 and are discussed in Section 6.

2. In Search for the Unit of Language Processing

2.1. The Role of the Sentence in Language Processing

According to the literature, the sentence plays a major role in language processing across different modalities (oral vs. written, production vs. reception). Several models of language processing assume that the first step of sentence construction is functional and consists in building the overall functional structure of the sentence out of abstract lexical items (Garrett 1975, 1980), which Bock and Levelt (1994) roughly identify as grammatical encoding. Garrett's model is considered operational for writing if enriched with a monitoring–editing component (Fayol 1991). Sentence parsing, dealing incrementally or in parallel with strings of words in an utterance, is a general processing strategy that includes assembling and checking while the contents are still unfolding (in reception) or to be unfolded (in production) (see Mitchell 1994).

The literature also provides evidence of the relationship between syntax and reading or oral performance. Syntactic awareness has been shown to influence reading skills and to enhance reading fluency and comprehension (Bowey 1986; Cain 2007; Mokhtari and Thompson 2006). A sentence superiority effect is considered to determine the accuracy of word identification, in which syntactic representation plays a crucial role (Snell and Grainger 2017). Using EEG recording, Wen et al. (2019) demonstrated a sentence superiority effect due to the interactive processing of ongoing words in reading by skilled readers. Studies on writing have also focused on the role of syntax and the place of sentences. Foulin (1998) argued that the sentence seems to be the most plausible unit of conceptual and linguistic planning.

2.2. Defining the Sentence and Its Scope

Despite the role attributed to the sentence, defining it as a linguistic unit remains paradoxically difficult. The concept has been defined from the point of view of grammatical acceptability by native speakers (Chomsky 1965), by its prosody, grammatical independence, and predicative value (Zawadowski 1971), or in a functional perspective as an articulate sound symbol embodying some volitional attitude of the speaker towards the listener (Gardiner 1922). Sentence completeness and semantic units are also identified as definitional criteria. However, Lyons (1977) distinguished between grammatical and contextual completeness, in line with the distinction between the sentence as an abstract structure and the utterance as a sequence of oral performance. Unity of meaning and prosodic (oral) and typographic (written) marking are recurrent criteria in grammar manuals, and are sometimes also mentioned in the linguistic literature. Drawing on the previous approaches, a sentence can be defined as "a unit of communication consisting of a structured and ordered sequence of word(s), the enunciation of which produces an utterance, and which the enunciator decides to make a sentence" (Van Raemdonck et al. 2011, p. 103, cited in Van Raemdonck 2013, p. 209). The sentence is thus seen as the result of an intentional act on the part of the enunciator, by virtue of her/his grammatical skills. Combettes (2011, p. 14) recalls the evolution of grammatical notions in French, which moved over the course of the 18th century from the period, a unit of speech tailored for meaning, structured via connectors or parataxis and enunciated with a single breath, to a choppy style consisting of juxtaposed short sentences, thus reinforcing the feeling of a syntactically autonomous, semantically self-contained sentence. He points out that this "grammaticalization" of the sentence is subordinate to philosophical and logical objectives designed to establish the link between thought and language, and not to language learning or practice objectives. In this perspective, the sentence is a construct, a sort of grammatical artifact used to segment the information flow. As mentioned by Ruwet (1967, p. 366), "In generative grammar, the notion of sentence is held to be a primitive, undefined term in the theory of grammar". Thus, the paradox of the sentence is that it is "necessary, but sometimes impossible to find" (Le Goffic 2011, p. 15): sentences, defined as the fundamental language units of meaningful communication, are necessary reference templates. However, their realization by children or in oral speech, for instance, is highly different from that found in idealized sentences.

Thus, the development of speech analysis and the grammar of speech deconstructed the notion of grammatical sentence, based on empirical observations of the sentence's inability to constitute a unit of (oral) discourse analysis (Brazil (1995) and O'Grady (2010) for English; Blanche-Benveniste et al. (1990) and Berrendonner (2004) for French). Indeed, oral speech is made up of disfluencies, fragments, false starts, etc. Grammars of speech propose new categories such as intonational or illocutionary units (Benzitoun et al. 2010), which are merely "functional increments" (Brazil 1995). According to this literature, the sentence is therefore no longer seen as the most fundamental processing unit of language: it is indefinable and therefore non-operational.

A possible explanation of this divergence regarding the role attributed to the sentence is that syntactic performance is manifested differently in speech, whose processual units are arguably underdetermined, or distorted due to the time constraints associated with the online packing of information into sentence-like processual units (Le Goffic 2011), and writing, where sentences appear to be an explicit building block of text construction. According to Goody (1987, p. 264), *greater formality* and *greater syntactical elaboration* are among the seven features of written language, whereas *short clauses, unembedded dependent clauses, run-on clauses, repetition, truncation, hedges, dislocations*, etc., are identified as features of speech by Givón (2002, p. 75). Literacy seems not only to have played a key role in the invention of the sentence as a concept (Seguin 1993) but also to remain anchored in our understanding of sentence patterns. In written texts, the sentence is formalized and can be defined not only as a syntactic structure constituting a unit of meaning, but also as a sequence that is explicitly marked as such by a capital letter at the beginning and a period at the end (or any other strong punctuation mark). In accordance with these principles, the sentence appears as one of the production objectives, together with text crafting.

2.3. Information Processing and Processing in Writing

However, while the sentence can be seen as a tool for segmenting written text, this does not mean that it constitutes a unit for processing written production. Other units have been identified in the linguistic literature, such as multiword sequences, which are formulaic form–meaning pairings that are memorized and actualized in common patterns (Sinclair 1991, p. 108). These might be good candidate units for incremental processing. Frank et al. (2012) argued that hierarchical relationships are not necessary for sentence processing: in their view, the sequential approach is congruent with the formulaic conception of language, where multiword sequences are combined by making selections in parallel streams. While multiword sequences constitute about half of language production according to Sinclair's "idiom principle", it remains to be understood how and in what form the other half are processed, and how the two relate in order to produce language. Previous work has shown that only a small proportion of multiword units correspond to written production units (Cislaru and Olive 2017; Gilquin 2024).

The fact is that process and product are dissociated in writing, unlike in oral speech. This raises the issue of language processing units across language modalities on the one hand, and in the articulation between the process and product in writing on the other hand. Some authors argue that the basic units of processing are shared across language modalities. For instance, in their Linear Unit Grammar, Sinclair and Mauranen (2006) advocate for instance a single way of processing speaking, writing, reading and listening, through small units of meaning called chunks. Substantiating such a claim nevertheless entails empirically accessing these processing units. Chafe (1992), for instance, built on the understanding of what goes on in speech to focus on the special dynamics of information flow as it is created by writers and perceived by readers. He used aloud reading of a text written in English to identify possible intonational units that reflect the organization of *information flow* in written texts in comparison with speech. Such spontaneous segmentation in reading produces syntactically relevant sequences with an average length of 5.7 words, even though they go beyond the boundaries of simple/plain constituents. These intonational units are derived from a reception perspective of written texts, with no access to the writing process itself.

The development of research on the writing process gave access to spontaneous segmentation procedures specific to the written production process: first, through think-aloud protocols asking writers to comment on their strategies, and then via the use of keystroke logging or tablets to record the flow of writing in the alternation between language production and writing pauses. In this context, bursts of writing were identified. These are language sequences produced between two pauses (Chenoweth and Hayes 2001). Studies of writing consider pauses to be moments of reflection in preparation for the next segment of discourse, and they are dependent on memory, and syntactic, semantic, and discursive constraints (Matsuhashi 1981; Schilperoord 2002; Spelman Miller 2006). As a result, pauses are expected to be located mainly at the boundaries of basic units of language processing. The literature on the writing process has provided evidence for both the hypothesis of infra-sentence processing (e.g., Kaufer et al. 1986; Cislaru and Olive 2018) and the relevance of syntactic units in processing (Medimorec and Risko 2017; van Hell et al. 2008). Through a think-aloud protocol, Kaufer et al. (1986) noticed that writers do not produce sentences, but rather sequences segmented by pauses and meta-comments. In other words, writing typically occurs by constructing sentences from bursts of writing. In Kaufer et al.'s corpus, 36% of sequences ended at clause boundaries, 26% ended at phrase boundaries that were not also clause boundaries, and 39% ended at places that were neither clause nor phrase boundaries. On the whole, the length of bursts of writing evolves with the (ortho) graphic, lexical, and syntactic skills (for a review, see Olive 2014). Based on keystroke logging data, van Hell et al. (2008) examined the duration of pauses before various types of sentences considered to be more or less complex. They showed that pauses are longer at higher hierarchical levels; however, at the same hierarchical level, their length varies depending on the nature of the sequences that follow them. According to Medimorec and Risko (2017), the duration of writing pauses increases with the complexity of the language units that follow them. Drawing on the study of sequences that are segmented by pauses and their linguistic nature, Cislaru and Olive (2018) found that, in more than half of the cases, writers pause at the end of sequences that are not syntactically saturated, i.e., that do not take the form of a word, phrase, or clause. Some of these unsaturated bursts are due to the revision process, which uses language sequences to modify the text already produced.

3. Writing and Sentence Processing in Children

3.1. Sentence Conceptualization and Processing in Children

As already mentioned, segmentation and the sentence are key notions for mastering literacy. Ferreiro (1978) demonstrated that the first representations of the sentence in children, aged 4, do not include any segmentation criteria. While typical children aged 10 have not yet fully acquired writing conventions, they are already capable of an extensive and cohesive textual output. Focusing on the French context, graphically marked sentences constitute reading cues in text segmentation strategies (Masseron 2019) and assessment standards for the quality of written texts (Rondelli 2013). The sentence is taught at school (Gerlaud 2016), and 9- to 11-year-olds need to be able to produce a real story, understand syntax and word order, and structure a sentence by integrating several elements (Kail 2000). In terms of reading, children of this age are already expected to recognize these units and to modulate their intonation accordingly. In a descriptive study aimed at identifying sentence conception in children producing written text, Berninger et al. (2011) showed that, while most children in the first year of school manifest syntactic awareness and can write a complete sentence, up to 9% produce fragments or run-ons (two or more main clauses unseparated typographically).

According to Arfé and Pizzocaro (2016), sentence generation measures are the most sensitive to developmental and individual differences in writing. As demonstrated by Dockrell and Connelly (2016), while children generate more correct sentences orally, written sentence generation is significantly associated with reading and spelling and not with standardized oral language measures. In a longitudinal study of spelling and text composition

in 3rd- to 4th-grade French children, Bressoux et al. (2023) noted inter-individual differences linked to lexical spelling difficulties (evidenced by a higher number of lexical errors) that were associated with shorter texts, while morphological errors only had a negative effect on the completeness of the texts.

3.2. Writing in Children with Dyslexia

Dyslexia affects 5–10% of the school-age population (Peterson and Pennington 2012) and is found more frequently in boys than in girls (ratio 3:1) (Rutter et al. 2004).

