Linguistic and Behavioral Alignment in Writing: A Scoping Review

Olena Vasylets[1] & Javier Marín[2]

[1] University of Barcelona, Barcelona | Spain

[2] University of Murcia, Murcia | Spain

Abstract: This article reports on a scoping review of the literature exploring the alignment between behavioral and linguistic units in L1 and L2 writing. Behavioral units in writing were assessed using keystroke logging measures of pauses, bursts, and revisions. Linguistic units were operationalized based on lexical and syntactic definitions from the literature. Nine empirical studies met the inclusion criteria. Most of these studies focused on L1 English writing by adult participants, although some explored other first languages, such as German, Dutch, and French. The identified L2 studies focused exclusively on English. Due to the limited number and high heterogeneity of the studies, no definitive conclusions can be drawn. However, meaningful links between behavioral and linguistic units were detected. In addition to confirming some previously known phenomena, the studies provided new evidence on online processing during pauses, revealed certain idiosyncrasies in L1 versus L2 writing cognition, and offered new insights into the nature of revision. We provide a critical interpretation of the results, propose new research directions, and recommend solutions.

Keywords: keystroke logging, bursts of writing, pauses, revision, linguistic analysis, idea units

journal of WRITING RESEARCH

Vasylets, O., & Marin, J. (2025) Linguistic and behavioral alignment in writing: A scoping review. *Journal of Writing Research*, *16*(3), 375-404. https://doi.org/10.17239/jowr-2025.16.03.02

Contact: Olena Vasylets, Department of Modern Languages Faculty of Philology and Communication, University of Barcelona Gran Vнa de les Corts Catalanes, 585 08007 Barcelona | Spain – vasylets@ub.edu - https://orcid.org/0000-0003-0241-1279

The advent of keystroke logging technology has promised a significant breakthrough in the study of writing. This technology offers the possibility of providing an objective and unobtrusive real-time record of the unfolding writing process, measured by various indicators such as pauses, bursts, and revisions (Beinhorn & Hollenstein, 2023; Leijten & Van Waes, 2013); Vandermeulen et al., 2020). However, merely registering writing behaviors is insufficient to interpret the underlying cognition. To gain deeper insights, it is essential to establish a meaningful relationship between the behavioral units measured by keystroke logging and linguistic units. This scoping review aims to assess the current knowledge on this issue by analyzing available publications.

First, we briefly describe behavioral units, expressed in terms of pauses, bursts, and revisions, and discuss the challenge of aligning keystroke measures with specific writing processes. Next, we define the most common linguistic units and describe their connection with textual complexity. Finally, we present the results of our scoping review, discuss implications, and provide suggestions for future research.

1. Behavioral units

Investigating the temporal dynamics of writing primarily requires studying pauses and execution periods (Olive, 2010). Accordingly, keystroke logging measures of pauses, bursts, and revisions have been considered to be the primary behavioral indicators associated with underlying writing cognition. In what follows, we provide definitions and relevant empirical findings about the behavioral units operationalized in terms of the above-mentioned measures.

1.1 Pauses

In Wengelin's (2006) definition, a writing pause is a "transition time between two keystrokes, which is longer than what can be expected to be necessary for the time needed to merely find the next key" (p.111). The importance of pauses in writing is hard to underestimate. According to Alamargot et al. (2007), "pausing can occupy up to 60 or 70% of total composition time" (p. 13), thus representing an integral part of writing. Although writers can pause for different reasons (e.g., a simple distraction can cause a disruption in composing), pauses are considered indicative of both micro- and macro-level mental activities. In terms of measurement, the important aspects of pauses are their length, frequency, location, and distribution.

Regarding length, longer pauses have been related to macro operations, such as paragraph organization, or to more cognitively demanding operations, such as text revision (Olive & Kellogg, 2002). Conversely, shorter pauses are typically associated with motor-related issues or micro-level processes, such as lexico-grammatical choices. Pause duration has also been shown to increase with grammatical unit level, with the longest pauses typically found between paragraphs, followed by sentences, word units, individual words, and parts of words (Schilperoord, 2001). This pattern for pause length is believed to indicate different levels of processing at different discourse boundaries

(Chanquoy et al., 1996). Pause location is also considered meaningful. For example, pauses between sentence boundaries are related to global text planning, while within-word pauses are associated with typing skills (Van Waes et al., 2021).

The most commonly accepted temporal threshold for a writing pause is 2 seconds (Wengelin, 2006). However, some scholars have questioned the appropriateness of this threshold, highlighting the potential moderating effects of external and internal factors on pauses (Van Waes et al., 2012). For example, pauses can vary depending on task characteristics (Schilperoord, 1996) or the mode of production (pen versus keyboard) (Van Waes & Schellens, 2003). Personal characteristics, such as typing skill (Barkaoui, 2016), writing expertise (Chukharev-Hudilainen et al., 2019), language proficiency (Ivaska et al., 2025, this issue; Thorson, 2000), or dyslexia (Sumner et al., 2013) can also influence pausing behavior.

