Introduction to Special issue

Operating Units in Written Language Performance: Linguistic and Behavioral Perspectives

Georgeta Cislaru[1], Claire Doquet[2] & Thierry Olive[3]

[1] Université Sorbonne Nouvelle | France
[2] Université Bordeaux | France
[3] Universié de Poitiers | France

Abstract: This special issue of *Journal of Writing Research* addresses the fundamental question of performance units in writing: how can we characterize these units, and which theoretical paradigms allow us to describe them? Despite their core role in the writing activity, there is no consensus on the nature of written performance units. In order to progress on this issue, the different articles presented in this special issue shed light on performance units, their description, definition and role in text construction. Different methodological and theoretical approaches, based on behavioral data, with pauses as a central indicator, illustrate how linguistic structures produced in these units constrain written production.

Keywords: written performance, linguistic markers, behavioral markers, pausing behavior

Although Foulin (1995) noted about 30 years ago the pre-theoretical nature of studies on pauses, he also underlined that the study of pauses could be a heuristic way of exploring the cognitive system of written production, since pauses can be traces of cognitive processes that are not directly accessible. This position is in line with the postulate that pauses "offer observable clues to the covert cognition processes which contribute to discourse production" (Matsuhashi 1981, p. 114). In addition, pauses are directly linked to observable linguistic data, which they segment into sequences of varying length, content and format, going from "So…" to "They decided to…" and even "…legislation? If the legislation is put in place under strict rules and regulations, then it won't…". The presence of these fluently produced language sequences shows that language performance is sequential in its release. In this perspective, the study of written performance units would mean matching language units to cognitive units, in order to better understand both language processing and the writing activity. Thus, since they are heuristically identified by observing language performance, fluently produced linguistic sequences can be considered as performance units.

Addressing performance units in writing means questioning how writing develops through time, in other words how writers progressively construct their texts by adding new content and by revising the text they have already produced. Such a perspective not only involves focusing on the process of writing rather than on the final product, it also requires examining the different textual operations such as deletions, substitutions, insertions and displacements carried out to assemble and revise the segments produced to eventually form the final text. The aim is then to reconstruct the writing process by placing the writing operations in chronological order, considering pauses, execution periods and the language sequences produced. This perspective introduces temporality into the study of writing and attempts to recover the dynamic dimension of discourses, allowing for a new understanding of written production.

Time, chronology, continuous discourse and its interruptions have long been studied in speech (Goldman-Eisler, 1968; Fox Tree, 2001). Such approaches are adapted to the materiality of spoken discourse, characterized by the coincidence between the production of the discourse and its final form. Oral communication cannot rely on drafts: temporality constrains and structures the discourse as it unfolds. Looking into these dynamics is therefore central for speech analysis, since uttering discourses includes stammering and hesitations. Thus, silence and pauses are *ex officio* inherent to enunciation.

By contrast, for its reception, writing-as-a-product is presented as complete and finished. It conceals the revisions, corrections, and restructurings that marked its genesis. This explains why many studies on writing initially considered it primarily as a product of planning (Ochs, 1979; Tannen, 1982), focusing on coherence markers. These studies, along with those focusing on the semiotic specificities of writing (Vachek, 1973), highlight the disjunction between production and product. The question of the elaboration of writing has been addressed since the 1990s in psychology (Foulin, 1995)

as well as in linguistics and textual genetics (Doquet, 1999). Much research is being done on the corresponding cognitive processes (see below). Such studies have mainly been enabled by the use of tools for recording writing and its chronology, in both its handwritten and its typed forms (Doquet, 2004; Leblay, 2007; Wengelin et *al.*, 2009; Leijten & van Waes, 2016; Chesnet et *al.*, 2018; Cislaru & Olive, 2018; Lindgren & Sullivan, 2019). Thanks to access to the temporal variables that are recorded, the lens shifted from the finished written text to the act of writing, from statement to enunciation, from describing static forms of writing to questioning units of written performance.

The aim of this special issue is to bring together recent work approaching the question of performance units in writing from both a linguistic and a behavioral perspective, in order to provide a state of the art and identify emerging research directions. Before presenting the articles in this issue, we offer a brief review of behavioral approaches that have adopted a processual point of view on the dynamics of text composition based on the analysis of pauses, and trace the linguistic tradition that has questioned the notion of language performance units in speech and writing.

The dynamics of text composition

Understanding how writing develops from a process perspective requires first identifying the cognitive processes necessary to compose a text, which are invisible to the observer. Four main cognitive components (or writing processes) have been identified (Hayes, 2012; Olive, 2014). One component is related to conceptual preparation (or text planning) and involves retrieving knowledge from the writer's long-term memory, seeking information from external sources (books, on the internet, etc.), and organizing and hierarchizing this information. A second component, translating, intervenes at the linguistic level to transform the conceptual message into a verbal one by defining the orthographic, morphological and syntactic properties of the linguistic units being processed. The control or revision component evaluates the output of the preceding processes. Revision can be mental, focusing on parts of the message yet to be transcribed or it can involve editing text that has already been written. The fourth component, called execution, ensures the transcription of the text. Different models of execution have been proposed based on the execution modality, such as handwriting (van Galen, 1991) or typing (Logan & Crump, 2011).