The most common hypothesis behind the effects of dyslexia is the presence of phonological impairment, which suggests that children with dyslexia fail to learn to read because they do not acquire the ability to make connections between mental representations of letters (graphemes) and speech sounds (phonemes). Spelling checks and spelling errors play an important role in the writing process of dyslexic children. However, some authors argue that higher-level language weaknesses underly the phonological deficit (Snowling 2001; Snowling et al. 2020), while others insist on the role of semantic and syntactic deficits (Bishop and Snowling 2004) and show that more than one deficit converges in dyslexia, such as phonological and naming-speed deficits (Krasowicz-Kupis et al. 2009). The question of discourse segmentation is thus all the more acute as many phenomena seem to interfere with the perception and production capacities of dyslexic children. Berninger et al. (2008) showed that children with dyslexia have significant spelling difficulties and are slow in handwriting. At a linguistic level, this shows a sublexical phonological deficit (Sprenger-Charolles and Serniclaes 2003), and it can be argued that the fragmentation of the typical processual units is solely imputable to this. Using keystroke logging data to compare groups of dyslexic and non-dyslexic children aged 11, Morken and Helland (2013) showed that the same cognitive skills affect reading and writing in dyslexia, and while revisions take longer in dyslexic children, the result is poorer than in typically developing children. Sumner et al. (2013) argued that children with dyslexia are poorer writers because they pause more often. Along similar lines, Lalain et al. (2012) reported longer pauses when reading and speaking in dyslexic children and hypothesized difficulties in planning semantic and syntactic units. Frequent pauses and misspelling were also shown to be related to shorter bursts produced by 11-year-old children with specific language impairments (Connelly et al. 2012).

Research questions and purpose

Given the artifactual nature of the sentence on the one hand, and the segmentation of the information flow into infra-sentential units observed in previous studies of oral and writing processes on the other, what are the basic units of written language processing? Which linguistic patterns do children rely on to incrementally package information when writing? How can these patterns inform the modelling of language processing? We expect both typical and dyslexic children to operate with sub-sentence units during the writing process. A better understanding of written production strategies will shed light (i) on the nature of the mechanisms underlying sentence and text production, and (ii) on how syntax intervenes in language processing at this level of production. Comparing sentence production in typical and dyslexic children will allow us to gain new insights regarding how dyslexia impacts writing performance, especially with respect to the handling of basic production units and the potential impact of spelling and incremental parsing on syntax processing.

4. Materials and Methods

4.1. Data

For the present study, we focused on two groups: typical children and children with dyslexia (mean age 10). Our data are extracted from a larger project that focuses on the construction of written texts.

Typical children. This group included 29 children in CM2 (the 5th and final year of primary school). In this grade, children are generally 10 years old, which was the case

for 24 of them, while 3 were already 11, one was 9, and one was 12. There were 15 girls and 14 boys in the group. The children were asked to write a story within a limited time (20 min) under the usual conditions of their class. We obtained 29 narrations of different length, nature, and quality, as is usually the case in a class of all-ability children.

Children with dyslexia. We tested 11 children with dyslexia, who were recruited from the Child and Adolescent Psychiatry Department, Robert Debré Hospital (Paris, France). Inclusion criteria were as follows: a normal mean intelligence quotient (IQ between 80 and 115, as evaluated using the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV); normal vision; and a reading age (assessed using the ELFE test (Evaluation of Fluency Reading), www.cognisciences.com (accessed on 20 May 2023), Grenoble) of at least 2 years behind the chronological age. The group of dyslexic children comprised only boys; nine of them were 10 years old, one was 13 and another was 14 years old. They were asked to write a story based on a series of pictures (see Appendix A). We thus obtained 11 short narrations.

Data collection. All writing data were produced in French. They were collected with a keystroke-logging program, Inputlog (Leijten and Van Waes 2013). The software recorded data such as keystrokes, mouse movements, the temporality of the process (pauses, production time, etc.), the written units produced, and the revision operations affecting them (deletion, reformulation, moving, and addition). Thanks to real-time recording, we had access to two different datasets: the texts produced by the children and the elements of the production process, in line with the approaches developed in Cislaru (2015). In both groups, the collected texts gave rise to two sub-corpora: an original corpus (OC, Appendix B), in which the original formatting and writing were preserved, and a corrected corpus (CC, Appendix C), in which spelling, grammar and text structuring were reviewed by a corrector with regard to French academic standards. It should be pointed out that the correction of the texts was carried out without any problematization of the notion of sentence and was therefore neutral from this point of view.

4.2. Product Analysis: Sentence Description in Finished Texts

The finished texts allowed us to see how sentences were mobilized in efforts to structure them. In order to gain an overview of sentence management by 10-year-olds typing on a computer, we first identified the number of marked sentences signaling an intentional act on the part of the enunciator in the original corpus. This was performed by automatically counting strong punctuation marks. Capital letters were also considered as potential—but not sufficient—markers of sentence segmentation (see also Berninger et al. 2011). Sequences introducing direct speech were counted as simple sentences when typographically separated from the reported speech sequence.

Once all the marked sentences were identified, we defined sentence subtypes in order to determine the degree of syntactical complexity handled by children. In doing so, we followed the categories usually taught to children at school. Marked sequences containing a conjugated verb, or equivalents such as verbless sentences (non-attested in the corpus), were considered as simple sentences. Sequences containing a conjugated verb and connected by a coordinating conjunction ("such as", or "and", etc.) were considered to be coordinated sentences, which was in accordance with the categories taught to children. Constructions containing at least two conjugated verbs and one subordinate clause, whether embedded or not, were considered to be complex sentences. Given the large number of juxtaposed constructions and the diversity of relationships implied by parataxis (Béguelin et al. 2011), we decided to include them in this category as well. Drawing on Berninger et al. (2011), we also took into account potential fragments (under-sentence units) and run-ons (two or more concatenated main clauses with no typographical demarcation).

The data obtained were compared to the corrected texts to assess the degree of correspondence between the syntactic structure marked by the children and that identified by the correctors.

4.3. Process Analysis: Bursts of Writing and Revision Processes

The processual data allow us to observe the duration and location of pauses, bursts of writing, and revision operations.

A burst of writing is defined as a language sequence produced during the writing process and segmented by two pauses (Chenoweth and Hayes 2001). We consider bursts as the raw material used to compose texts (Kaufer et al. 1986; Cislaru and Olive 2018). To identify bursts of writing in the Inputlog output, we exploited the inter-key intervals (IKIs), as computed from the recorded keylogs, to segment the writing flow. We used individualized thresholds for each group. These individualized thresholds were computed in the following way: we pooled all the IKIs of a group together, and looked for the quantile associated to 2s, which is a shared reference value in the literature (cf. Chenu et al. 2014). Then, we applied this quantile to the IKI distribution of each individual distribution to distinguish the pauses. Therefore, instead of having a fixed threshold, we have a fixed proportion of pauses in the IKIs. We distinguished three types of bursts:

- Bursts of writing (P-burst), aiming at producing text: *[pause] but now she [pause]*;
- Revision bursts (R-burst), aiming at modifying previously produced text: *[pause] and the last one [pause]*—followed by more than one burst—*[pause] then [pause]*, where 'then' modifies the first burst as 'and then the last one';
- Immediate revision bursts (RB-burst), which modify the last word(s) of the immediately produced burst: *[pause] the [pause] the other [pause]* ('the other' replaces 'the').

Our approach to revisions was product-oriented and focused on the effect of the revision on the text and sentence segmentation (Faigley and Witte 1981). In order to study revisions, we relied on two types of revision bursts as well as the calculation of all revision events in the text at the character level, as manifested by one or multiple successive strokes of the backspace key: La⌫⌫Je␣pense␣que␣la␣violense⌫⌫ce␣. A visualization of the data exploited at this level of processing is proposed at [PROTEXT WIP (http://syled.univ-paris3.fr/protext/PLAY-TEXTE/CORPUS-ENFANTS-Poitiers/index-inputlog.html)].

In order to assess the effect of revision processes on the product, we took into account the morphosyntactic nature of the initially produced sequences compared to the morphosyntactic nature of the finished product.

Linguistic data taken from both the product and process provide qualitative insights into types of sentences, the nature and segmentation of bursts, and the impact of revisions on text segmentation and sentence demarcation.

4.4. Statistics for Pause Analysis

To obtain a clearer understanding of the behavioral data from the process and their relationship to linguistic data in the two groups, we carried out a few statistical analyses. First, we attempted to characterize which contextual factors cause pauses to occur. To do so, we defined six simple contexts: within a word (baseline condition), between two words, after a weak punctuation mark, before and after a strong punctuation mark, and before a sequence of revision events (e.g., one or multiple successive strokes of the BACK key). Since the participants produced very few weak punctuation marks, we subsequently dropped this factor. The probability of pausing in each context was assessed for each participant and was determined as the probability that a pause would be found in a given context (for instance, the baseline condition may be associated with a very low probability of pausing, but since there are many such contexts in the entirety of the text, it may be that a large quantity of pauses are found in that context). We refer to the first probability (probability of a pause given the factor occurrence) as the 'factor impact' (the impact of the factor on the production process), and to the second probability (probability of a factor given a pause) as the 'factor incidence'.

Second, since pauses are dependent on the definition of a threshold, and since this threshold depends on a common reference of 2 s that may not be equally relevant for both the typical and dyslexic children, we also wanted to rely on an analysis that was independent of pauses. To do so, we applied a parametric model of the entire distribution of

the inter-key intervals (IKIs), as computed from the recorded keylogs. For each participant, the distribution of these IKIs was modelled according to a two-mode Gaussian mixture configured in logarithmic space, as is customary in the literature, where these two modes were, respectively, associated with a fluent and a disfluent component of the writing process (Campione and Véronis 2002; Hall et al. 2022; Van Waes et al. 2021; Almond et al. 2012; Baaijen et al. 2012). This model is associated with five parameters: the means and standard deviations of each Gaussian mode, and the weight of the first mode. These parameters were therefore extracted for each participant for profiling purposes and to enable comparison between the two groups.

5. Results

This section begins by presenting the analysis of finished texts and then examines the writing process, including pause analysis, burst analysis and revision analysis.

5.1. Product Analysis: (Un)marked Sentences in the Original and Corrected Corpora

5.1.1. Typical Children

Overall, 3 out of 29 children did not use strong punctuation marks (=no marked sentences). However, one used capital letters for proper names, and another used a capital to mark the beginning of the text (examples 1 and 2 in Appendix B). Out of the total studied, eight children did not use capital letters to mark the beginning of sentences, but one of the eight children used capital letters for proper nouns, and two other children (three per total) used capital letters to mark the beginning of the text. This indicated that they were able to produce these marks, but that they did not always use them to mark sentences.

We then looked at the number of sentences in the corrected versions of the texts (see Appendix C). In many cases, the correctors decided to mark sentence segmentation. The corrected version of the texts showed that the children did indeed produce elements that made it possible to segment or at least reconstitute sentences, most often starting from the presence of a nominal group or equivalent and a verb (Table 1). However, the texts produced in the corrected version still differed in quality at this level, with writers who produced a limited number of syntactic structures and writers who provided a variety of syntactic structures that enabled sentences to be reconstructed.

Table 1. Descriptive statistics: texts produced by typical children, original, and corrected corpora.

	Original Corpus			Corrected Corpus		
Statistics	**Sentences**	**Words**	**Paragraphs**	**Sentences**	**Words**	**Paragraphs**
Total	177	4277	70	336	4357	57
Minimum	0	39	1	2	41	1
1st Quartile	4	104	1	8	107	1
Median	6	136	2	12	139	1
3rd Quartile	8	182	3	15	185	2
Maximum	23	321	12	21	314	10
Average	6.1	147.5	2.4	11.6	150.2	2
Variance (n − 1)	24.5	4008.1	5.3	24.6	3916.8	3.3
Standard deviation (n − 1)	5	63.3	2	5	62.6	1.8

The question is whether this disparity in the number of sentences (either marked or corrected) could be due to an overall disparity in the children's writing skills: children who are able to write a longer text in the context of the task should produce more sentences than children who write a shorter text. To check this, we compared to what extent the number of sentences and the number of words correlated, as they would in any normal text. If we considered the corrected sentences, we found such a correlation ($r = 0.73$, $p < 1 \times 10^{-5}$), as expected; yet, if we considered the correlation between the number of words and the number of actual, explicitly marked sentences, then the correlation collapsed and ceased to

be significant (*r* = 0.30, *p* = 0.12). This shows that the disparity in sentence marking strategy is not a by-product of the disparity in writing skills.