1.2 Bursts

A burst represents a continuous production of text between interruptions (Baaijen & Galbraith, 2018; Feltgen & Lefeuvre, 2025, this issue). Bursts can occupy up to 50% of writing time. The early study by Kaufer et al. (1986) was the first to identify bursts, showing that adult L1 writers composed text in approximately nine-word segments separated by pauses. Primary measures related to bursts include their type, number, and length (Baaijen et al., 2012). The literature identifies various burst types, including p-bursts, defined as uninterrupted writing units delineated by pauses of a predefined length (e.g., 2 seconds); r-bursts, which are writing units bound by revisions (Hayes & Chenoweth, 2006; Kaufer et al., 1986); s-bursts, representing writing production interrupted by a transition to an external source (Leijten et al., 2014); and recently, Baaijen and Galbraith (2018) identified i-bursts, which start with a mouse or arrow key movement away from the leading edge, followed by text production.

Another important characteristic of a burst is its length, which is believed to be indicative of language production skills. Chenoweth and Hayes (2003) related burst length to the translating process, which gives prelinguistic ideas their corresponding linguistic form. With higher language and writing proficiency, there is more mental space for the translating process, allowing writers to activate various processes simultaneously, produce longer bursts, and increase textual quality (Alves et al., 2016). Burst length has been found to be constrained by typing skills (Alves et al., 2007). It has also been shown that with sufficient automatization of typing, higher-order processes of planning or revision can be activated simultaneously with translation during burst production (Olive, 2014). Empirical studies have shown that learners with special needs produce shorter bursts (Connelly et al., 2012) and that L1 writers tend to produce longer bursts than L2 writers (Chenoweth & Hayes, 2001).

1.3 Revision

Revision, defined as "making any changes at any point in the writing process" (Fitzgerald, 1987, p.484), is commonly regarded as a central part of writing. Revision can refer to both the product and process of writing, and can be either external (revision of the produced text) or internal (changes in the writer's mind). It can apply to different areas (spelling, lexico-grammatical choices, content) and can occur linearly or recursively at different stages of composing (draft, final version) (Allal & Chanquoy, 2004). Proficient revision is conceptualized as a problem-solving, cognitively costly, and strategic activity (Allal et al., 2004), and is considered indicative of writing development, with skilled writers revising more effectively than novices (Hayes et al., 1987). Other factors that can influence revision behaviors include writers' familiarity with the genre/topic or reading ability (McCutchen et al., 1994). For efficient revision, an interplay between linguistic and contextual knowledge is vital (Chanquoy, 2001), and the quality of revision is defined by how well writers adapt their text to their goals (Flower et al., 1986).

Primary keystroke measures related to revisions are defined in terms of revision operations (deletion, insertion), the level of revision (word, phrase, sentence), or spatial location (revision at the leading edge of the text, away from the leading edge) (Lindgren & Sullivan, 2006). It is worth noting that the operationalization of revisions tends to vary across studies. For instance, in Van Waes et al. (2014), revision was defined in terms of backspaces and delete keys, while insertions were ignored. Conversely, Conijn et al. (2021) included both deletions and insertions. Similarly, Eklundh and Kollberg (2003) provided an overview of the spatial and temporal locations of all insertions and deletions to analyze revision.

In conclusion, the behavioral units operationalized in terms of pauses, bursts, and revisions have proven to be promising in understanding the cognitive activities underlying writing. However, interpreting these measures is challenging due to various factors, including, inter alia, the moderating roles of contextual and individual differences. In the following section, we delve deeper into the problem of aligning keystroke measures with specific cognitive processes.

1.4 The challenge of measure-process alignment

As mentioned previously, the interpretation of behavioral units is not straightforward (Schilperoord, 2001). In speech, pauses are typically considered a direct reflection of planning, but pauses in writing are more complex. Because writing is slower and leaves a visible trace, pauses during composing may reflect not only planning but also revision. Additionally, a combination of various processes can take place during a pause (Olive, 2010). Bursts, typically associated with linguistic formulation, can also indicate an attempt at text improvement, classifiable as revision. Measures of revision (e.g., deletions, insertions) may reflect semi-automatic error correction (low-order

revision) or a strategic and purposeful attempt to modify/improve content (high-order revision).

Several authors have highlighted the challenge of aligning behavioral measures with specific writing processes and offered various solutions (Baaijen et al., 2012; Chenu et al., 2014; Galbraith & Baaijen, 2019; Wengelin et al., 2019). For example, Chenu et al. (2014) emphasized refining pause thresholds to distinguish between higher-order mental processes and lower-order graphomotor processes. Baaijen et al. (2012) proposed sorting the keystroke log into activities corresponding to global models of writing (Hayes, 2009), such as global planning, drafting, or post-drafting revision. This approach isolated pretext planning processes and distinguished between while-writing and post-draft revisions. Additionally, they suggested refining measures of pauses, bursts, and revisions, including a more detailed classification of bursts based on how they are initiated and terminated. Baaijen et al. (2012) found that writers who paused longer at sentence boundaries produced shorter but more well-formed bursts. Based on these findings, Galbraith and Baaijen (2019) argued that the key feature of a burst during extended composing is its cleanliness of production.

Wengelin et al. (2019) proposed a different set of procedures to improve behavior-cognition alignment. They advocated for a multi-method approach, such as the triple task (Olive et al., 2002), and triangulating keystroke data with eye-tracking data (Wengelin et al., 2009) or data from self-report instruments (Sala-Bubaré et al., 2021). Importantly, Wengelin et al. (2019) noted that while self-report methods could be useful for investigating revision processes, their use for investigating pauses was questionable. For example, think-aloud protocols could cause reactivity by inducing writers to pause longer and more frequently.