Another goal of cognitive research is to understand the time course of these processes, i.e., how these processes unfold during a writing session. There is now ample evidence that planning predominantly occurs in the first half of a composition session, revision processes occur mainly in the latter half and translating or formulation processes occur almost continuously throughout (for a review, see Olive 2004). How these processes are coordinated is also crucial. Although producing a written segment of text seems to intuitively start with conceptualizing the message to be written, followed by its formulation and then by its transcription (and possibly with revision), these processes are not triggered in a strictly linear way. Instead, their implementation is recursive, each of

the writing processes being able to interrupt another (Flower & Hayes, 1981; Ransdell & Levy, 1994). Additionally, the writing processes can occur simultaneously with information cascading between levels of processing: although each segment of text is prepared in a linear way from planning to execution, as soon as a level of processing has sent the results of its operations to the next process, it begins preparing the next segment of text, whether it adds new content or modifies the text already written (Olive, 2014). For example, adult writers can plan their text or revise it at the same time as they are transcribing a segment of text (Alamargot et al., 2007; Wengelin et al., 2024). Similarly, they can plan the end of a sentence during the writing of its beginning, while novice writers plan each part of a sentence sequentially, with long planning pauses between segments (Chanquoy et al., 1990). In fact, sentence preparation can be conceived as incremental: sentences are created as smaller units with the complete sentence not being fully planned at the onset of execution. However, when sentence complexity increases, writers can prepare longer units before they begin to write (e.g., Damian & Stadthagen-Gonzalez, 2009; Nottbusch, 2010; Roeser et al., 2019). More specifically, some lexical retrieval and grammatical encoding of the beginning of a sentence are prepared before their execution while the remaining parts of the sentence can be prepared concurrently to transcription, mostly due to projection principles (Auer, 2005) and semantic priming (Foss, 1982), which foster language choices. Such parallel and advance planning has also been demonstrated at the lexical level during the transcription of words (Maggio et al., 2011) as well as at the syllable level when producing words (Sausset et al., 2013). All these findings suggest that writers can rely both on parallel and sequential planning of the text. It also indicates that writers may process different types of units to adapt to the demands that text composition places on working memory (Olive, 2014).

While research on cognitive processes has provided valuable insights into their nature and their temporal organization and coordination, a critical question remains about the linguistic and/or textual units on which these processes operate. One approach to address this question has been to analyze writing pauses, drawing on studies of oral fluency and disfluency, which interpret hesitations in spontaneous speech as decision-making moments (see above).

Segmenting writing with pauses

In writing, pauses serve several functions, and writing research has focused on those that allow cognitive writing processes to take place. Such pauses serve at least two primary functions (Schilperoord, 2002). First, they allow for the preparation of the next segment of text, whether at the conceptual or linguistic level. Second, they enable the revision of already-produced text. Notably, pauses preceding revisions account for approximately 45–50% of the total pause time and tend to be longer when the revised segments are more extensive or involve higher-level processing (van Waes & Schellens, 2003). The duration of pauses also varies depending on the type of revision: for instance, semantic

revisions are typically preceded by longer pauses than spelling corrections (Matsuhashi, 1987).

Studies have further revealed that conceptualization, linguistic formulation, and revision processes can occur with similar probability during pauses (Olive et al., 2009). This suggests that pauses are moments of significant cognitive activity, involving a range of processes that support both the production and the revision of text.

The study of pauses occupies an important place in the search for units of writing, since by segmenting texts spontaneously, pauses provide insights into the nature of the units that writers process. For instance, if pauses are privileged moments of planning, they should occur more frequently at the boundaries of the "units" of language planning. In speaking, Butterworth (1975) showed that discourse is structured in cycles alternating fluent phases and hesitations, organized around semantic configurations. During fluent phases of production, characterized with few and short pauses, speakers are believed to plan their discourse essentially at the morphosyntactic level, while during the disfluent phases they focus on the conceptual aspects of their discourse. Furthermore, since fluent phases often involve the production of multiple sentences, conceptual planning is thought to expand several statements.

Other researchers have emphasized the function of pauses in syntactic planning. Pauses are not only more frequent at the boundaries of syntactic units but are also longer depending on the structure they precede. Specifically, pauses are longer at higher hierarchical levels (Medimorec & Risko, 2017) and, within the same hierarchical level, their duration varies according to the nature of the clause that follows (van Hell et al., 2008). This suggests that the duration of writing pauses may therefore increase with the syntactic complexity of the language units that follow them (see above or Ailhaud et al., 2016). However, while sentence boundaries seem to generally attract pauses, other levels such as connectors, strong punctuation marks or sentence subjects are marked by comparable pauses, depending on the writer and the text produced (Cislaru et al., 2024). Pauses also open windows onto the macro- and microstructural planning of discourse. For instance, pauses between paragraphs are suggested to be related to knowledge management, while those between sentences would mainly stem from syntactic and lexical processes, which often interact with conceptual planning. Pauses between clauses would mostly reflect linguistic planning, specifically the syntactic organization of discourse, while within-clause pauses would primarily be influenced by the predictability of lexical items. In short, research strongly supports the idea that syntactic boundaries impact pauses. These are longer as the hierarchical level of the unit to be produced increases (phrase, clause, sentence, paragraph) and as discourse complexity increases.

As indicated above, since central cognitive operations such as planning, or formulation, can occur during transcription, studying the segments of texts that are written during execution periods between two consecutive pauses can also inform on the units of processing. These sequences, called bursts of writing (Kaufer et al. 1986), have been described in terms of writing behavior and fluency.