We thus took a closer look at the types of graphically marked sentences in the original texts; we identified simple, coordinated, and complex sentences.

Simple sentences. We counted 56 simple sentences, representing less than a third of the 177 marked sentences. Seven of these simple sentences represented reported speech and in most cases were introduced by simple sentences containing speech verbs such as "a dit" [said]. It seemed justified to consider about a fifth of simple sentences as being determined in their production by the change in enunciative plane. Many simple sentences were introduced by connectors such as "Et" [and] (6), "Puis" [then] (3), "Quand" [when] (1). "Et" and "Puis" are frequently used in oral speech to segment information and to introduce a succession of events. They are also frequent in writing by beginners, before starting to be replaced by more diverse options (Schneuwly 1988).

Globally, simple sentences do not appear to be the preferred material for composing texts, despite their renowned grammatical simplicity. Nor are simple sentences preferred for the incipit: in our sample, only three texts opened with simple sentences.

Coordinated sentences. There were 21 coordinated sentences. These used "et" [and], "mais" [but], or "car" [since, as], and one sentence also contained a paratactic structure. It is worth mentioning that while "car" is not specifically a coordinator from a linguistic point of view (for instance, it cannot coordinate two subject–noun phrases), it was used as such by the children. Two other coordinated sentences also included relative clauses (see complex sentences below). Coordinating relations at the clause level were nonetheless present in the data, but these were generally part of complex sentence units containing at least one temporal, causal, or complementary subordinate clause.

Coordination markers were also used to introduce simple sentences, as mentioned above.

Complex sentences. Among complex sentences, 39 contained parataxis. Parataxis can be marked by punctuation or can be left unmarked. Generally, the children who used parataxis did so more than once within the same sentence. Example 1, with four paratactic connections, illustrates this type of pattern (occurrences of parataxis are segmented by |; connectors are underlined).

(1) *il c'etait fait punir,* | *il disait que c'etait de la faute de trois fille,* | *une des trois fille s'appeler léa <u>mais</u> maintenant elle n'ais plus dans cette école* | *la deuxième s'appeler enola* | *elle aussi elle n'ai plus dans l'ecole <u>et puis</u> la dernière s'appeler amendine* [P4CM2N2, 10 yo, M]
he got punished, | he said it was the fault of the three girls, | one of the girls [' name] was léa but now she isn't at this school any more | the second was enola | she is not at this school anymore either and then the last was amendine

"Et" and "Puis" can also introduce complex sentences; some children used one or the other to introduce the majority of the marked complex sentences (P10CM2N2, P12CM2N2, and P25CM2N2).

More than one child combined several sequences with conjugated verbs within the same typographically marked sentence, whereas the adult proofreader segmented them into two or more sentences. The use of parataxis enables children to integrate several main clauses within the same sentence (Berninger et al. 2011) call them *run-ons*. Generally, this writing strategy is devoted to the development of the narrative plot, where events often follow one another right up to the denouement, as shown in the excerpt below:

(2) *Mais j'ai insisté il mon frapé et j'ai eu mal j'ai appelé la police quand elle est arrivé ils ont voulu s'échapé il n'on pas pu il m'on dit qu'il allait s'occupé d'eux et que je devait rentrait chez moi* [P14CM2N2, 9 yo, M]
But I insisted, and I was hurt I called the police when the police arrived they tried to escape, but they couldn't they told me they would take care of them and that I should go home

To summarize, the final texts show different writing strategies in terms of sentence production. There are inter-writer differences, which probably reflect different writing

skills and grammatical competences. However, we also note the deployment of some pattern strategies within the same text, which seem to be reproduced by several children. Writers who do not mark sentences apply the strategy of concatenation to the whole text (see Appendix B for examples). Written output can, to a certain extent, be related to actual sentences by adult editors, meaning that children do indeed produce 'proto-sentences', even if they are not explicitly marked. They do so in proportion to the amount of text they produce. However, the number of explicitly marked sentences is not significantly dependent on a text's length. If the length of the text produced can be considered a proxy for the children's varying writing skills, it means that the degree to which they have internalized the notion of sentence (as reflected by explicit sentence marking) is decorrelated from these writing skills. This raises the question of the representations of the sentence in didactic terms on the one hand, and that of the process of production on the other.

5.1.2. Dyslexic Children

Overall, 5 out of 11 children did not use strong punctuation marks (=no marked sentences). One, however, started a new line to segment the text. Altogether, seven children did not use capital letters to mark the beginning of sentences, but one out of the seven used capital letters to mark the beginning of the text. In two cases, where full stops were used to segment sentences but were missing at the end of the text, we decided to count the last sequence as a sentence as well. Given the greater age differences in this group, we paid particular attention to the age of the writers for all observations in order to avoid the bias of the age gap within the group. We noted that the 14-year-old produced a longer text (59 words, i.e., nearly one-third longer than the average), but did not segment it into sentences. No significant difference was noted for the other child, aged 13.

The ratio of children not producing marked sentences was higher than it was for typical children (45.5% vs. 10%); the difference between the two groups is significant, even though the sample sizes involved are small (p = 0.02 according to a one-sided Fisher exact test).

We then looked at the number of sentences contained in the corrected versions of the texts. Similarly to typical children, children with dyslexia produced enough elements to enable the corrector to segment or, at least, reconstitute sentences, most often starting from the presence of a nominal group or equivalent and a verb (Table 2).

Table 2. Descriptive statistics of the original and corrected corpora (dyslexic children).

	Original Corpus		Corrected Corpus	
Statistics	**Words**	**Sentences**	**Words Corrected**	**Sentences Corrected**
Total	393	23	430	44
Minimum	18	0	19	3
1st Quartile	30	0	32	3.5
Median	34	1	35	4
3rd Quartile	38	4	44	4
Maximum	59	5	67	6
Average	36	2.1	39	4
Variance (n − 1)	130	4.5	170	0.8
Standard deviation (n − 1)	11	2.1	13	0.89

While word segmentation also had to be corrected in some cases, in other cases it was impossible for the corrector to identify the words in the strings produced, as seen in Example 3. This seemed to impact not only the number of words but also the number of sentences in the corrected versions. The standard deviation was reduced mainly due to the reconstruction of sentences in the initially non-segmented texts.

(3)　*Il y a un garçon qui joue pour tirer sur la d[b]alle. Il tire en l'air et ça casse la vitre. Il cecivon se fait gronder. Le père les gronde et le père ramène une cèsse aouti [caisse à outils].* [10 yo; reading: 5th percentile]
There is a boy who plays at shooting the d[b]all. He shoots in the air and it breaks the window. He cecivon gets scolded. The father scolds them and the father brings back a cèsse aouti [toolbox].

Following the abovementioned classification, we distinguished three types of graphically marked sentences: simple, coordinated, and complex. However, in the texts produced by dyslexic children, parataxis was a preferred strategy compared to the use of complex sentences.

Simple sentences. We counted 9 simple sentences, representing nearly 40% of the 23 marked sentences—a higher percentage than found among typical children. They were attested in four texts, but two texts of average length contained six of them. Given the age gap in this group, it is important to point out that these texts were all produced by children of the same age, 10 years old. All the simple sentences had an SV(O) form. Simple sentences were used in the incipit in three texts, and in an additional text marked by a semi-colon, which was the opposite of what was observed in the texts produced by typical children. Connectors such as "Et" [*and*] or "Puis" [*then*], frequently used by typical children to introduce simple sentences, were absent in most of the texts, but in one "puis" was used ambiguously after a triple exclamation mark and with no capital letter, although the child used capital letters elsewhere in his text: "[Mr Durand] les gronde!!! puis répare sa fenetre" [*Mr Durand scolds them!!! then repairs the window*].

Coordinated sentences. The corpus contained 5 coordinated sentences, which were found in two of the texts using several simple sentences. The only conjunctions used for coordination were "et" [*and*], and in one case "mais" [*but*] was ambiguously used: "mais sauf que" [*but except that*]. The sentences were rather simple, and the subject was sometimes resumed, even in cases where it is usually omitted (Example 4):

(4)　*Tom joue a la bale. Emlie prene le balon <u>et</u> il case la viter. Emile et Tom ont prer. Le papa de Tom et Emile les gonde <u>et</u> il le répare.* [10 yo; reading: 5th percentile]
Tom plays ball. Emlie takes the ball and he breaks the window. Emile and Tom are afraid. Tom and Emile's dad scolds them and he fixes it.

Parataxis. Among marked sentences, 6 contained parataxis. Generally, children who used parataxis did so more than once within the same sentence (the same strategy was observed in typical children). In the sequence below, some paratactic structures (segmented by | here) connect simple SV sequences and some concern non-marked reported speech:

(5)　*harli jou o balon. il tir | il mark | il cas la.vitre jen di a harli | tu a fai une gros betis. son papa a di | vou save fe une betis. ge ve voir oooo vousave fe coi. la se grav les sanfen | coman on vaver pour acheter une nouvelle vitro* [9,7 yo; reading: 40th percentile]
harly plays ball. he shoots he scores he breaks the .window jean said to harly you've done something very stupid. his dad said you've done something stupid. i'm going to see what you've done. this is serious, kids, how are we going to buy a new window

Complex sentences. There was one explicitly marked complex sentence, which was rather a collection of coordinated sequences, with one relative and one temporal clause. The uses of "mais sauf que" and "puis" mentioned above could also be counted here.

If we look at the syntactic structure of the five texts in which sentences were unmarked, we can see that only paratactic or coordinated constructions were used: for instance, the longest text (59 words) contained five coordinators that related verbal constructions. In this respect, the output of dyslexic children differed from that of typical children who, even if they did not mark sentences graphically, used complex syntactic relationships and employed connectors accordingly.

5.2. Process Analysis

In this section, in order to observe the dynamics of spontaneous segmentation during the writing process in relation to sentence production and marking sentences, we give an overview of writing events and propose an analysis of bursts and revisions.

5.2.1. Process Analysis

- Overview

Table 3 points out some significant variations between the two groups, particularly in terms of the number of keyboard events and bursts of writing. Note that the number of words applied to the data from the process does not match the number of words in the product. Defining words for the process data is a particularly difficult task, notably due to the pervasiveness of spelling revisions. To offer an operational and automatable definition, we considered any string of characters produced between spacing characters to be a word. This rather restrictive definition explains the lower number of words in this table.

Table 3. Global descriptive statistics of the process data: typical/dyslexic children.

	Number of Events	Number of Words	Number of Sentences	Number of Revisions	Number of Bursts	Mean Burst Length
Mean	455/305	50/37	2.2/2.5	24/18	47/78	11.4/4.1
Standard deviation	181/183	24/20	2.9/2.2	11.7/8.6	21/46	8.4/1.2
Minimum	121/188	15/14	0/0	3/6	10/26	6/3
Q1	314/208	30/26	0/0	17/13	31/54	7/3
Median	448/242	46/30	1/4	24/17	43/62	8/3
Q3	558/297	69/37	3/4	30/20	55/82	10/4
Maximum	872/854	98/89	11/6	58/40	102/208	51/7

Next, we considered the IKI distribution parameters for each participant. To account for the small sample sizes involved, a Mann–Whitney U test was performed for each parameter to compare the parametric values of the two groups (typical and dyslexic), for which the corresponding *p*-values are reported. We also report the effect size, as assessed using the rank-biserial correlation *r* (Table 4). The difference between the two groups proved to be significant only for the mean of the second mode, even accounting for the Bonferroni correction for multiple comparisons (*p* values should then be below $0.05/6 = 0.008$). Moreover, the effect size (-0.54) is reliable (its magnitude is higher than 0.5). This difference leads to an insightful interpretation: under the hypothesis that the two modes refer to a fluent/disfluent dichotomy (Roeser et al. 2021), the weight of the first mode does not display any difference between the two groups, which means that the proportion of the disfluent component in the IKI distribution is the same for both. What changes, however, is the typical duration of these disfluencies, which is longer for dyslexic children. For dyslexic children, the disfluencies that inevitably arise are more problematic and more systematically disrupt the production process.