In sum, while the alignment of writing behaviors and processes is undeniably challenging, the issue is not unresolvable, and different sets of procedures have been used, including data triangulation. Another promising solution could lie in the alignment of behavioral units and linguistic units. In the following section, we provide definitions and empirical findings for the main linguistic units and establish their connection with the complexity of the produced text.

1.5 Linguistic units: Definition and types

Words are considered the basic linguistic unit (Biber et al., 1999). In written discourse, the relevant construct is an orthographic word, which can be defined as a written sequence delineated by white spaces at each end but without spaces in the middle.

In addition to individual words, there are multiple classes of multi-word units, including lexical bundles, collocations, and repeated segments. Lexical bundles, defined as frequently recurring sequences of words in a register (e.g., *I don't know if*; *I just wanted to*), are identified based on their rate of occurrence and distribution across texts (Biber & Barbieri, 2007). To qualify as a lexical bundle, a multi-word sequence

must be extremely common, must be used in multiple texts, and must not be idiomatic in meaning or represent an idiosyncratic use of a particular speaker or writer.

Collocations, which represent a different type of multi-word unit, do not have one simple and precise definition (Nation, 2001). Collocations are commonly defined as habitually occurring lexical combinations characterized by restricted co-occurrence of elements and relative transparency of meaning (Laufer & Waldman, 2011). Another common type of multi-word unit is repeated segments, defined as strings of at least two graphical units that occur together at least twice in a text or corpus (Lafon & Salem, 1983; Olive & Cislaru, 2015). Repeated segments represent discourse routines typical in a particular type of discourse and can commonly be found in professional discourse. Repeated segments lend themselves to both grammatical and semantic analysis (Cislaru et al., 2013).

Linguistic units can also be identified from the perspective of grammar. In this tradition, clauses, phrases, T-units, and sentences represent the central units in written discourse (Biber et al., 1999; Hunt, 1976). A clause is a unit structured around a verb phrase. A clause can be embedded within a larger structure, with the superordinate clause termed the main clause and the embedded clause termed the dependent clause. Another central discourse unit is a phrase. There are several major phrase types, each with a lexical word as the head and a number of accompanying elements. The main types of phrases are noun phrases, verb phrases, adjective phrases, and adverb phrases (Biber et al., 1999). The T-unit, defined as "one main clause plus whatever subordinate clauses happen to be attached or embedded within it" (Hunt, 1966, p. 735), was originally designed to assess L1 development in children. The T-unit has been widely used in the analysis of both L1 and L2 written discourse.

Finally, in grammatical tradition, sentences are regarded as the maximal syntactic units of language (see Ulasik et al., 2025, this issue). In written discourse, however, sentences can be identified rather straightforwardly by applying graphic and syntactic criteria. Graphically, a sentence can be identified by relying on the final punctuation mark (e.g., a full stop). Syntactically, a sentence is structurally autonomous: it is linked to the surrounding context merely by discursive relations, with the elements within the sentence being related to one another by morphosyntactic rules (Biber et al., 1999).

Another unit of discourse is the idea unit, defined by Chafe (1985) as a "spurt of discourse" (p. 106), which verbalizes the information active in the speaker's mind at its onset. Use of this discourse unit would allow to gauge the dimension of propositional complexity which refers to the amount of information (expressed as the number of idea units) which a speaker or writer encodes to convey the intended message (Ellis & Barkhuizen, 2005). Although originally conceptualized as a unit of speech, Halliday (1985) and Chafe (1985) agreed that punctuation in written production reflects covert prosody and can be used as a cue to the boundaries of idea units. While theorists have primarily conceptualized this unit as phonological (Halliday, 1985), others have treated the idea unit as a semantic unit of planning and a unit of cognitive processing and

memory (Chafe, 1985). Chafe (1985) also posits that "some of the differences between written and spoken language can be understood only with reference to the notion of idea units" (p. 106). Online pressures can limit speakers, but the increased time and editability of writing allow the creation of "expanded idea units" (Chafe & Danielewicz, 1987, p. 86). Halliday (1985) and Chafe (1985) also emphasized the close (but not fixed) association between the idea unit and the clause, suggesting that a prototypical idea unit would typically take the form of a clause.

Importantly, the number and characteristics of linguistic units are crucial for assessing text complexity. According to typological linguistics, complexity can be defined by the number of linguistic units and the interconnections between them (Dahl, 2004). Bulté and Housen (2014) summarize this idea with the following points: (1) more units indicate higher complexity (e.g., a text with more words or idea units is more complex); (2) longer linguistic units are more complex (e.g., greater word, clause, or idea length indicates higher complexity); (3) greater embedding (e.g., a higher subordination index) indicates higher complexity; (4) more varied or diverse linguistic units indicate higher complexity of production; (5) more marked, infrequent, sophisticated, or later-acquired features indicate higher complexity. Empirical studies applying these assumptions have shown that more complex academic writing in L1 is characterized by a higher incidence of low-frequency words (McNamara et al., 2010), longer and less familiar words (Crossley et al., 2011), more specific words (McNamara et al., 2010), and a higher incidence of sophisticated phrases (Crossley et al., 2012). Similarly, more advanced L2 writing is characterized by higher lexical sophistication (Crossley, 2020; Crossley & McNamara, 2012), a greater range of lexical phrases (Kyle & Crossley, 2015), and longer and more varied syntactic structures (Lu, 2011; Ortega, 2003).