Two main types of bursts have been studied in detail: production bursts and revision bursts (e.g., Chenoweth & Hayes, 2003; Baaijen et al., 2012). The former increment the text on its edge, the latter intervene on the text already produced. Revision bursts are generally fewer and shorter than production bursts. Their writing speed is also faster, and they are preceded by longer pauses, suggesting greater mental preparation than production bursts. As with production bursts, the unit of a revision burst also requires further investigation to uncover the different spans of revision.

In addition, it has been shown that burst length, and therefore the text segments that are produced or revised during these bursts, varies according to the writers' skills. For instance, bursts are shorter when transcription is not well mastered. This is the case in handwriting in children and adults as well as in typing. Consequently, training transcription in children leads to longer bursts (Alves et al., 2016; Limpo & Alves, 2017). Second, linguistic formulation skills also influence bursts. Thus, skilled writers write with longer bursts than novice writers (Friedlander, 1989; Kaufer et al., 1986). Similarly, when students write a text in their mother tongue, bursts contain more words than when the same students compose in their second language (Chenoweth & Hayes, 2001, 2003; Barkaoui, 2019; Olive et al., submitted). Children with a specific language disorder also produce with shorter transcription periods than children with typical language development (Connelly et al., 2012).

Pauses in writing therefore not only segment the process of writing and the writers' mental operations: they create bursts of language and, therefore, allow researchers to identify text segments that may represent units of written performance. Further research is however needed to better understand the nature of bursts of writing, and particularly how these bursts change with different pause thresholds. One possible approach to enhance the alignment between behavioral cues and cognitive processes would be to divide up keystroke log information into activities corresponding to the components of global writing models and to identify sets of measures to be aggregated across the text as a whole (Galbraith & Baaijen, 2019: 322).

Performance units as the fundamental units of writing

Numerous studies have thus attempted to identify the fundamental unit of language, whether in the reception or production of oral discourse (most often) or in the production of written discourse (more rarely). Some initial research attempted to describe the language units manipulated by the writers to produce texts (Kaufer et al., 1986). However, the alignment between linguistic aspects and behavioral data remains to be developed. Previous findings in speech (Gee & Grosjean, 1983) and in writing (Doquet, 2011; Cislaru & Olive, 2018) have shown that these units are not fully accountable for by traditional syntactic theories and require specific linguistic and psycholinguistic models.

Competence vs performance: modelling units in speech

Although taking pauses into account in writing goes back as far as the middle of the last century (Van Bruggen 1946), the issue of encoding or decoding units precedes research on the writing process. It is also rooted in speech research, where it was discussed from various viewpoints: competence and/or performance, linear vs hierarchical models, syntax, prosody, semantics and their intertwining. In Goldman-Eisler's linear model, the planning unit is the word; in hierarchical domains, the planning unit is deemed syntactic (the proposition, i.e., the constant meaning of a clause or sentence, independent of the chosen form; Taylor, 1969). Studies such as Piolat's (1983) focus on pauses in spoken discourse and their role in the planning and coherent insertion of meaning. For Piolat (1983: 378), "the pause bears witness to the cognitive processes of 'selection', 'decision' and 'encoding' of the semantic, syntactic and lexical aspects that precede and accompany verbal emission."

Selkirk's (1984) Sense Unit Condition is a competence-based theory operating with the notion of intonational phrase, which must form a semantic unit. The intonational phrase is restructured in relation with the length of the constituents, flow pace, style, and its borders may not coincide with the constituents' frontiers. In a similar vein, Chafe (1994) relates the linguistic expression of immediately online available information to intonation units. The anti-attachment hypothesis (Watson & Gibson, 2005) is a performance-based theory which shows that "intonational boundaries are partly a product of planning and recovery processes by the sentence production mechanism and are therefore likely to occur before and after large constituents" (p. 280). Biber et al. (1999, pp. 1069–1070) identify the C-unit, which may be a clausal unit (close to a T-unit in its definition, see below) or a non-clausal unit, that cannot be analyzed as part of a clausal unit, but represents more than a third of the units in conversations.

Grosjean and colleagues were among the pioneers in this field, and their thoughts on performance units are well worth a look. Grosjean et al. (1979) showed that the performance structures identified by the location of pauses are invariant from one task to another (normal reading and reading without catching one's breath, at different speeds, and segmentation task). Later on, Grosjean et al. (1981) showed that these structures are not language- or modality-specific (e.g., American Sign Language). In a search for the "temporal organization of sentences", Gee and Grosjean (1983) matched a psychological and a (mainly prosodic) linguistic point of view to identify units called "phonological phrases", or "ɸ-phrases." These units are prosodically defined, and their size is between the word and the syntactic phrase. Grosjean and Dommergues (1983) focused on performance structures and their psychological reality. They argued that three main properties of performance units can be distinguished: structures composed of basic units of roughly equivalent length, hierarchical structures, and structures that present a certain degree of symmetry. The principle of hierarchy remains open to discussion; however, it is formulated as follows:

> "The second property of performance structures is that they are hierarchically organized. This is because the basic units themselves group together into larger units, which in turn group together to form still larger units. Thus, […] two higher-level groups then join to form the full structure." (p. 520)

Thus, speakers segment sentences and texts into units which, when articulated to each other, form the sentence or text in question. The same should be true in writing, with the main difference being the need to plan and choose appropriate sequences. Drawing on Gee and Grosjean's (1983) article, Abney (1991) introduced the concept of *chunk*, defined as "a single content word surrounded by a constellation of function words, matching a fixed template"; they are identical to "prosodic patterns", pauses being "most likely to fall between chunks" (p. 257). Viewed from a different angle, however, the content word can echo the informational dimension, while the prosodic role in segmentation is in line with Chafe's hypotheses mentioned above.