Table 4. Average and standard deviation of each parameter for both groups.

Parameter	Mean (Typical)	Standard Deviation (Typical)	Mean (Dyslexia)	Standard Deviation (Dyslexia)	Mann-Whitley U (*p*-Value)	Rank-Biserial *r*
Mode 1	0.37	0.13	0.48	0.27	0.65	-0.10
Standard deviation 1	0.46	0.11	0.58	0.14	0.02	-0.49
Mode 2	1.30	0.51	1.88	0.57	0.0077	-0.54
Standard deviation 2	0.68	0.09	0.72	0.13	0.25	-0.70
Mode 1 weight	0.60	0.10	0.59	0.10	0.65	0.09
% of revisions	0.12	0.08	0.08	0.04	0.09	0.35

- Grammatical boundaries and factors impacting pausal segmentation

We therefore looked at the impact of several factors on pausal segmentation, focusing on grammatical boundaries. For each participant, we assessed the weight of each factor in relation to the probability of a break exceeding the threshold for a multivariate logistic regression model featuring a baseline value and these five factors. The resulting analysis is summarized at the group level by Figure 1 and in the corresponding descriptive statistics.
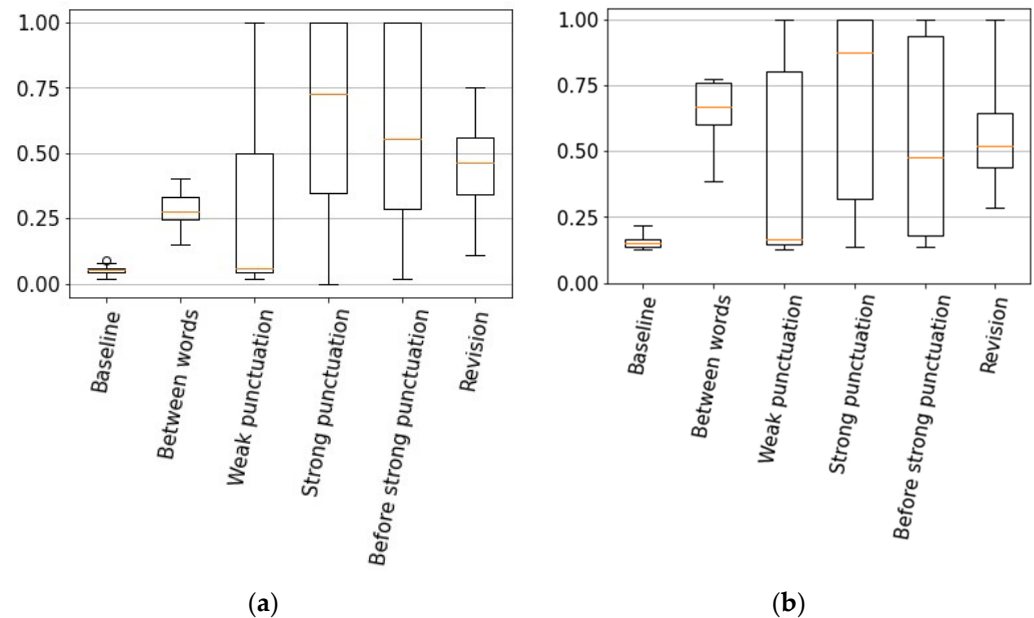

(a) (b)

Figure 1. (**a**) P-burst knowing factor (typical children); (**b**) P-burst knowing factor (dyslexic children).

The weight of strong punctuation was highly variable, notably because some participants systematically stopped before or after a period, while others did not produce any periods at all. The differences between the two groups are summarized in Table 5. The weak punctuation factor did not play a significant role here because weak punctuation marks remained a rarity in the children's texts.

Table 5. Factor impact comparison.

	Baseline	Between Words	Before Period	After Period	Before Revision
Mean (typical)	0.05	0.28	0.58	0.64	0.45
Mean (dyslexia)	0.15	0.64	0.5	0.63	0.59
p-value	8.6×10^{-10}	3.5×10^{-9}	0.86	0.63	0.26
r	-1	-0.99	0.04	-0.11	-0.33

This analysis evidences two main results. First, for typical children, the period is a very reliable predictor of bursts, and the occurrence of a revision event also tends to produce a halt. For dyslexic children, the probability of stopping between and within words is also very high, and significantly more so than for typical children. This connects to the observation, noted above, that the number of events per burst is very low (between 3 and 7; cf. the average burst length in Table 3), and that the second mode in the 2-mode Gaussian mixture analysis is longer for dyslexic children. This may explain why we find a greater abundance of breaks in the two contexts where typical children are not so likely to pause (baseline and between words). This suggests that, in dyslexia, any kind of disfluency (e.g., spelling hesitation, lexical retrieval), that is, any active cognitive monitoring, has the potential to disrupt the production process.

However, if a context is rarely encountered throughout a text, it will not be able to explain much of the segmentation into bursts, even though it has a strong influence on the

likelihood to pause. To quantify this, we computed the proportion of pauses found in each of our six contexts. The results are displayed in Figure 2 and are summarized in Table 6.
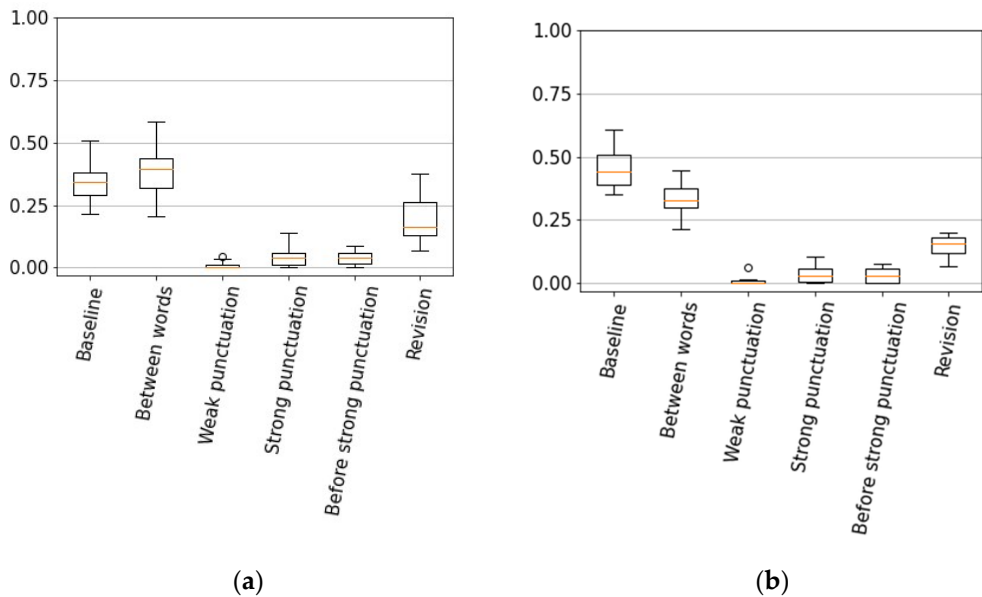

Figure 2. (**a**) P-factor knowing burst (typical children); (**b**) P-factor knowing burst (dyslexic children).

Table 6. Factor incidence comparison.

	Baseline	Between Words	Before Period	After Period	Before Revision
Mean (typical)	0.34	0.38	0.04	0.04	0.19
Mean (dyslexia)	0.46	0.33	0.03	0.03	0.15
p-value	7×10^{-5}	0.06	0.37	0.44	0.23
r	-0.77	0.39	0.19	0.17	0.26

First, these results corroborate and quantify the observation that, although sentence boundaries attract pauses (see Figure 1), they do not exhaust the segmentation of the production process. In the group of typical children, the probability that a pause accompanies a strong punctuation mark (either before or after producing it) is 8% on average and does not exceed 12%. By contrast, revisions and pauses between words account together for the majority of pauses (58%).

Second, the only significant difference between the two groups (with a threshold at 0.01 because of the Bonferroni correction) is the greater abundance of pauses found in the baseline condition. This derives from the previous analysis: since the baseline probability to pause is higher in dyslexic children, they have a greater propensity to pause anywhere; and since the most prevalent context to pause in a text is within words, we witness a substantial increase in the proportion of such pauses. The high prevalence of these pauses within words explains why the processual units become degraded.

One caveat stemming from the analysis is that the two groups are not associated with comparable numbers of pauses. Indeed, the threshold is set with respect to a fixed quantile in the IKI distribution, determined separately for each group, as explained in Section 4.3. As a result, the proportion of pauses among the IKIs is equal to 8% for the typical children, and 17% for the dyslexic children (this proportion is determined by looking at the percentage of pauses above 2 s when the IKIs are pooled over the whole group). Therefore, this over-abundance of pauses among dyslexic children compared to typical children may inflate the proportion of pauses occurring within words, which are usually shorter. In order to test the potential impact of this discrepancy between the two groups, we raised the threshold for dyslexic children until the proportion of pauses among the IKIs matched that of the

typical children (92% of non-pauses IKI). After this was performed, the mean proportion of pauses occurring within words dropped from 0.46 to 0.40, the difference with the typical children was no longer significant (p = 0.18), and the effect size dropped from −0.77 to −0.28. Therefore, it would seem that the increased proportion of baseline pauses was in part due to the shift in the IKI distribution toward higher values, as evidenced by the higher values for the second mode mean in the dyslexic group.

5.2.2. Burst Analysis

We now take a closer look at bursts of writing, which are the empirical data derived from pausal segmentation.

- Typical children

The 29 texts were produced in 3302 bursts, with 114 bursts per text on average. This represented about 19 bursts per sentence produced. The average was skewed by cases where no sentence was typographically marked, or by complex sentences using parataxis to concatenate several main clauses with their subordinates.

Overall, 28% were revision bursts, which modified a previous passage on the same computer screen. Additionally, 14% were immediate revision bursts, which modified the right border of the burst preceding them.

Example 6 illustrates burst segmentation via the production of non-marked sentences. The asterisk (*) marks the location of the pausal segmentation, and the brackets signal modified sequences. Sentence frontiers are marked in accordance with the corrected version of a text [|]. The observations in this case study are broadly valid for all the children's productions in our corpus. We can see that pausal segmentation does not follow syntactic segmentation and that it produces linguistic sequences at different levels of syntactic analysis. However, sentence frontiers tend to be associated with pauses. There are certain trends within this, such as pausing before and sometimes after the subject of the sentence, as observed by Feltgen et al. (2023), but these are not systematically detected. Reported speech tends to be produced separately, as observed by Cislaru and Olive (2018).

(6) *il c'*etai*t fait punir*, | il* disait que c'etait de la faute de trois fille* ,* | une des trois fille s'appeler* léa* mais maintenant elle* n'ais plus da*ns *[dans] cette* école* | la deuxième s'appeler* eno*la*| elle aussi elle n'['*'a]*ai plus dans l'ecole* et [puis] la derni*ère s'appeler* [puis*] amendine [P4CM2N2, 10 yo, M]
he got punished, | he said it was the fault of the three girls, | one of the girls[' name] was léa but now she isn't at this school any more | the second was enola | she is not at this school anymore either and then the last was amendine

Table 7 illustrates the writing process, including spontaneous segmentation into bursts, the length of the pause before each burst, and the length of each burst in terms of number of characters. To preserve overall legibility, we have only marked the most important revisions.