To conclude, assessing the number, type, and interrelationships of linguistic units provides direct cues to text complexity. By extension, establishing the relationships between linguistic and behavioral units can further illuminate the interpretation of cognitive signals captured through keystroke logging techniques.

2. Method

2.1 Review Design

Aiming to take stock on the existent findings concerning the links between behavioral and linguistic units in written discourse, we conducted a scoping review. According to Munn et al. (2018), scoping reviews are useful to identify the available evidence in the field, clarify key concepts/definitions, examine design and methodology adopted by studies on a certain topic, and identify and analyze knowledge gaps, inter alia.

Search Procedures and Selection of Studies

To identify relevant studies, we searched major databases in linguistics, psychology and education, including Educational Resources Information Center (ERIC), EBSCOhost, Linguistics and Language Behavior Abstracts (LLBA), PsycINFO, ScienceDirect, PsychArticles, ProQuest Dissertations. We also consulted search engines (Google Scholar and Google) and the websites of peer-reviewed journals with the orientation in L1 and L2 writing research, linguistics, psychology and education. Various combinations of the following key words were employed: *keystroke logging*, *L1*, *L2*, *writing*, *text quality*, *complexity*, *accuracy*, *fluency (CAF) measures*, *alignment*, *behavioral units*, *linguistic units*. References in primary studies, reviews and meta-analysis on similar topics were also consulted to ensure a complete study selection.

Inclusion criteria

To filter the studies after the primary search, the following inclusion criteria were employed:

1. Study must employ keystroke methodology.
2. Study must focus on L1 or L2 writing, or on both.
3. Study can include both healthy participants and/or participants with cognitive/learning disabilities.
4. Study must explore a relationship between behavioral (keystroke) measures and discrete linguistic measures of written texts.
5. Linguistic units must correspond to the lexical/syntactic units recognized in the literature (e.g., repeated segments, clauses, T-units, idea units, etc.).
6. Linguist data for the analysis must come from a coherent text.
7. Writing task can be produced under a variety of conditions, including laboratory, classroom or any kind of naturalistic condition.
8. Writing task must be performed using a keyword (computer-based writing modality).
9. Study must be a primary study published as an article in a peer-reviewed journal, book chapter in a volume from a prestigious editorial, PhD dissertation, conference proceedings, or a published report of a research project funded by a prestigious official body.
10. Study must be published in English.
11. Study must be published between 2000 and 2023.

Exclusion criteria

The exclusion criteria were as follows:

1. Study does not report sufficient information on the way keystroke measures were defined.

2. Study does not report sufficient information on the way linguistic units were defined and assessed.
3. Study does not report sufficient information on the participants.
4. Study does not report sufficient information on the writing task or procedure.
5. Study in which linguistic properties of the written text are gauged by means of holistic assessment, rather than employing discrete linguistic units.
6. Linguistic data in the analysis come from discrete isolated units (e.g., unconnected individual sentences).
7. Study published before 2000.

Coding procedure

A coding scheme was developed in order to systematically capture the characteristics of each study in the sample. The following broad categories of features were determined:

1. Type of the publication (e.g., journal article, book chapter, etc.)
2. Design (qualitative, quantitative, classroom-based, experimental, etc.).
3. Study aim/focus.
4. Type of writing (L1/L2) and target language.
5. Participants.
6. Procedure and writing task.
7. Operatinalization of behavirol units.
8. Definition/ measurement of linguistic units.
9. Software to capture behavirol units.
10. Analysis of linguistic units (tools, etc.).
11. Main findings.

The coding scheme underwent several rounds of revisions and was piloted before its applications.

3. Results

After applying the inclusion/exclusion criteria filters, nine studies were identified as meeting the goals of this scoping review. Table 1 presents information based on our main coding categories, including the study author(s) and type of publication (under the 'study' column), study design, aim, number and type of participants, procedure and writing task(s), behavioral and linguistic data analysis, software used (under the 'data analysis' column), and the main results. The reviewed studies comprised three book chapters, three journal articles, two conference proceedings, and one research report. In terms of design, the majority of the studies (7 out of 9) were quantitative experiments conducted under laboratory conditions, while two publications (Kerz et al., 2020; Olive & Cislaru, 2015) were corpus-analysis studies using naturalistic data. While exploring the relationship between behavioral and linguistic units was the primary aim

in some studies (e.g., Kerz et al., 2020; Leijten et al., 2015, 2019; Olive & Cislaru, 2015), others had additional goals, such as investigating the mediating effects of working memory in L2 writing of test-takers (Révész et al., 2017) or examining the relation between pauses, genre, and transcription fluency (Medimorec & Risko, 2017).

Most publications focused on L1 writing, with L1 English being predominant (n=3 studies). Other languages explored included L1 French (Olive & Cislaru 2015), L1 Dutch (Leijten et al., 2015), and L1 German (Mahlow et al., 2024). Research also focused on L2 writing, exclusively targeting English. For example, Révész et al. (2017) explored L2 English writing by Chinese learners, while Kerz et al. (2020) analyzed L2 English texts written by L1 German participants. Only one study (Leijten et al., 2019) contrasted L1 Dutch and L2 English writing.