Sinclair and Mauranen (2006) developed a theoretical model for conceptualizing grammar, Linear Unit Grammar, at the junction of speech and writing. They also underlined the open dimension of the segmented units:

> "Structures thus have completion points, which need not put an end to the hearer's task of processing a given string until it's finished" (p. 33).

Modeling units in writing

In writing research, the question of the relevant encoding or decoding unit is often considered from a purely practical point of view, aimed at identifying the right unit or units for measuring the fluency or quality of the text produced, in relation to pausal segmentation. But planning units are not necessarily identical to production/segmentation units, and the identification of the latter is related to the research objective. Another approach is to look at the way in which a given predefined unit is produced (character, word, clause, sentence, paragraph...). In this perspective, prior to the study of the writing process as such, Hunt (1970, p. 4) defined the T-unit as "one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it". The length and complexity of these units are used to assess writing quality in L1 and L2, in children and adults (Scott, 1988, 2009).

While she laid the foundations for an approach to the writing process considering pausal segmentation, Matsuhashi (1981) was mostly interested in pause length and location – for instance, in observing pauses at the boundaries of T-units – rather than in the definition of what a written performance unit is. Nonetheless, her viewpoint is that of text analysis as she identified eight units of language: abstraction level, sentence roles, paragraph, initial modifying structure, lexical cohesion in the initial simple subject, lexical cohesion in the object of an adjective phrase, syntax, and content words. In addition to recognizing the writing process as a complex activity, this functional approach opens the way to a textually informed conception of written production units, which has unfortunately been too little exploited so far.

In one of the first studies that examined the segmentation of texts in bursts, Kaufer et al. (1986) sought to better understand how writers construct sentences when writing texts. Through four studies based on verbal protocols, they showed that more skilled writers produced longer bursts (11.3 words) than less skilled writers (7.2 words), and that writers required around 2.8 bursts to compose a sentence. From a qualitative point of view, the authors observed that bursts corresponded to 1/3 of syntactic proposition boundaries and 1/5 of sentence boundaries. They therefore hypothesized that writers retrieve ideas from long-term memory, grouped in a format that enables formulation processing to produce sentences.

In the frame of real-time writing, Spelman Miller (2002) focused on three measures to observe the kind of *units of production* that emerge: pause duration, pause frequency and the length of text produced between two pauses, called productivity by the author. Beyond the practical aim, the author was interested in the relevance/confirmation of specific grammatical units such as words, clauses, etc. What is interesting in Spelman Miller's model is the notion of *potential completion points* to define pause location, in order to "reflect the idea that the development of the text is not predetermined but open to adjustment in a number of different ways as the writer reworks the text" (2002: 259; see also Cislaru & Olive, 2018; Ulasik & Miletic, 2024). Based on a text linguistic frame, Spelman Miller (2002) categorized pauses according to their grammatical locations within the stretch of text. However, this approach is top-down and asymmetrical, the categories globally corresponding to traditional grammatical units, although enriched with discourse framing devices such as conjuncts and disjuncts and topic at a local level. Spellman Miller nevertheless shows the importance of clearly defining the units of analysis in the study of the writing process.

In an attempt to further comprehend the performance units of writing, Cislaru and Olive (2018) described the text segments contained in bursts of writing. They observed that a large majority of bursts were made up of incomplete or syntactically non-finite sequences, the right bound of which engages a syntagmatic scheme that has to be completed by the following text segment. They also described the formal, semantic and relational regularities of bursts and showed that written production is underpinned by constructional logics, in unsaturated preconstructed formats. This suggests that some text segments, or performance units, may be retrieved from the writer's long-term memory, while others may be fully generated (Cislaru & Olive, 2017). Cislaru and Olive (2018) also proposed to describe these bursts as units agglomerating elements around a junction attractor, which could implement a semantic relation at the textual level. The question of the nature of written performance units therefore remains open.

More recent work exploits advance in natural language processing (NLP) tools to annotate and thus convert parts of speech, chunks or sentences into process analysis units. Based on English and Dutch data, Leijten et al. (2019) offered a mainly methodological perspective showing that levels of language structure above the character can inform the analysis of the writing process. Cislaru et al. (2023) and Manseri et al.

(2024) proposed a chunk annotation and automatic analysis comparing burst and chunk boundaries to find a significant proportion of matches (between 43% and 75%) that are sensitive to the grammatical nature of the forms: thus, adjectives and nouns increase the chance of matching boundaries.

Beyond the lexeme and below the sentence, prefabricated sequences and constructions have also been a subject of interest in recent years (Cislaru & Olive, 2017, 2018; Gilquin, 2020, 2024). Their stabilized status in memory, due to entrenchment, makes these sequences prime candidates for one-shot written actualization, in a single burst of production. The results of all these studies are mitigated, with some sequences or contexts of use confirming the hypothesis of unitary actualization, while other occurrences are segmented by long pauses. These observations call, on the one hand, for further studies to test the hypothesis of a correspondence between prefabs and units of written performance and, on the other, for better consideration of the complexity of the writing process and the various parameters which may influence the nature and boundaries of these units.

While the levels of linguistic solidarity likely to underlie the configuration of written performance units are numerous and varied, another question emerges, that of the possibility of identifying performance units transcending these levels and constituting a homogeneous class despite linguistic heterogeneities.