Long pauses, i.e., pauses at least three times longer than the reference duration (in italics in Table 7), often occur before a sentence subject, a strong punctuation mark, or a connector. In some cases, these data confirm Medimorec and Risko's (2017) results concerning pause rates at sentence boundaries predicting sentence length, but they do not do so systematically. For instance, adding the connector "puis" [then] after "et" [and] induces a long pause (19.7 s), although there is not, strictly speaking, a sentence boundary. Revisions may also impact pause location and length. The end of the sentence, starting with "et" [and], is produced after 69 intervening bursts.

To sum up, even though sentences do segment the text into units that are relevant from a processual perspective, the segmentation into sentences does not exhaust the much wider diversity of processual units that surface in textual production.

The question is, how is sentence marking—when it is effective—integrated into the writing process? The analysis of the production of strong punctuation marks shows that, out of a total of 177 sentences produced by typical children, 131 strong punctuation marks were produced as a single burst, 18 as a burst beginning with a strong punctuation mark, and 2 as a

burst with punctuation in the middle. These results are consistent with the findings regarding writing in adults (Cislaru and Olive 2018, 2021). It is important to note that more than half of these were production bursts, while a third were revision bursts. However, there were significant differences between writers, with some preferring to produce revision punctuation bursts, and others producing punctuation bursts. These data suggest that producing a strong punctuation mark needs a decision of completion that is related to a delayed sentence closure, which may be interpreted as intention/decision to make a sentence.

Table 7. Burst segmentation and pause length during the production of a complex sentence (Example 6). Dashed lines indicate the end of a sentence according to the adult editor. Revisions are in red strikethrough. Long pauses (at least three times longer than the threshold) are marked in italics.

Pause Length (In Seconds)	Burst Length (N° of Characters)	Burst
13.775	6	il c's
1.264	4	étai
1.638	13	t fait punir
3.135	6	~~quand~~
2.465	2	~~sa~~
2.106	1	
4.446	0	
1.497	2	~~d~~
2.277	3	il
1.435	64	disait que c'etait de la faute de trois fille ~~une qui s'appeler~~
21.56	0	
2.403	0	
7.69	1	,
2.293	30	une des trois fille s'aooekler
1.124	0	
5.694	7	ppelerv
1.451	1	
1.077	4	léa
50.373	21	mais maintenant elle
1.341	14	n'ais plyus da
9.454	0	
1.996	14	us dans cette
1.092	6	ecokle
1.17	3	~~le~~
4.929	8	~~l'autre~~
1.95	1	~~l~~
1.17	4	a de
1.154	14	uxième s'appel
2.917	1	e
2.527	1	~~r~~
2.137	1	
8.58	4	eno:
1.436	3	la
1.888	11	elle aussi
1.934	9	elle n'''a
1.233	0	
1.498	21	'ai plus dans l'ecole
14.804	1	
59.499	1	.
69 bursts later		
17.987	12	et la derni
2.138	14	ère s'appeler
19.687	5	puis
4.992	1	
2.901	8	amendine

- Dyslexic children

The corpus produced by dyslexic children included 685 bursts, among which were 39 revision bursts (5.7%) and 73 immediate revision bursts (10.7%). Five children produced several revision bursts and two children, respectively, produced one and two R bursts each. Four children did not produce any R bursts. In contrast, all the children produced immediate revision bursts. Compared to typical children, dyslexic children produced 5-fold fewer revision bursts, but only nearly one-quarter less immediate revisions.

Example 7 and Table 8 illustrate the writing process behind a text composed of marked sentences. In Example 7, pauses are projected onto the text and marked by an asterisk (*); square brackets indicate immediate revisions:

(7) *Tom* joue* a la* b*ale*.*[eilme*Eilme*] [E]mli*e *p*r*ene* le* balon* et* il* ca*se* la* viter*. *Em*ile* et* [t]om*[T*] ont* [p*r*e*?*?*] p*rer* .* Le papa de* Tom* et Emile* les* gon*de* et il l*e* r*é*pare*.
[10 yo; reading: 5th percentile]
Tom plays ball. Emlie takes the ball and he breaks the window. Emile and Tom are afraid. Tom and Emile's dad scolds them and he fixes it.

Table 8. Burst segmentation and pause length during the production of a text by a dyslexic child (Example 7, [10 years old; reading: 5th percentile]). Sentence ends are marked with dashed lines. Long pauses (at least three times longer than the threshold) are marked in italics.

Pause Length (In Seconds)	Burst Length (N° of Characters)	Burst
	4	Tom
5.550	4	joue
3.736	1	
3.252	5	a la
2.161	1	b
1.705	4	ale
12.914	−1	
1.991	2	.
8.414	6	eilme
2.061	−1	Eilme
1.923	−2	Mli
1.626	2	e
39.371	1	p
12.996	1	r
6.958	4	ene
3.272	3	le
4.293	6	balon
4.738	3	et
6.295	3	il
3.023	2	ca
2.895	3	se
3.024	3	la
11.579	6	viter
5.443	−1	
1.537	2	.
5.274	2	Em
3.748	3	ile
2.430	1	
1.834	3	et
2.984	4	tom
5.592	0	T
28.092	4	ont
6.456	1	p
8.736	1	r

Table 8. *Cont.*

Pause Length (In Seconds)	Burst Length (N° of Characters)	Burst
2.806	1	e
3.510	1	?
4.780	1	?
4.172	−5	
15.371	1	p
1.576	3	rer
2.410	1	
2.975	1	
10.938	2	.
5.871	11	Le papa de
3.719	4	Tom
1.924	9	et Emile
4.244	4	les
3.917	3	gon
3.075	3	de
6.853	7	et il l
2.480	2	e
1.696	1	r
7.666	1	e
2.479	4	pare
2.509	1	.

Table 8 illustrates the spontaneous segmentation into bursts and lists the length of pauses before each burst.

Bursts are very short, often being shorter than a word, not least because of their nature as immediate revisions correcting or attempting to correct spelling. With the exception of a pause between the subject and the verb, the longest pauses (in italics) are located before the production of words whose spelling is incorrect or subject to immediate correction. As shown in this example, sentence boundaries do not constitute a privileged attractor for long pauses, which echoes the observation made for typical children.

Bursts and the production of strong punctuation marks. The above example and table also show the way punctuation marks are segmented per se during the writing process, as hinted by the Gaussian mixture. Indeed, out of a total of 23 sentences produced by children with dyslexia, 16 full stops, 2 question marks, and 1 triple exclamation mark were produced in a single burst. One occurrence concerned a burst finishing with a full stop. Four of them were revision bursts (three full stops and a question mark), one was an immediate revision burst, and the rest were all production bursts. It could be noticed that, as in the typical children, punctuating sentences was a full-fledged task for the writers. Proportionally, children with dyslexia used a greater variety of strong punctuation marks, but the data collected are too limited to draw any conclusions.

5.2.3. Revision Analysis

Significant differences were noted between the writing processes developed by typical and dyslexic children. The revision processes did not seem to be deployed in the same way by the two groups. It is important here to understand how they affected sentence processing.

- Typical children

The revisions made during the writing process affected a number of areas: lexical or grammatical spelling, lexis and its reformulations, syntactic structures, or typography. It could even be the complete rewriting of the text (change of narrative). However, revisions affecting syntactic structures were not very frequent in our corpus.

The excerpt below shows how, even when the child makes more substantial revisions that go beyond typographical or spelling hesitations, the syntactic structure is preserved (revisions are in red strikethrough). Thus, changing "avaient un" [*had a*] to "connaisseaient

un" [*knew a*] and "était méchant" [*was bad*] to "s'appelait Louis" [*named Louis*] does not change the syntactic macro-structure SVO+Relative Clause.

(8) 0:<P~~IER~~>:2
1:<ierre et Paul étaient deux a>:30
29:<mis>:33
32:<. Ces deux amis ~~avaient un~~ >:49
48:<connaiss>:57
56:<aient un garçon qui >:77
76:<~~était méchant~~>:77
76:< s'appellait Louis >:96 [P28CM2N1, 10 yo, F]

In the rest of the text by the same writer, a potential change in the syntactic structure was identified: the possibility of a relative clause "qui é{tait}" [*who w{as}*] was abandoned for anaphora using the 3rd-person pronoun "il" [*he*], and finally replaced by the demonstrative use of anaphora with the noun phrase "ce dernier" [*the latter*]. This change could be explained by the avoidance of a second relative clause. However, while this final modification impacted the relationships between clauses (the relative clause was abandoned for parataxis), it did not change the local syntactic structure:

(9) 95:<, ~~qui é~~>:98
97:<~~il~~>:98
97:<ce dernier était méchant. ~~Il embéter~~>:125
124:<l taper les autres et n'avait pas d'amis. [P28CM2N1, 10 yo, F]

All in all, in the 29 texts produced, we only identified 13 modifications involving a change in local syntactic structure (Table 9). This number should be compared to the number of events and number of revisions mentioned above (Table 3).

Table 9. Detail of revisions having an impact on the syntactic structure.

Initial Version	Modification	Impact
Ce j[our] [*this day*]	Elle [*she*]	Temporality → subject
ses parents [*her/his parents*]	/	Deletion
Une fois [*once, on day*]	Je suis Lilou [*I am Lilou*]	Temporality → subject (incipit)
Il a [*he has*]	Un [*a*]	Sentence base → article (Noun Phrase)
Je ma [*I my/myself*]	Il était une fois [*once upon a time*]	Subject → temporality (narrative form)
mais les [*but the*]	le petit dit [*the little one says*]	Coordination → parataxis
Mais [*but*]	et après [*and then*]	Coordination → coord. & temporality
ne revient plus jamais [*never ever returned*]	il devient un [*he became a*]	Verbal clause → sentence
les trois filles s [*the three girls …*]	il disait les trois filles [*he said the three girls*]	Main sentence → Reported speech
que c'était [*that it was*]	que les filles se moquer [*that the girls were mocking*]	Cleft sentence → complement clause
qyuand le maitreudertiam [*when the teacher*]	qui tape [*who hit*]	Temporal → relative clause
Il [*he*]	et y en a un [*and there is one*]	Subject → coordinated sentence
EN [*in*]	et il [*and he*]	Location → coordinated sentence

Six rewriting events mark some hesitation after the production of the first letter of a new phrase, as in text 2, where "L" is abandoned for "Des foie il y a des élèves…" [*Sometimes there are pupils…*]. "L" might be interpreted as the first letter of the definite article "Le", potentially introducing a subject, and it is also replaced by the temporal marker "Des foie" [*sometimes*], followed by the presentative "il y a" [*there are*]. The six rewriting events should be compared to the number of marked sentences and take into account the fact that only 177 sentences are explicitly marked in these productions, while the textual contents enable the number of sentences to be practically doubled in the corrected versions, where 336 sentences are reconstituted.

The fact that revision most often occurs at a sub-sentence level, within the syntactic frame of the sentence, indicates that processual units are not equivalent to sentences but constitute a lower level of organization.

- Dyslexic children

Two features of dyslexic children's productions prevent a true analysis of revisions focusing on sentence structuring: the limited number of revision bursts and the production of very short bursts. These factors make it impossible to anticipate the projected syntactic structures underlying a processual sequence. Dyslexic children concentrate on spelling. It should be noted, however, that one child corrected a lower-case letter into a capital letter, thus marking the beginning of a sentence, and that five occurrences of strong punctuation marks were recorded as revisions.

Although language processing is impaired (as exemplified by the more fragmented bursts and the spelling difficulties), the segmenting of the text into sentences remains operational for some children and may be subject to revision. However, the difficulties above seem to impact the nature, length, and level of complexity of the sentences produced.