Regarding participants, most studies focused on young, healthy adults (university students). Mahlow et al. (2024) included a wider age range, with L1 German participants aged 32 to 64. Olive and Cislaru (2015) focused on adult professionals, analyzing reports written by six L1 French social workers of unspecified age. Only one study (Leijten et al., 2015) focused on elderly writers, comparing the L1 Dutch writing of a participant with dementia (79 years old) to that of a healthy match (81 years old). The number of participants in experimental studies varied widely, from large samples like n = 189 in Goodkind and Rosenberg (2015) and n=101 in Medimorec and Risko (2017) to smaller samples like n = 30 in Révész et al. (2017). Three studies declared themselves as exploratory and included small samples of n = 18 in Leijten et al. (2019), n = 7 in Mahlow et al. (2024), n = 2 in Bowen (2019), and n = 2 in Leijten et al. (2015).

The type of writing, writing conditions in the lab, and the number of writing samples collected from each participant also varied. Some studies included less cognitively demanding tasks like knowledge-telling (Leijten et al., 2019) or video retelling (Mahlow et al., 2024), while others required participants to produce argumentative essays (Bowen, 2019), IELTS Academic writing tasks (Révész et al., 2017), or both narrative and argumentative tasks (Medimorec & Risko, 2017). In some cases, the research focus determined the writing task type. For example, to factor out task complexity effects, Goodkind and Rosenberg (2015) used 10-12 prompts of varying cognitive load for each participant. Leijten et al. (2015) used picture description tasks typical in dementia studies. The two studies using naturalistic written data varied greatly in corpus size. Olive and Cislaru (2015) analyzed a small corpus of six reports by L1 French social workers, while Kerz et al. (2020) examined a large corpus of 3,454 reflective learning journals by L1 German learners of L2 English.

Regarding behavioral units, some studies focused exclusively on pauses (Medimorec & Risko, 2017) or revisions (Bowen, 2019). However, most studies operationalized behavioral units in multiple ways, focusing on both pauses and revisions (Kerz et al., 2020; Leijten et al., 2015) and also on bursts (Leijten et al., 2019; Olive & Cislaru, 2015).

There was considerable variation in the definition and measurement of linguistic units. The word was the central linguistic unit in most studies. Leijten et al. (2019) used part-of-speech tagging and analyzed word length and frequency. Several studies (e.g., Révész et al., 2017; Medimorec & Risko, 2017; Kerz et al., 2020) included word diversity and sophistication in their analyses. Some focused exclusively on lexically defined units, such as repeated segments (Olive & Cislaru, 2015) or multiword units (Goodkind & Rosenberg, 2015). However, most studies considered both lexically and syntactically defined units (clauses, sentences) (Kerz et al., 2020; Medimorec & Risko, 2017; Révész et al., 2017). Mahlow et al. (2024) introduced the concepts of version and transforming sequence in their exploratory study. They defined versions as points in the production history of a text (i.e., specific text-produced-so-far) and transforming sequences as the textual material combined with edit operations that constitute the difference between two adjacent versions. They also introduced text history and sentence histories as analytical concepts.

In terms of software, Inputlog was most commonly used (Leijten et al., 2015, 2019), although some studies used other alternatives, such as Scriptlog (Mahlow et al., 2024). There was greater variation in the analysis of linguistic units. While some studies used Inputlog's affordances (Leijten, 2015, 2019), others (Goodkind & Rosenberg, 2015; Olive & Cislaru, 2015; Medimorec & Risko, 2017; Révész et al., 2017) employed alternative natural language processing tools (e.g., CohMetrix, Graesser et al., 2004).

Due to the heterogeneous nature of the studies in terms of specific aims, participants, and writing tasks, there was significant variation in the results. However, some results confirmed findings from previous writing investigations. For example, Medimorec and Risko (2017) and Révész et al. (2017) reported a positive correlation between pause rate and linguistic unit size, with increased pause rates from smaller (word) to larger (sentence, paragraph) units.

New evidence revealed a complex but meaningful pattern of relationships between behavioral and linguistic units. In addition to replicating the correlation between pauses and linguistic unit hierarchy, Medimorec and Risko (2017) showed that pauses in L1 writing were related to the hierarchy of linguistic units regardless of task genre or transcription skills. Specifically, word frequency decreased with increased pause rates at word boundaries, while pause rates at sentence boundaries had a marginally significant negative correlation with lexical diversity and a positive correlation with sentence length. Further findings about the relationship between pauses and lexical units in L1 writing come from Goodkind and Rosenberg (2015). They reported that pauses within multiword units varied depending on the task's cognitive demands, with longer pauses observed during the production of multiword units under greater cognitive demands.

In L2 writing, Révész et al. (2017) found that more frequent pausing between lower textual units was linked to less sophisticated lexis, lower syntactic complexity, and lower IELTS scores. Higher-order revisions, however, predicted more sophisticated

lexis. These findings align with Kerz et al. (2020), who reported that more fluent production correlated with higher syntactic complexity. Revealing results from the L1-L2 comparative study by Leijten et al. (2019) showed longer p-bursts, longer texts, and more fluent production in L1 Dutch compared to L2 English writing. There was higher variation in pause length before different word categories in L2 English, with longer pauses before nouns and conjunctions and shorter pauses before adjectives.

Insights into the nature of professional L1 writing come from Olive and Cislaru (2015), who analyzed morphosyntactic and semantic aspects of bursts and repeated segments in social workers' reports. They found less than 3% convergence but identified a high proportion of noun-based constructions among repeated segments and bursts. This nominal dimension of professional discourse resonates with Bowen's (2019) finding that noun groups were the most frequently revised constituents in academic writing by L1 English university students. Finally, in the study with participants with dementia, Leijten et al. (2015) reported that the participant with dementia had longer pauses when producing nouns, verbs, and pronouns, used more adjectives, and exhibited less regular use of connectives.