Modeling (re)writing units

To account for this unique material, one also needs to consider the specific nature of written enunciation, in particular its spatiality, and what it aims for: a fixed object that will be presented as a whole and will elicit reading. This specificity has significant implications for the very conception of the written discourse, which is produced to be read rather than heard. There is one more fundamental distinction between spoken and written language, related to the disjunction between the temporality of the utterance and that of the communication of the text: the role of metalinguistic activity in writing. Writing time is not the same as reading time; writing is punctuated by moments of looking back at the text and rectifying it (see for instance Bowen & van Waes, 2020), which involve the metalinguistic activity of the writer.

When writers reread their own text, they do not engage in an interpretive collaboration (Eco, 1985) that characterizes discovery reading: when they reread, they already (globally) know what their text contains. They are therefore in immediate control of the text, rather than engaged in a project of collaborating with it to construct meaning. One rereads in order to assess the conformity of one's writing with one's own expectations, be they semantic, morphosyntactic or discursive. But why does this happen in writing rather than orally? Unlike speaking, which unfolds and disappears instantly (without any draft), writing is governed by stabilization, even during the writing process itself, because the letters and words are there, visible at any moment (and modifiable for the final text). Writing presents something to be seen (Doquet et al., 2022), and this

"presenting to be seen" is important from the perspective of the addressee, who, in reading, has access to a referent, but also sees signs and, inseparably, their graphic substance (see Olive & Passerault, 2012, who underline the visual spatial dimension in writing both from the product and the process perspectives). The fact that this substance is permanent, unlike in oral communication which is ephemeral, paves the way for metadiscursive activity by the writer-reader. Writing, and the (re)reading that accompanies it, involves a doubling of the reader, who is both engaged in understanding the text and in paying attention to the words. This metalinguistic activity inherent to writing is an important parameter for describing the activity as it may impact the writing events. Thus, a statement in progress will be studied differently depending on whether it is oral or written (Lemke, 2021).

Description of the special issue
In this context, this special issue proposes new reflections on the nature of performance units of writing. It does not offer an exhaustive description of these units. Likewise, it does not present a detailed modelling of the written performance units or of how texts are constructed. The articles that are collected aim to illustrate how different authors have addressed the question of written performance units in their research.

Olena Vasylets and Javier Marнn assert the need to define a meaningful relationship between the behavioral units (pauses, bursts and revisions) measured by keystroke logging and linguistic units. They conduct a scoping review to identify existing findings concerning the links between the two types of units, based on 9 studies in English, published between 2000-2023 and collected from major databases in linguistics, psychology and education. The findings concern different types of processing at linguistic levels (word groups, clauses, paragraphs...) and potentially different cognitive processing in L1 and L2 writing. A series of avenues are proposed to further develop research in this field. Beyond the generalizability and replicability of research on behavioral units and their matching with linguistic units, the authors mention the importance of multiplying the linguistic categories observed, taking into account various external and internal factors that can moderate writing behaviors and performance, such as writing task and topic, discourse types, writing conditions and medium (e.g., handwriting).

Another important question for a better understanding of written performance units concerns the typology of languages. Beyond an understanding of the linguistic specificities of languages, contrasting different languages makes it possible to question the universality of the units of written performance. In this context, Ilmari Ivaska, Outi Toropainen and Sinikka Lahtinen explore the differences between Finnish and Swedish, two languages that stem from different language families. They also question the impact of language proficiency by comparing first and second language writers. To do this, they analyze the distribution of pauses. Their results clearly show that the least skilled second language writers produce more pauses and in particular intra-word pauses, while the more skilled writers pause more often between sentences. For the authors, this suggests

different units of processing, with the less skilled writers focusing on lexical units while the more skilled writers deal with larger units. The authors also show that typological differences in the first language affect pause behavior more than the skills of the writers. At the same time, the authors problematize the choice of pause threshold selected for the analysis. They suggest that the choice of a pause threshold of 2 seconds probably does not allow for a more precise approach to lexical processing.

The article by Quentin Feltgen and Florence Lefeuvre seeks to highlight the way in which information is packaged into production bursts. The authors show that the writing process is determined or, at least, constrained by linguistic structure. In line with Linear Unit Grammar, the authors consider that cognitive pauses in writing are a close counterpart to oral disfluencies. They examine the written actualization of clitic subjects in French, which are atonal pronouns (*je, tu, il* – I, you, he) and therefore need to be associated with a conjugated verb. The written production of clitics appears to favor the position at the beginning of bursts; also, clitics are not subject to immediate revisions themselves, but attract revisions from the sequence following them, often produced after a deletion operation. These findings evidence that linguistic units are in turn characterized not only by their syntactic role (here, that of the Subject), but also by the way they allow language users to handle the information flow both through incremental utterances and repairs.

A large number of empirical studies have focused on the dynamics of sentence production. However, these studies have led to little modelling. In their paper, Malgorzata Anna Ulasik, Cerstin Mahlow and Michael Piotrowski present a model they developed to describe how sentences are gradually constructed from writing bursts and revision periods, what they call the text history. Their model is based on the premise that the final text is a linear sequence of sentences. They therefore try to model how writers gradually construct sentences from the mapping of three layers —a transformation layer, a sentence layer, and a burst layer— which relates pauses and bursts to syntactic structures affected by transforming sequences. Ulasik et al. present a set of tools to analyze sentence production cycles and align the different versions from a corpus of text keylogs. At the same time, they offer a visualization tool that reports on the different steps that lead to sentences. This implemented model thus makes it possible to automatically process large corpora of keylogs.