6. Discussion

Our two-fold product-process study of the way children handle sentences allowed us to point out the tension between the sentence as a production objective and spontaneous segmentation into bursts of writing.

6.1. Sentences in the Finished Product

Sentences are seldom graphically marked in written texts produced by typical and dyslexic children. Sentence demarcation is not acquired by all children, and this is even more true for dyslexic children, although the latter produce shorter texts that would allow them to take the time to produce capital letters and strong punctuation marks.

Despite the fact that sentences are intended to play a role in text segmentation, most children draft the information flow through elocutionary periods when writing, in line with pre-sentence norms (see Combettes 2011), and do not systematically parse by means of sentences, proceeding instead as in oral speech (Blanche-Benveniste et al. 1990) by using concatenated main clauses, which are called run-ons by Berninger et al. (2011). While the percentage of run-ons in typical children corresponded to the results of Berninger et al., in our data, dyslexics were four times more likely to operate using run-ons.

The corrected versions of the texts highlight the representation of linguistic norms in terms of written text, syntax, and sentences. While many typical children produce linguistic strings that are perfectly adapted to the expectations of oral communication, their proficiency in writing sentences remains limited. Even children who typographically mark sentences using capital letters and strong punctuation marks sometimes fail to segment them into sequences corresponding to the written production norms applied by correctors. These results confirm the idea of the sentence as a grammatical artifact that is intentionally demarcated by the enunciator (Le Goffic 2011; Van Raemdonck et al. 2011). In the case of dyslexic children, the relationship between phonological and syntactic processing remains to be elucidated.

Previous work on English data found that single independent clauses were more likely to occur in multi-sentence constructions (like texts), while single-sentence constructions favored the production of an independent clause plus a dependent clause (Berninger et al. 2011). We observed, on the contrary, that simple sentences were not preferentially produced by the typical children; however, they were more strongly favored by the dyslexic children. This difference, which could be due to the dynamics of production at work through shorter sequences in dyslexics, would need to be verified using other corpora.

While simple sentences are not the preferred material for a text's incipit and composition, it may be hypothesized that segmenting information into simple sentences when writing a story is difficult for children. The use of complex structures in typical children, and in some rare cases in dyslexic children, to set out the development and climax of the plot, even going so far as to produce a series of several main propositions thanks to parataxis, suggests that sentence segmentation obeys semantic–textual representations. For the child writer, the plot functions as a unit of meaning, which is semiotized in a typographically marked sentence. This runs counter to the hypothesis of the sentence being the most plausible unit of conceptual and linguistic planning (Foulin 1998).

6.2. Language Processing When Writing: Infra- and Supra-Sentential Units vs. Sentences

When writing, children spontaneously segment the information flow through bursts of writing. Our data validate the function of bursts as parsing units in production (Kaufer et al. 1986; Chenoweth and Hayes 2001). As empirically attested in linguistic sequences segmented by pauses in language production, bursts echo the segmentation principles of information flow identified by Chafe (1992) and Sinclair and Mauranen (2006).

The relationship between burst segmentation and sentence demarcation is similar for both groups of children. Sentence boundaries are often marked by pauses, which occur most of the time before the strong punctuation mark—if any is used. There are proportionally as many pauses found before/after periods in dyslexic and in typical children. Although the sentence segmentation of a text typically overlaps with pauses, these pauses amount to a marginal proportion of the pauses that delimit processual units. Also, pause duration is not necessarily higher at the sentence boundary compared to segmentation within the sentence. Thus, in writing, the sentence does not appear to be the basic unit of language production.

During the writing process, typical children revise a lot, but very few revisions affect the syntactic sentence structure; a limited number of revisions affect syntactic relations between clauses due to the decision of whether or not to use a connector. The persistency of the syntactic structure throughout the process could confirm, at least partially, the validity of the first functional step concerning the establishment of a syntactic schema during language processing (Garrett 1975, 1980; Bock and Levelt 1994). Dyslexic children revise two to five times less, and only in a few isolated cases revisions introduce sentence boundaries (capital letters and strong punctuation).

We also found that supra-sentential units underpin writing production as run-ons are generated within the frame of event dynamics, thus echoing the narrative pattern, as mentioned above.

6.3. The Case of the Dyslexic Children: When Spelling Interferes with Sentence Processing

Our results confirm the writing difficulties of dyslexic children (Berninger et al. 2008). At the level of process, dyslexic children show a more pronounced inclination to pause outside of any context that usually favors a production break (typically found between words or before revisions). Dyslexic children have a slower pace of typing, because of longer disfluencies; interestingly, we showed that their mechanical, fluent writing competency is on par with that of typical children. This latter observation partly challenges the still-current hypothesis of insufficient automaticity and motor control due to cerebellar deficiency in dyslexic children (Nicolson and Fawcett 1990) and should be subjected to targeted studies.

At the linguistic level, the bursts produced by the dyslexics are shorter than those produced by typical children and do not allow us to identify the configuration of sentences as an emanation of the first functional processing step identified by language processing models and confirmed for typical children. Connected to pause length and place, this observation might hint at difficulties in the conceptualization processes. This is in line with the conclusions of Chanquoy et al. (1990) concerning writing behavior and syntactic production in children and adults.

At the product level, the texts produced by dyslexic children are short, with little narrative or syntactic development, which confirms Morken and Helland's (2013) findings. As might have been expected, some of the texts they produce contain a large number of dysgraphic errors, some of which prevent words from being recognizable. These results show that phonological (Bruck 1992; Snowling 1995) and spelling (Plisson et al. 2013; Sumner et al. 2016) deficits specific to dyslexics are concomitant with deficits in syntactic production. These results might support the process-disruption hypothesis, according to which spelling directly interferes with written composition: while recent studies have demonstrated that this hypothesis is inoperable in typical children (Rønneberg et al. 2022), it could be confirmed in dyslexics. Meanwhile, our data clearly demonstrate that syntactic production is affected in dyslexic children and that spelling is not the only area of difficulty, which is in line with the observations of Bishop and Snowling (2004) and Krasowicz-Kupis et al. (2009). At this level of phonological difficulties, the effect of syntactic awareness can be inoperative. These results are in line with those obtained, also for French, by Van Reybroeck (2020), who reported specific difficulties in terms of syntactic awareness and grammatical spelling in dyslexic children.

6.4. Challenges and Limitations

Our study faced some challenges and limitations.

First, we observed substantial interpersonal and inter-group differences at the level of sentence generation, in line with Arfé and Pizzocaro's (2016) results. We drew on the statistical Gaussian mixture model to standardize the results. However, some features can be improved in further studies. The study conditions at the hospital constrained data collection parameters for the group of dyslexic children. In order to deepen our knowledge of the impact of spelling on sentence generation in this specific group of the population, we need to acquire more complex and balanced data. For all children, the individual differences in holistic and compositional language processing (McConnell 2023) may influence sentence composition and segmentation in reading and writing.

Another limitation concerns the potential impact of the typing environment on sentence marking is as follows: without a specific test, it is difficult to know whether unmarkedness is due to the influence of the keyboard, linked to the need in French to use two keys to produce the period symbol, or to a "natural" production process that reflects oral practices, meaning that marking sentences is not systematically required for oral language use (Blanche-Benveniste et al. 1990; Brazil 1995; Berrendonner 2004; O'Grady 2010).

A fine-grained sentence analysis combined with a revision analysis would further support the syntax awareness hypothesis (Bowey 1986; Cain 2007; Mokhtari and Thompson 2006). A detailed study into the processing of verbal tenses, temporal markers, and character introduction and presentation would also give us a better grasp of the impact of supra-sentential units specific to the discourse genre on language processing.

Conversely, a detailed description of bursts' linguistic structure will help to better understand the transition from bursts to sentences and then to text. We must consider that, due to revisions, the sentence is not specifically the result of burst concatenation, while both production and revision bursts overlap with language chunks in about 75% of cases, according to recent findings (Cislaru et al. 2023).

Concerning writing in dyslexics, eye-tracking approaches—which are rather difficult to implement in keystroke-logging studies due to the instability of the gaze between keyboard and screen—might provide support for a multiple-cause hypothesis. Indeed,

recent imaging studies have clearly shown that phonological alterations in dyslexic children are associated with abnormalities in cortical structures, and several studies have reported deficits in terms of hearing (Tallal 1980), visual perception, and eye movement performance (Stein et al. 1988).

Finally, given the length and overall quality of the texts, the question of motivation to write (cf. Rasteiro and Limpo 2023) seems to be an issue to cope with for all dyslexic children, as well as for some typically developing children.

7. Conclusions

Starting from the premise that the sentence is a grammatical artifact of literacy, and building on Sinclair and Mauranen's hypothesis that language processing relies on segmenting the information flow into smaller units, we investigated the way in which children aged 9–11 handle sentences in terms of language processing. To investigate the processual units, we used keylog recordings of a writing task, which allowed us to segment the textual process into bursts of production. We found that not all children produced sentences, even though they all produced a text, and that their degree of conformity with sentence-related conventions did not correlate with their ease of writing, as proxied by text length. We furthermore showed that sentence boundaries do attract pauses, yet most pauses are located before a strong punctuation mark, at an infra-sentential level, mainly being located between words or before revision events. The study of such revised sequences also showed that the units that can be swapped for each other are mostly located at a lower syntactic level than that of the sentence.

We also considered dyslexic children in relation to hypothesis that dyslexia can be considered a language processing impairment that potentially impacts sentence processing. It was found that the dyslexic children's writing process is more fragmented than that of the typical children, yet they do not display any significant difference with respect to their handling of proto-sentences. Instead, they show differences in terms of the degree of elaboration and explicit marking. Language processing seems to be impaired at a lower level, which is consistent with our hypothesis that the processual units are best defined at a level below that of the sentence. This might also be a trace of conceptualization difficulties.

Summing up, it appears that processing units of language cannot be equated with sentence processing in writing. The information flow is usually produced through smaller bursts that each carry a piece of meaning or correspond to a specific operation in text crafting and revision. The way these pieces are joined together depends on the children's writing skills, writing strategies, and on the artefactual conventions that they are taught and come to adopt.

Author Contributions: Conceptualization, G.C. and M.P.B.; methodology, G.C., Q.F. and M.P.B.; selection/screening of subjects, E.K. and R.D.; formal analysis, Q.F., M.P.B. and G.C.; investigation, G.C. and M.P.B.; resources, M.P.B. and G.C.; data curation, M.P.B., G.C., Q.F.; writing—original draft preparation, G.C., Q.F. and M.P.B.; review draft, E.K. and R.D.; supervision, G.C. and M.P.B.; project administration, G.C. and M.P.B.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) (Comité de Protection des Personnes CPP, Nord Ouest III, n° 2021-A00489-32, 11 May 2021).

Informed Consent Statement: Written informed consent was obtained from all children involved in the study and the children's parents after the experimental procedure had been explained to them.

Data Availability Statement: Data available on demand.

Conflicts of Interest: The authors declare no conflicts of interest.