4. Discussion

4.1 Interpretation of the research findings

Despite the limited number of empirical studies, a discernible pattern of nuanced and meaningful relationships between behavioral and linguistic units in writing has emerged. The reviewed studies have replicated and extended several previously known phenomena, such as the boundary effect (i.e., increased pause duration with higher grammatical unit levels), longer p-bursts in more proficient writing, the complex nature of revision processes, and the nominal nature of academic discourse.

Importantly, new insights have also been obtained. For example, findings by Leijten et al. (2015, 2019) highlighted the added value of linguistic diversification in between-word pausing patterns, which adds depth to the interpretation of pausing. Their exploration of the links between behavioral units (pauses) and linguistic units (parts of speech) revealed differences in linguistic processing and writing cognition between healthy and cognitively impaired elderly writers. This suggests that behavioral-linguistic writing alignment could serve as a diagnostic tool for the onset and progression of cognitive decline.

A fine-tuned analysis of pause-linguistic alignment has also led to new discoveries regarding online processing in L1 writing. By aligning pauses with different linguistic units (words, sentences, paragraphs), Medimorec and Risko (2017) found that pauses at various locations correspond to different psycholinguistic processes. Their study indicated that pauses at word boundaries are linked to online lexical processing, while pauses at sentence boundaries relate to syntactic processing. They emphasized the

benefit of examining pauses at different locations separately, rather than considering overall pause rates and durations across the text.

Comparing behavioral-linguistic alignment in L1 versus L2 writing has yielded new insights into the nature of writing cognition in L1 and L2 writers. For example, Leijten et al. (2015) showed higher variation in pause length before different word categories in L2 English, suggesting a higher cognitive load in L2 writing. Medimorec and Risko (2017) found that increased pause rate in L1 writing was positively related to lexical sophistication and syntactic complexity. Conversely, Révész et al. (2017) found that in L2 writing, more frequent pausing predicted less sophisticated lexis, lower syntactic complexity, and lower overall writing task scores. These findings suggest that pausing time is invested differently in L1 versus L2 writing. For L1 writers, pauses may reflect cognitive effort in conceptualization and formulation, positively affecting text quality. This association between cognitive load and pausing behaviors was also empirically confirmed by Goodkind and Rosenberg (2015), who reported longer pauses during the production of multiword units under greater cognitive demands in L1 writing. For L2 writers, pauses might indicate cognitive struggles, leading to lower linguistic quality. To gain further insights into the idiosyncrasies of L1 and L2 writing cognition, more studies with within-learner designs and reinforced methods (e.g., triangulation with eye-tracking and/or stimulated recall) are needed.

Additionally, insights were obtained regarding the nature of revision in writing. For instance, Révész et al. (2017) confirmed the value of revision in L2 writing, showing that higher-order revisions were related to higher textual quality. Studies on behavioral-linguistic alignment have also demonstrated the potential of dynamically studying the revision process (Bowen, 2019; Mahlow et al., 2024).

Despite these promising insights, high variation in study research foci, participants, tasks, measures, and analysis methods, as well as certain methodological limitations, restrict our interpretation of the significance of behavioral-linguistic alignment in writing cognition. In what follows, we provide a critical analysis of some of the limitations in previous studies and offer suggestions for future research.

4.2 Suggestions for Future Research

Enhancing research generalizability and replicability

First, it is important to note the limited generalizability of previous research. Most available studies have predominantly focused on L1 English writing of young adults without cognitive impairment. Moreover, there is significant variability in the writing tasks used, ranging from less cognitively demanding picture descriptions to more challenging argumentative essays.

To gain further insights, future research should diversify in several areas: type of writing (L1 or L2), language (including studies in languages other than English), participants (varying in age, cognitive abilities, writing expertise, level of L2

proficiency, etc.), and writing tasks (covering different genres and types of discourse, as well as different conditions such as writing under testing situations versus writing for pleasure - for some recent examples, see Kruse, 2024; Tian et al., 2024). A variety of study designs would also be beneficial. In addition to experimental cross-sectional laboratory-based studies, classroom-based research where writing tasks are performed under more ecologically valid conditions would be valuable. Longitudinal studies are also needed to track changes in the potentially dynamic relationship between writing behaviors and production over time.

Increasing the generalizability of findings is crucial, but so is enhancing replicability. Recently, several authors have highlighted a lack of research replicability (Bohannon, 2015; Oberauer & Lewandowsky, 2019; Stroebe, 2016). The causes of the "replication crisis" are complex and multifactorial and are beyond the scope of this paper. However, methodological factors such as data collection, analysis, and reporting have been identified as central explanatory factors for the low rate of replication (Open Science Collaboration, 2012).

To maximize the replicability of findings and resolve some current inconsistencies and debates, recent research in experimental psychology, particularly studies on word processing, has been characterized by the tendency to conduct mega-studies involving large corpora with thousands of participants (Keuleers & Balota, 2015). Although the methodology of these mega-studies cannot be easily applied to writing research, the idea of accumulating comparable data about keystroke patterns and their alignment with linguistic units is promising. The accumulation of large datasets could set a reliable benchmark for the interpretation of research outcomes. (For some recent examples of corpus studies, see, Hasund & Hasselgerd, 2022; Miaschi et al., 2021).