In his article, Jason Wirtz discusses creative writing, a type of writing that is rarely studied from an empirical point of view. The participants in the study, all enrolled in a Master of Fine Arts program, and who had shown experience in creative writing, wrote texts following the method of free writing. This study thus raises the question of the segmentation of texts based on a variable other than pauses, in this case, the moments of return to the text. It should also be noted that this study combines eye movement analysis with a qualitative approach. By crossing these data, Wirtz shows how both implicit and explicit cognition participate in the regulation of activity by participating in the decisions

to re-read the text and by creating the conditions for dispositionally guided text production.

In the frame of textual genetics, Rudolf Mahrer and Giovanni Zuccarino analyze a typewritten text by Jacques Derrida and challenge the status of writing pauses as the primary indicator for delineating text production units. Their approach thus diverges from most contributions in this issue, minimizing the notion of "textual flow" as the primary production unit. The authors aim to distinguish textual production units and characterize them, not based on their linguistic constituency, but as genetic units, i.e., according to how they articulate and, for some, burst forth in the constitution of the text. They consider the corpus from the perspective of editing, understood as "the mental process of shaping the form of a text sequence to be emitted". The authors propose a typology of writers based on their manner of textualization, contrasting fluidity (advancing a textual program) with disfluency, in which they distinguish three modes of textualization: stopping, repetition, and repair. Based on these categories, they describe recurrent scriptural sequences in the studied corpus, akin to textual production routines. These routines reinforce the observation of non-coincidence between the order in which the elements of the text are inscribed and their arrangement in the final text. Mahrer and Zuccarino conclude that a fundamental trait of writing as a genetic process is the adaptation of the order of textualization to the order of the text.

The articles collected in the present issue illustrate the diversity of approaches to written performance units. They all address the question of the temporality of the writing activity and, in an interconnected way, the segmentation of writing activity and text. They also highlight some of the factors that can affect this temporality and its segmentation. Pauses, as markers of scriptural rhythm and temporality, play an important role in these analyses and are examined in relation to other dimensions of writing, at the crossroads of linguistic, procedural and cognitive studies. This opens up interesting prospects. A fruitful way forward would be to continue pooling work from a variety of language contexts and writing production situations, thus enriching the empirical data for an exhaustive description and more advanced theorization of written performance units.

Acknowledgements

References

Abney, S.P. (1991). Parsing By Chunks. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-Based Parsing*. Springer Nature. https://doi.org/10.1007/978-94-011-3474-3_10

Ailhaud, E., Chenu, F., Jisa. H., (2016) A developmental perspective on the units of written French. In J. Perera et al. (Eds.), *Written and spoken language development across the lifespan: Essays*

in honor of Liliana Tolchinski (pp. 287-305). Springer Nature. https://doi.org/10.1007/978-3-319-21136-7_17

Alamargot, D., Dansac, C., Chesnet, D., & Fayol, M. (2007). Parallel processing before and after written text production. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 13–29). Elsevier.

Alves, R. A., Limpo, T., Fidalgo, R., Carvalhais, L., Pereira, L. Б., & Castro, S. L. (2016). The impact of promoting transcription on early text production: Effects on bursts and pauses, levels of written language, and writing performance. *Journal of Educational Psychology*, *108*(5), 665–679. https://doi.org/10.1037/edu0000089

Auer, P. (2005). Projection in interaction and projection in grammar. *Text*, *25*(1), 7–36. https://doi.org/10.1515/text.2005.25.1.7

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke Analysis: Reflections on Procedures and Measures. *Written Communication*, *29*(3), 246–276. https://doi.org/10.1177/0741088312451108

Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes? *Studies in Second Language Acquisition*, *41*, 529–554.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education Limited.

Bowen, N., & Van Waes, L. (2020). Exploring revisions in academic text: Closing the gap between process and product approaches in digital writing. *Written Communication*, *37*(3), 322–364.

Buttherworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*, 75–87.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press. https://doi.org/10.1017/s0142716400010523

Chanquoy, L., Foulin, J.-N., & Fayol, M. (1990). Temporal management of short text writing by children and adults. *Cahiers de Psychologie Cognitive*, *10*(5), 513–540.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in Writing: Generating Text in L1 and L2. *Written Communication*, *18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, *20*(1), 99–118. https://doi.org/10.1177/0741088303253572

Chesnet, D., Solier, C., Bordas, B., & Perret, C. (2024). A quick briefing on the new version of Eye and Pen (version 3.01): news and update. *Reading and Writing*, *37*(2), 253–264. https://doi.org/10.1007/s11145-022-10267-5

Cislaru, G., Eshkol-Taravella, I., & Almeida-Barreto, S. (2023). Automatic processing of real-time recorded writing: pausal segmentation versus chunking. *RADH 2023 - Recent Advances in Digital Humanities*, November 17–18, Timisoara, Romania.

Cislaru, G., Feltgen, Q., Khoury, E., Delorme, R., & Bucci, M. P. (2024). Language Processing Units Are Not Equivalent to Sentences: Evidence from Writing Tasks in Typical and Dyslexic Children. *Languages*, *9*(5), 155. https://doi.org/10.3390/languages9050155

Cislaru, G., & Olive, Th. (2017). Segments répétés, jets textuels et autres routines. Quel niveau de pré-construction? *Corpus*, *15*, 61-89. https://doi.org/10.4000/corpus.2846

Cislaru, G., & Olive, T. (2018). *Le processus de textualisation*. De Boeck. https://doi.org/10.3917/dbu.cisla.2018.01

Damian, M. F., & Stadthagen-Gonzalez, H. (2009). Advance planning of form properties in the written production of single and multiple words. *Language and Cognitive Processes*, *24*(4), 555–579. https://doi.org/10.1080/01690960802346500

Doquet, C. (1999) Analysis of the writing processes of 10-year-old children at school, with the help of the "Genèse du texte" word processing program. In M.-F. Crété & E. Espéret (Eds.), *Writing and learning to write at the dawn of the 21st Century: Proceedings of the 1998 Writing Conference* (pp. 127–132). Maison des Sciences de l'Homme et de la Société de Poitiers.