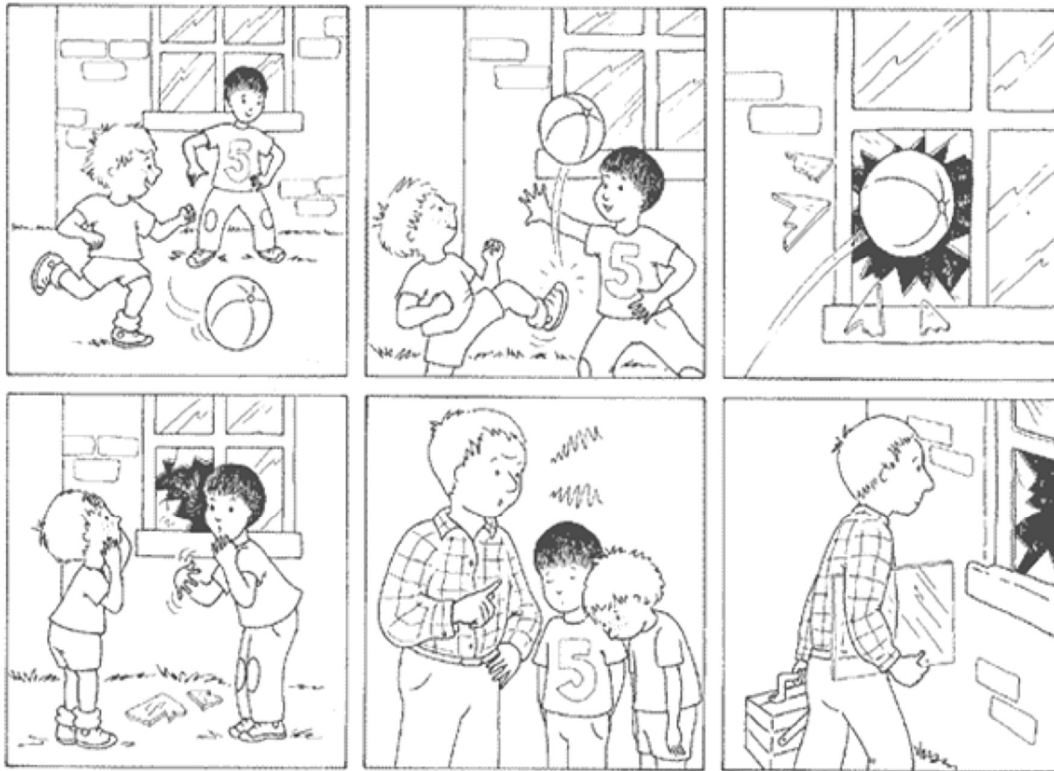## Appendix A. Material for Writing (Children with Dyslexia)

Figure A1. Picture sequence n°11, lutinbazar.fr.

Appendix B. Examples of Texts Produced by Typical Children

(1) *c'est etait une bagarre avec moi Baséhou contre Noah et Thibaut ca commencer par Noah il ma taper j'ai rendu aprés Baséhou à dit je vien t'aider Thibaut à dit moi aussi Baséhou à mit un coup Thibaut à dit moi aussi j'ai donné une claque Noah a conier Baséhou au mur j'ai poussé Noah vers l'herbe il est tomber sur l'herbe Noah c'est lever je les repouser il c'est relever il ma mit un coup de pied Thibaut a taper Baséhou la pousser contre le mur lui a taper il sont partie* [P16CM2N1, 10 yo, M]

(2) *Il etais une fois un garcon qui se battais toot le temps il cherchais toujours a se batre il frappais il insulter les mere et il ettais toujour punis il avais un bon fon mais il ne pouvez pas en nanpecher sa mamam le punisais de tout mais sa ne changer rien unjour il rencontra son pere et le pere lui dit si tu fait encore ca je vais t envoyer au maroc dans un ecole ou il auron le droit de te fraper ou sinon tu finis ton anne trankil et tu auras tous ce que tu voudras le petit a dis ca ne ce fais pas d acheter le gent le petit dit je te jure papa c eu qui me cherche apres il vont dire au prof que sait moi uje prefere aller au maroc et etre avec ma famille et me fer fraper au lieu de rester seul ma famille me manque papa tu peut m emener au maroc papa sil te plait si tu finis l anne bien oui et ensuite le garcon ne se battais plus il avais des bonne note et quand il termina l anne il alla au maroc et il devient ingenieur chez apple et il devin richce* [P03CM2N1, 10 yo, M]

Appendix C. Examples of Corrected Versions of the Texts—Corrected Versions of Examples (1) and (2)

(1b) C'était une bagarre avec moi, Baséhou contre Noah et Thibaut. Ça commençait par Noah. Il m'a tapé, j'ai rendu. Après Baséhou a dit: « je viens t'aider », Thibaut a dit: « moi aussi ». Baséhou a mis un coup. Thibaut a dit: « moi aussi ». J'ai donné une claque. Noah a cogné Baséhou au mur. J'ai poussé Noah vers l'herbe, il est tombé sur l'herbe. Noah s'est levé, je l'ai repoussé. Il s'est relevé, il m'a mis un coup de pied. Thibault a tapé Baséhou, l'a poussé contre le mur, lui, a tapé. Ils sont partis. [P16CM2N1]

(2b) Il était une fois un garçon qui se battait tout le temps. Il cherchait toujours à se battre. Il frappait, il insultait les mères, et il était toujours puni. Il avait un bon fond, mais il ne pouvait pas en n'empêcher. Sa maman le punissait de tout, mais ça ne changeait rien. Un jour, il rencontra son père, et le père lui dit: "si tu fais encore ça, je vais t'envoyer au Maroc dans une école où ils auront le droit de te frapper". "Ou sinon tu finis ton année tranquille, et tu auras tout ce que tu voudras". Le petit a dit: "ça ne se fait pas d'acheter les gens". "Le petit dit: "je te jure papa, c'est eux qui me cherchent, après ils vont dire au prof que c'est moi. Je préfère aller au Maroc et être avec ma famille, et me faire frapper au lieu de rester seul. Ma famille me manque. Papa, tu peux m'emmener au Maroc, Papa s'il te plaît". "Si tu finis l'année bien, oui". Et ensuite, le garçon ne se battait plus. Il avait de bonnes notes. Et quand il termina l'année, il alla au Maroc et il devint ingénieur chez Apple et il devint riche. [P03CM2N1]

References

Almond, Russell, Paul Deane, Thomas Quinlan, Michael Wagner, and Tetyana Sydorenko. 2012. A preliminary analysis of keystroke log data from a timed writing task. *ETS Research Report Series* 2: i-61. [CrossRef]

Arfé, Barbara, and Eleonora Pizzocaro. 2016. Sentence Generation in Children with and Without Problems of Written Expression. In *Written and Spoken Language Development across the Lifespan. Literacy Studies*. Edited by Joan Perera, Melina Aparici, Elisa Rosado and Naymé Salas. Berlin and Heidelberg: Springer, vol. 11, pp. 327–44. [CrossRef]

Baaijen, Veerle M., David Galbraith, and Kees de Glopper. 2012. Keystroke analysis: Reflections on procedures and measures. *Written Communication* 29: 246–77. [CrossRef]

Béguelin, Marie-José, Mathieu Avanzi, and Gilles Corminboeuf. 2011. *La Parataxe*. Lausanne: Peter Lang Verlag.

Benzitoun, Christophe, Anne Dister, Kim Gerdes, Sylvain Kahane, Paola Pietrandrea, Frédéric Sabio, and Jeanne-Marie Debaisieux. 2010. Tu veux couper là faut dire pourquoi—Propositions pour une segmentation syntaxique du français parlé. Paper presented at the Congrès Mondial de Linguistique Française, New Orleans, LA, USA, July 12–15, pp. 2075–90.

Berninger, Virginia W., Kathleen H. Nielsen, Robeftr D. Abbott, Ellen Wijsman, and Wendy Raskind. 2008. Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology* 46: 1–21. [CrossRef] [PubMed]

Berninger, Virginia W., William Nagy, and Scott Beers. 2011. Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. *Reading & Writing* 24: 151–82. [CrossRef]

Berrendonner, Alain. 2004. Grammaire de l'écrit vs grammaire de l'oral: Le jeu des composantes micro- et macro-syntaxiques. In *Interactions Orales en Contexte Didactique*. Edited by Alain Rabatel. Lyon: PUL, pp. 249–62.

Bishop, Dorothy V. M., and Margaret J. Snowling. 2004. Developmental dyslexia and specific language impairment: Same or different? *Psychol Bull.* 130: 858–86. [CrossRef] [PubMed]

Blanche-Benveniste, Claire, Mireille Bilger, Christine Rouget, and Karel van den Eynde. 1990. (with the participation of Piet Mertens). *Le français parlé. Etudes Grammaticales*. Paris: CNRS Éditions.

Bock, Kathryn, and Willem Levelt. 1994. Language production. Grammatical encoding. In *Handbook of Psychology*. Edited by Morton Ann Gernsbacher. San Diego: Academic Press, pp. 945–84.

Bowey, Judith A. 1986. Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology* 41: 282–99. [CrossRef]

Brazil, David. 1995. *A Grammar of Speech*. Oxford: Oxford University Press.

Bressoux, Pascal, Bernard Slusarczyk, Ludovic Ferrand, and Michel Fayol. 2023. Is spelling related to written composition? A longitudinal study in French. *Reading & Writing* 37: 615–39. [CrossRef]

Bruck, Maggie. 1992. Persistence of dyslexics phonological awareness deficits. *Developmental Psychology* 28: 874–86. [CrossRef]

Cain, Kate. 2007. Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics* 28: 679–94. [CrossRef]

Campione, Eselle, and Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. Paper presented at the Actes de Speech Prosody 2002, Aix-en-Provence, France, April 11–13, pp. 199–202.

Chafe, Wallace. 1992. Information flow in speaking and writing. In *The Linguistics of Literacy*. Edited by Pamela Downing, Susan D. Lima and Michael Noonan. Amsterdam and Philadelphia: John Benjamins, pp. 17–29.

Chanquoy, Lucile, Jean-Noël Foulin, and Michel Fayol. 1990. Temporal management of short text writing by children and adults. *Cahiers de Psychologie Cognitive* 10: 513–40.

Chenoweth, Norton Ann, and John R. Hayes. 2001. Fluency in Writing: Generating text in L1 and L2. *Written Communication* 18: 80–98. [CrossRef]

Chenu, Florence, Francois Pellegrino, Harriet Jisa, and Michel Fayol. 2014. Interword and intraword pause threshold in writing. *Frontiers in Psychology* 5: 189. [CrossRef]

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cislaru, Georgeta, ed. 2015. *Writing(s) at the Crossroads: The Process/Product Interface*. Amsterdam and Philadelphia: John Benjamins.

Cislaru, Georgeta, and Thierry Olive. 2017. Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction? *Corpus* 17: 1–21.

Cislaru, Georgeta, and Thierry Olive. 2018. *Le Processus de Textualisation*. Bruxelles: De Boeck.

Cislaru, Georgeta, and Thierry Olive. 2021. Que peut nous apprendre l'écriture enregistrée en temps réel au sujet des figures de construction? *L'Information Grammaticale* 169: 21–29.

Cislaru, Georgeta, Iris Eshkol-Taravella, and Sarah Almeida Barreto. 2023. Automatic processing of real-time recorded writing: Pausal segmentation versus chunking. In *RADH 2023—Recent Advances in Digital Humanities*. Timisoara: Universitatea de Vest din Timișoara.

Combettes, Bernard. 2011. Phrase et proposition: Histoire et évolution de deux notions grammaticales. *Le français aujourd'hui* 173: 11–20. [CrossRef]

Connelly, Vince, Julie E. Dockrell, Kirsty Walter, and Sarah Critten. 2012. Predicting the Quality of Composition and Written Language Bursts From Oral Language, Spelling, and Handwriting Skills in Children With and Without Specific Language Impairment. *Written Communication* 29: 278–302. [CrossRef]

Dockrell, Julie E., and Vince Connelly. 2016. The Relationships Between Oral and Written Sentence Generation in English Speaking Children: The Role of Language and Literacy Skills. In *Written and Spoken Language Development across the Lifespan. Literacy Studies*. Edited by Joan Perera, Melina Aparici, Elisa Rosado and Naymé Salas. Berlin and Heidelberg: Springer, vol. 11, pp. 161–77. [CrossRef]

Faigley, Lester, and Stephen Witte. 1981. Analyzing revision. *College Composition and Communication* 32: 400–14. [CrossRef]

Fayol, Michel. 1991. From Sentence Production to Text Production: Investigating Fundamental Processes. *European Journal of Psychology of Education* 6: 101–19. [CrossRef]

Fayol, Michel. 2013. L'enfant confronté à l'écrit. In *L'acquisition de l'écrit*. Edited by Michel Fayol. Paris: Presses Universitaires de France, pp. 33–51.