Refining and expanding behavioral and linguistic measures

Future research should strive for greater consistency in defining behavioral units. As Olive and Cislaru (2015) emphasized, focusing on bursts and repeated segments, "the notion of burst thus has to be fixed, and the ideal pause duration settled, before they can be assigned the status of psycholinguistic counterparts of formulaic language" (p. 120). Similarly, Leijten et al. (2015) argue that the concept of pause requires further refinement, as it remains too broad even when subdivided into different levels (e.g., character, word, sentence) (see also Hall et al., 2022). Olive and Cislaru (2015) also highlight the benefit of more fine-grained semantic analysis of bursts and lexical units.

Another point to consider is that current studies have explored only a limited range of linguistic units, primarily focusing on words (parts of speech), some types of lexical phrases (e.g., repeated segments or broadly defined multiword units), or sentences. Future research should broaden the scope of linguistic units to include, for example, idea units. As previously explained, idea units are primarily semantic discourse units traditionally linked to cognitive processing and memory. Exploring the alignment between behavioral measures and idea units could bring new insights into writing

cognition. Of importance, practical guidelines for segmenting written production into idea units are available (Vasylets, 2017). These guidelines define an idea unit as semantically integral (conveying a message that constitutes an indivisible block of meaningful information) and meaningful (making sense cognitively to both the speaker/writer and the listener/reader). Syntactically, an idea unit can be part of a clause, a clause (a prototypical idea unit), or a constituent of a larger syntactic construction, such as a T-unit or sentence. When there is a strong conceptual dependency between the main and subordinate clauses (e.g., restrictive relative clause), they form a single idea; with a loose conceptual relationship (e.g., adverbial clause in the initial position), the main and subordinate clauses represent separate ideas. Like repeated segments, idea units can be analyzed both semantically and grammatically. In terms of empirical findings, early research in L1 writing reported that ideas tend to be longer and more complex in written discourse compared to speech (Chafe & Danielewicz, 1987). More recent L2 studies have shown that writers and speakers complexify their discourse differently, with speakers producing more idea units and writers producing more complex ideas in the same task (Vasylets, 2017). Vasylets et al. (2017) also reported that more complex written tasks resulted in a higher incidence of complex ideas compared to simpler tasks. These advances in theory and empirical research make idea units a viable construct for further investigating linguistic-behavioral alignment.

Future studies should also incorporate analyses of textual quality and provide psycholinguistically justified explanations of their findings. As mentioned in the introduction (see the sub-section 1.5), the connection between the characteristics of linguistic units and discourse complexity is relatively straightforward and can be easily assessed. Information about textual quality would add an additional dimension to the findings and provide both theoretical and pedagogical implications.

Accounting for moderating factors

Another potential enhancement in the design of keylogging experiments could involve considering various external and internal factors that can moderate writing behaviors and performance. Among the external factors are the type of writing task (e.g., narrative versus argumentative), conditions (timed versus untimed writing, writing with or without sources), and writing topic or type of discourse (Berman & Verhoeven, 2002; Cahyono et al., 2024; Plakans & Gebril, 2013).

Research has shown that several cognitive characteristics (such as working memory) and affective states (e.g., self-efficacy) are relevant in writing (Ahmadian & Vasylets, 2021; Papi et al., 2022; Paul et al., 2021; Van Drie et al., 2021; Vasylets & Marнn, 2021). Writing anxiety is another emotional state whose effects on writing behaviors and production may be significant (Cheng et al., 1999). For instance, Horwitz et al. (1986) found that students with high levels of writing anxiety tended to write shorter

texts compared to their less anxious peers. This reduction in text length may affect pauses, bursts, or other behavioral measures obtained through keystroke logging.

To our knowledge, individual writer differences have not been included as a potential mediating factor in studies exploring behavioral-linguistic alignment (but see Wirtz, 2025, this issue, for a notable exception). The current methodology of Linear Mixed Models (Baayen, 2008; Winter, 2016) offers the possibility of including individual writer differences as random or fixed factors, thus eliminating any possible noise from these variables.

Comparison with handwriting

The interpretation of measures from keystroke logging could benefit from study designs that include planned comparisons between pen-and-paper and computer-based writing modalities. Writing processes may be influenced and even contaminated by the idiosyncrasies of the writing instrument, as demonstrated for both young writers (Mangen & Balsvik, 2016) and adults (Bouriga & Olive, 2021; Christensen, 2004, 2009). It has also been shown that writing modality can affect the allocation of cognitive resources during writing (Vasylets et al., 2022). Handwriting and typewriting may differ in their cognitive demands, influenced by factors such as familiarity with the particular writing modality (Vasylets & Marнn, 2022). For example, Bouriga and Olive (2021) found that typewriting was more cognitively demanding for undergraduate students, which was attributed to their low typing skills.

It is important to consider that effects consistent across modalities can be considered more reliable than those specific to one modality, due to potential task effects. To isolate the modality effect, future studies could investigate whether the patterns of pauses, bursts, or revisions change for the same participants performing the same task in different writing modalities. Any common patterns emerging from this planned comparison could be attributed to the central processes common to both writing modalities. Conversely, potential differences would reveal some idiosyncratic aspects of handwriting and typewriting.