Doquet, C. (2004). Genѣse des textes, genèse des scripteurs : regards sur la langue dans l'écriture. In A. Piolat (Ed.), *Ecriture : approches en sciences cognitives*. Presses de l'Université de Provence.

Doquet, C. (2011). *L'Écriture débutante. Pratiques scripturales à l'école élémentaire.* Presses Universitaires de Rennes.

Doquet, C., Lefebvre, J., Mahrer, R., Testenoire, P.-Y. (2022). Oral/écrit. Quelles places dans les modѵles linguistiques? *Linguistique de l'écrit, Special Issue, 3*, 9–32. https://doi.org/10.19079/lde.2022.s3.1

Eco, U. (1979). *The Role of the Reader: Explorations in the semiotics of texts*. Indiana University Press / Hutchinson.

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication, 32*(4), 365. https://doi.org/10.2307/356600

Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition, 6*(1), 35–53.

Foss, D. J. (1982). A discourse on *semantic priming. Cognitive Psychology, 14*(4), 590–607.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Foulin, J.-N. (1995). Pauses et débits: les indicateurs temporels de la production écrite. *L'année psychologique, 95*(3), 483–504. https://doi.org/10.3406/psy.1995.28844

Friedlander, A. (1989). *The writer stumbles: Some constraints on composing in English as a second language*. Carnegie Mellon.

Galbraith, D., & Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing Writing: Insights from Keystroke Logging and Handwriting* (pp. 306–325). Brill. https://doi.org/10.1163/9789004392526_015

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*(4), 411–458.

Gilquin, G. (2020). In Search of Constructions in Writing Process Data. *Belgian Journal of Linguistics, 34*, 99–109. https://doi.org/10.1075/bjl.00038.gil

Gilquin, G. (2024). The Processing of Multiword Units by Learners of English: Evidence from Pause Placement in Writing Process Data. *Languages, 9*(2), 51. https://doi.org/10.3390/languages9020051

Goldman-Eisler, F. (1968) *Psycholinguistics: Experiments in Spontaneous Speech*. Academic Press.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du franзais: vitesse de parole et variables composantes, phénomѵnes d'hésitation. *Phonetica, 31*(3–4), 144–184.

Grosjean, F., & Dommergues, J.-Y. (1983). Les structures de performance en psycholinguistique. *L'année Psychologique, 83*(2), 513–536.

Hayes, J. R. (2012). *Modeling and remodeling writing. Written Communication, 29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hunt, K. W. (1970). Do Sentences in the Second Language Grow like Those in the First? *TESOL Quarterly, 4*(3), 195–202. https://doi.org/10.2307/3585720

Immonen, S., & Mѧkisalo, J. (2010). Pauses reflecting the processing of syntactic units in monolingual text production and translation. *HERMES-Journal of Language and Communication in Business, 44*, 45–61. https://doi.org/10.7146/hjlcb.v23i44.97266

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English, 20*, 121–140.

Kellogg, R. (2015). *The Psychology of Writing*. Oxford University Press.

Leblay, C. (2007). L'avant-texte comme texte sur le vif. Analyse génétique d'opérations d'écriture en temps réel. *Langue française, 155*(3), 101-113.

Leijten, M., Van Horenbeeck, E., & Van Waes, L. (2019). Analysing keystroke logging data from a linguistic perspective. In K. Sullivan and E. Lindgren (Eds), *Observing writing: Insights from keystroke logging and handwriting* (pp. 71-95). Amsterdam: Brill.

Leijten, M., & Van Waes, L. (2016). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication, 33*(4), 358–392. https://doi.org/10.1177/0741088313491692

Lemke, J. L. (2021). *Space, Time, and the Human Sciences: A Semiotic Approach to Meaning in Discourse*. Routledge.

Limpo, T. & Alves, R. A. (2017). Written language bursts mediate the relationship between transcription skills and writing performance. *Written Communication, 34*(3), 306–332.

Lindgren, E., & Sullivan, K. P. H. (2019). *Observing writing: Insights from keystroke logging and handwriting*. In R. Fidalgo & T. Olive (Series Eds.), *Studing in Writing, Vol. 38.* Brill.

Logan, G. D., & Crump, M. J. (2011). *Hierarchical control of cognitive processes: The case for skilled typewriting. Psychological Review, 118*(3), 437–460. https://doi.org/10.1016/b978-0-12-385527-5.00001-2

Maggio, S., Lété, B., Chenu, F., Jisa, H., & Fayol, M. (2012). Tracking the mind during writing: Immediacy, Delayed, and Anticipatory Effects on Pauses and Writing Rate. *Reading and Writing, 25*, 2131–2151. http://dx.doi.org/10.1007/s11145-011-9348-1

Manseri, K., Silai, I.-M., Eshkol-Taravella, I., & Cislaru, G. (2024). Analysing the Dynamics of Textualisation through Writing Bursts and Pauses in French. *LIFT 2*. https://lift2-2024.sciencesconf.org/resource/page/id/5

Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. In Gregg, L. W., & Steinberg, E. R. (Eds.), *Cognitive processes in writing* (pp. 113–134). Lawrence Erlbaum.