Feltgen, Quentin, Florence Lefeuvre, and Dominique Legallois. 2023. Sujet clitique et dynamique de l'écrit: Un éclairage par les jets textuels. *Discours*, 32. Available online: http://journals.openedition.org/discours/12509 (accessed on 10 December 2023).

Ferreiro, Emilia. 1978. What is Written in a Written Sentence? A Developmental Answer. *The Journal of Education* 160: 25–39. Available online: http://www.jstor.org/stable/42750832 (accessed on 31 October 2023).

Foulin, Jean-Noël. 1998. To what extent does pause location predict pause duration in adults' and children's writing? *Cahiers de Psychologie Cognitive* 17: 601–20.

Frank, Stefan L., Rens Bod, and Morten H. Christiansen. 2012. How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences* 279: 4522–31. [CrossRef]

Gardiner, Alan H. 1922. The Definition of the Word and the Sentence. *British Journal of Psychology* 12: 352–61.

Garrett, Merrill F. 1975. The analysis of sentence production. In *The Psychology of Learning and Motivation*. Edited by Gordon H. Bower. New York: Academic Press, vol. 9, pp. 133–77.

Garrett, Merrill F. 1980. Levels of processing in sentence production. In *Language Production, Volume I: Speech and Talk*. Edited by Brian L. Butterworth. London: Academic Press, pp. 177–220.

Gerlaud, Béatrice. 2016. La phrase: Entité insaisissable au lycée? *Lidil* 54: 151–66. [CrossRef]

Gilquin, Gaëtanelle. 2024. The Processing of Multiword Units by Learners of English: Evidence from Pause Placement in Writing Process Data. *Languages* 9: 51. [CrossRef]

Givón, Talmy. 2002. *Biolinguistics: The Santa Barbara Lectures*. Amsterdam and Philadelphia: John Benjamins.

Goody, Jack. 1987. *The Interface between the Written and the Oral*. Cambridge, MA: Cambridge University Press.

Hall, Sophie, Veerle M. Baaijen, and David Galbraith. 2022. Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing* 37: 329–57. [CrossRef]

Kail, Michèle. 2000. Perspectives sur l'acquisition du langage. In *L'acquisition du langage*. Edited by Michèle Kail and Michel Fayol. Paris: PUF, pp. 13–30.

Kaufer, David S., John R. Hayes, and Linda Flower. 1986. Composing Written Sentences. *Research in the Teaching of English* 20: 121–40. [CrossRef]

Krasowicz-Kupis, Grażyna, Aneta R. Borkowska, and Pietras Izabela. 2009. Rapid automatized naming, phonology and dyslexia in Polish children. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 15: CR460-9. [PubMed]

Lalain, Muriel, Luciana Mendoça-Alves, Robert Espesser, Alain Ghi, Céline De Looze, and César Reis. 2012. Lecture et prosodie chez l'enfant dyslexique, le cas des pauses. Paper presented at the TALN-RECITAL, Journées d'Études sur la Parole, Grenoble, France, June 4–8, pp. 41–48.

Le Goffic, Pierre. 2011. Phrase et intégration textuelle. *Langue Française* 170: 11–28. [CrossRef]

Leijten, Marielle, and Luuk Van Waes. 2013. Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication* 30: 358–92. [CrossRef]

Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.

Masseron, Caroline. 2019. De la textualité narrative aux faits syntaxiques dans un écrit scolaire. Peut-on articuler micro- et macro-syntaxe dans une perspective didactique? In *Types d'unités et procédures de segmentation*. Edited by Marie-José Béguelin, Gilles Corminboef and Florence Lefeuvre. Limoges: Lambert Lucas, pp. 45–63.

Matsuhashi, Ann. 1981. Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English* 15: 113–34. [CrossRef]

McConnell, Kyla. 2023. Individual Differences in Holistic and Compositional Language Processing. *Journal of Cognition* 6: 29. [CrossRef]

Medimorec, Srdan, and Evan F. Evan Risko. 2017. Pauses in written composition: On the importance of where writers pause. *Reading and Writing: An Interdisciplinary Journal* 30: 1267–85. [CrossRef]

Mitchell, Don C. 1994. Sentence parsing. In *Handbook of Psycholinguistics*. Edited by Morton Ann Gernsvacher. New York: Elsevier, pp. 375–409.

Mokhtari, Kouider, and Brian H. Thompson. 2006. How problems of reading fluency and comprehension are related to difficulties in syntactic awareness skills among fifth graders. *Reading Research and Instruction* 46: 73–94. [CrossRef]

Morken, Frøydis, and Turid Helland. 2013. Writing in dyslexia: Product and process. *Dyslexia* 19: 131–48. [CrossRef] [PubMed]

Nicolson, Roderick I., and Angela J. Fawcett. 1990. Automaticity: A new framework for dyslexia research? *Cognition* 35: 159–82. [CrossRef] [PubMed]

O'Grady, Gerard. 2010. *A Grammar of Spoken English Discourse: The Intonation of Increments*. London: Bloomsbury Publishing.

Olive, Thierry. 2014. Toward an Incremental and Cascading Model of Writing: A review of research on writing processes coordination. *Journal of Writing Research* 6: 173–94. [CrossRef]

Peterson, Robin L., and Bruce F. Pennington. 2012. Developmental dyslexia. *Lancet* 379: 1997–2007. [CrossRef] [PubMed]

Plisson, Anne, Daniel Daigle, and Isabelle Montesinos-Gelet. 2013. The spelling skills of French-speaking dyslexic children. *Dyslexia* 19: 76–91. [CrossRef] [PubMed]

Rasteiro, Isabel, and Teresa Limpo. 2023. Examining Longitudinal and Concurrent Links Between Writing Motivation and Writing Quality in Middle School. *Written Communication* 40: 30–58. [CrossRef]

Roeser, Jens, Sven De Maeyer, Mariëlle Leijten, and Luuk Van Waes. 2021. Modelling typing disfluencies as finite mixture process. *Reading and Writing* 37: 359–84. [CrossRef]

Rondelli, Fabienne. 2013. La phrase, segment textuel «de base»: Choix d'écriture d'élèves de cycle 3 et jugements des enseignants. *Le Français Aujourd'hui* 181: 71–81. [CrossRef]

Rønneberg, Vibeke, Mark Torrance, Per H. Uppstad, and Christer Johansson. 2022. The process-disruption hypothesis: How spelling and typing skill affects written composition process and product. *Psychological Research* 86: 2239–55. [CrossRef] [PubMed]

Rutter, Michael, Avshalom Caspi, David Fergusson, John L. Horwood, Robert Goodman, Barbara Maughan, Terrie E. Moffitt, Howard Meltzer, and Julia Carroll. 2004. Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *JAMA* 291: 2007–12. [CrossRef] [PubMed]

Ruwet, Nicolas. 1967. *Introduction à la grammaire générative*. Paris: Plon.

Schilperoord, Joost. 2002. On the Cognitive Status of Pauses in Discourse Production. In *Contemporary Tools and Techniques for Studying Writing*. Edited by Thierry Olive and C. Michael Levy. Dordrecht: Kluwer Academic Press, pp. 61–87.

Schneuwly, Bernard. 1988. *Le langage écrit chez l'enfant*. Neuchâtel and Paris: Delachaux et Niestlé.

Seguin, Jean-Pierre. 1993. *L'Invention de la phrase au XVIIIe siècle, contribution à l'histoire du sentiment linguistique français*. Leuven: Peeters.

Sinclair, John, and Anna Mauranen. 2006. *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam and Philadelphia: John Benjamins.

Sinclair, John M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Snell, Joshua, and Jonathan Grainger. 2017. The sentence superiority effect revisited. *Cognition* 168: 217–21. [CrossRef] [PubMed]

Snowling, Margaret J. 1995. Phonological processing and developmental dyslexia. *Journal of Research in Reading* 18: 132–38. [CrossRef]

Snowling, Margaret J. 2001. From language to reading and dyslexia. *Dyslexia* 7: 37–46. [CrossRef] [PubMed]

Snowling, Margaret J., Charles Hulme, and Kate Nation. 2020. Defining and understanding dyslexia: Past, present and future. *Oxford Review of Education* 46: 501–13. [CrossRef] [PubMed]

Spelman Miller, Krystian. 2006. The pausological study of written language production. In *Computer Keystroke Logging: Methods and Applications*. Edited by Kirk P. H. Sullivan and Eva Lindgren. Amsterdam: Elsevier, pp. 11–31.

Sprenger-Charolles, Liliane, and Willy Serniclaes. 2003. Acquisition de la lecture et de l'écriture et dyslexie: Revue de la littérature. *Revue française de linguistique appliquée* VIII: 63–90. [CrossRef]

Stein, John F., Patricia M. Riddell, and Sue Fowler. 1988. Disordered vergence control in dyslexic children. *British Journal of Ophthalmology* 72: 162–66. [CrossRef]

Sumner, Emma, Vincent Connely, and Anna L. Barnett. 2013. Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution. *Reading & Writing* 26: 991–1008. [CrossRef]

Sumner, Emma, Vincent Connelly, and Anna L. Barnett. 2016. The influence of spelling ability on vocabular choices when writing for children with dyslexia. *Journal of Learning Disabilities* 49: 293–304. [CrossRef]

Tallal, Paula. 1980. Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang* 9: 182–98. [CrossRef] [PubMed]

Tolchinsky, Liliana, and Anna Teberosky. 1998. The development of word segmentation and writing in two scripts. *Cognitive Development* 13: 1–24. [CrossRef]

van Hell, Janet, Ludo Verhoeven, and Liesbeth van Beijsterveldt. 2008. Pause time patterns in writing narrative and expository texts by children and adults. *Discourse Processes* 45: 406–27. [CrossRef]

Van Raemdonck, Dan. 2013. Aussitôt la détermination effectuée, et toute prédication dehors, on dit le syntagme clôturé. Etude de structures entre syntagme et sous-phrase. In *Les Fonctions Grammaticales. Histoire, Théorie, Pratiques*. Edited by Aboubakar Ouattara. Bruxelles: P.I.E. Peter Lang, pp. 209–21.

Van Raemdonck, Dan, Marie Détaille, and Lionel Meinertzhagen. 2011. *Le sens grammatical. Référentiel à destination des enseignants*. Bruxelles: Peter Lang.

Van Reybroeck, Marie. 2020. Grammatical spelling and written syntactic awareness in children with and without dyslexia. *Frontiers in Psychology* 11: 1524. [CrossRef] [PubMed]

Van Waes, Luuk, Marielle Leijten, Jens Roeser, Thierry Olive, and Joachim Grabowski. 2021. Measuring and assessing typing skills in writing research. *Journal of Writing Research* 13: 107–53. [CrossRef]

Wen, Yun, Joshua Snell, and Jonathan Jonathan Grainger. 2019. Parallel, cascaded, interactive processing of words during sentence reading. *Cognition* 189: 221–26. [CrossRef]

Zawadowski, Leon. 1971. Sentence, its grammatical definition. *Linguistics* 9: 95–112. [CrossRef]