Matsuhashi, A. (1987). *Writing in real time: Modeling production processes*. Ablex Publishing.

Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: on the importance of where writers pause. *Reading and Writing, 30*(6), 1267–1285. https://doi.org/10.1007/s11145-017-9723-7

Nottbusch, G. (2010). Grammatical planning, execution, and control in written sentence production. *Reading and Writing, 23*, 777–801. https://doi.org/10.1007/s11145-009-9188-4

Ochs, E. (1979). *Planned and Unplanned Discourse*. Brill.

Olive, T. (2004). Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist, 9*(1), 32–42. https://doi.org/10.1027/1016-9040.9.1.32

Olive, T. (2014). Toward a parallel and cascaded model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research, 6*(2), 173–194. https://doi.org/10.17239/jowr-2014.06.02.4

Olive, T., Alves, R. A. & Castro, S. L. (2009). Impact of children's handwriting skill on text composition. *Handwriting Today*, 14–17.

Olive, T., Bouriga, S., Bordes, C. & Chesnet, D. (submitted). *Impact of the writing tool and of translating on bursts of writing: Comparing handwriting vs. typing in low and high demands of translating*.

Olive, T., & Cislaru, G. (2015). Linguistic forms at the process-product interface: Analysing the linguistic content of bursts of production. In G. Cislaru (Ed.), *Writing(s) at the Crossroads: The process-product interface* (pp. 99–123). John Benjamins.

Olive, T. & Passerault (2012). The visuospatial dimension of writing. *Written Communication, 29*, 326–343. https://doi.org/10.1177/0741088312451111.

Piolat, A. (1983). Localisation syntaxique des pauses et planification du discours. *L'année psychologique, 83*(2), 377–394.

Ransdell, S. E., & Levy, C. M. (1994). Writing as process and product: The impact of tool, genre, audience knowledge, and writer expertise. *Computers in Human Behavior, 10*(4), 511–527. https://doi.org/10.1016/0747-5632(94)90044-2

Roeser, J., Torrance, M., & Baguley, T. (2019). Advance planning in written and spoken sentence production. *Journal of Experimental Psycholoy: Learning, Memory and Cognition, 45*(11), 1983–2009. https://doi.org/10.1037/xlm0000685

Sausset, S., Lambert, E. & Olive, T. (2013). Flexibility of orthographic and graphomotor coordination during a handwritten copy task: effect of time pressure. *Frontiers in Psychology 4* :866. https://psycnet.apa.org/doi/10.3389/fpsyg.2013.00866

Schilperoord, J. (2002). On the cognitive status of pauses in discourse production." In T. Olive, & C. M. Levy (Eds.), *Contemporary Tools and Techniques for Studying Writing (pp. 59-85).* Kluwer Academic Publishers. http://dx.doi.org/ 10.1007/978-94-010-0468-8_4

Scott, C. M. (1988). Spoken and written syntax. In M. A. Nippold (Ed.), *Later Language Development. Ages nine through nineteen* (pp. 49–95). PRO-ED.

Scott, C. M. (2009). Language-based assessment of written expression. In G. A. Troia (Ed.), *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 358–385). The Guilford Press.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press.

Sinclair, J. M., & Mauranen, A. (2006). *Linear Unit Grammar: Integrating speech and writing.* John Benjamins.

Spelman Miller, K. (2002). Units of production in writing: evidence of topic 'framing' in on-line writing research. *Reading Working Papers in Linguistics, 6, 255–272.*

Tannen, D. (1982). Oral and Literate Strategies in Spoken and Written Narratives. *Language, 58*(1), 1–21.

Taylor I. (1969) Content and structure in sentence production, *Journal of Verbal Learning and Verbal Behavior*, 8, 170–175.

Ulasik, M. A., & Miletić, A. (2024). Automated Extraction and Analysis of Sentences under Production: A Theoretical Framework and Its Evaluation. *Languages*, 9(3), 71. https://doi.org/10.3390/languages9030071

Vachek, J. (1973). *Written Language. General problems and problems of English*. Mouton.

Van Bruggen, J. (1946). Factors affecting regularity of the flow on words during written composition. *Journal of Experimental Education, 15*, 133–155.

van Galen, G. P. (1991). *Handwriting: A developmental perspective. Advances in Psychology, 84*, 151–171. https://doi.org/10.1017/cbo9780511663284.016

Van Hell, J. G., Verhoeven, L., & Van Beijsterveldt, L. M. (2008). Pause time patterns in writing narrative and expository texts by children and adults. *Discourse Processes, 45*, 406–427. https://dx.doi.org/10.1080/01638530802070080

Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics, 35*(6), 829–853. https://doi.org/10.1016/s0378-2166(02)00121-2

Watson, D., & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica, 59*(2-3), 279–300. https://doi.org/10.1111/j.1467-9582.2005.00130.x

Wengelin, E., Johansson, R., Frid, J., & Johansson, V. (2024). Capturing writers' typing while visually attending the emerging text: A methodological approach. *Reading and Writing: An Interdisciplinary Journal, 37*(2), 265–289. https://doi.org/10.1007/s11145-022-10397-w

Wengelin, A., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eye tracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods, 41*, 337–351. http://dx.doi.org/10.3758/BRM.41.2.337

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition* (pp. 41–62). John Wiley.