5. Conclusion

The goal of this scoping review was to identify and analyze empirical studies exploring the alignment between behavioral units, as measured by keystroke logging, and linguistic units, operationalized in terms of discrete lexical and syntactic units (e.g., repeated segments, clauses). Our search revealed a limited number of studies on behavioral-linguistic alignment. Most of these studies focused on L1 writing by adult writers without cognitive impairments, predominantly in English, although some also investigated other languages such as German, Dutch, and French. The studies on L2 writing focused exclusively on English. Due to the limited number and high heterogeneity of the studies, definitive conclusions cannot yet be drawn. However,

cumulative results clearly show a meaningful link between behavioral and linguistic units in writing.

In addition to confirming some known phenomena, these studies have contributed new insights, such as evidence for different types of linguistic processing at various text boundaries and potential differences in writing cognition between L1 and L2. The studies also highlighted the potential of behavioral-linguistic alignment to illuminate the dynamic unfolding of complex writing cognition, such as revision. Moreover, this method shows promise for understanding both L1 and L2 writing in participants from the general population and those with special needs.

Importantly, the small number of studies and methodological limitations, such as the small sample sizes and the exploratory nature of some investigations, preclude firm conclusions at this point. Nevertheless, we believe that the reviewed studies provide a strong foundation for future research. To enhance the replicability and generalizability of findings, future studies should incorporate several methodological improvements, such as refining outcome measures, using more sophisticated data analysis techniques, and controlling for confounding and moderating variables.

This line of research is promising but challenging. The challenges arise from the interdisciplinary nature of this research, which requires the integration of writing studies, linguistics, and computer science, among other fields. However, this interdisciplinarity is also a strength, as it allows for a more holistic, innovative, and effective approach to addressing the complex and multilayered phenomenon of writing.

References

Ahmaddian, M., & Vasylets, O. (2021). The role of cognitive individual differences in writing performance and written corrective feedback processing and use. In R.M. Manchyn & C. Polio (Eds*.), The Routledge Handbook of Second Language Acquisition and Writing* (pp. 139-152). Routledge. https://doi.org/10.4324/9780429199691-17

Allal, L., & Chanquoy, L. (2004). Introduction: revision revisited. In L.Allal, L Chanquoy, & P.Largy (Eds.), *Revision cognitive and instructional processes* (pp. 1-7). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-1048-1_1

Allal, L., Chanquoy, L., & Largy, P. (2004). *Revision Cognitive and Instructional Processes* (Vol. 8, p. 200). Springer. https://doi.org/10.1007/978-94-007-1048-1

Alamargot, D., Dansac, C., Chesnet, D., & Fayol, M. (2007). Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 13–29). Elsevier. https://doi.org/10.1163/9781849508223_003

Alves, R. A., Castro, S. L., de Sousa, L., & Strumqvist, S. (2007). Influence of typing skill on pause–execution cycles in written composition. In M.Torrqance, L. van Waes, & D. Galbraith (Eds.), *Writing and Cognition* (pp. 55-67). Brill. https://doi.org/10.1163/9781849508223_005

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. https://doi.org/10.1037/edu0000089

Baaijen,V.M., Galbraith, D.,&deGlopper, K. (2012). Keystroke analysis: Reflections on procedures andmeasures. *Written Communication*, *29*, 246–277. doi:10.1177/0741088312451108

Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction*, 36(3), 199–223. https://doi.org/10.1080/07370008.2018.1456431

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. *2001*, 390. https://doi.org/10.1017/CBO9780511801686

Baldridge, J. (2002). *The opennlp maximum entropy package*. OpenNLP download | SourceForge.net

Barkaoui, K. (2019). What can L2 writers` pausing behavior tell us about their writing processes? *Studies in Second Language Acquisition*, *41*(3), 529-554. https://doi.org/10.1017/S027226311900010X

Beinborn, L., & Hollenstein, N. (2023). *Cognitive plausibility in natural language processing*. Springer Nature. https://doi.org/10.1007/978-3-031-43260-6

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy*, 5 (1), 1-43. https://doi.org/10.1075/wll.5.1

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, *26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, G. (1999). *Longman grammar of spoken and written English*. Pearson Education. https://doi.org/10.1017/s002222670221 1627 Bohannon, J. (2015). Many psychology papers fail replication test. *Science*, *349*(6251), 910–911. https://doi.org/10.1126/science.349.6251.910

Bouriga, S., & Olive, T. (2021). Is typewriting more resources-demanding than handwriting in undergraduate students? *Reading and Writing*, *0123456789*. https://doi.org/10.1007/s11145-021-10137-6

*Bowen, N. (2019). Unfolding choices in digital writing: A functional perspective on the language of academic revisions. *Journal of Writing Research*, *10*(3), 465-498. https://doi.org/10.17239/jowr-2019.10.03.03

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Cahyono, B. Y., Irawati, R., Amalia, S. N., & Hidayat, L. E. (2024). Project-based Learning in EFL educational settings: A meta-analysis study in EFL/ESL writing. *Journal of Writing Research*, 16(1), 105-127.https://doi.org/10.17239/jowr-2024.16.01.04

Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In D.Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing (pp.* 105-123), Cambridge University Press.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & Samuels (Eds.), *Comprehending oral and written language* (pp. 83-113). Academic Press. https://doi.org/10.1163/9789004653436_007

Chanquoy, L. (2001). How to make it easier for children to revise their writing: A study of text revision from 3rd to 5th grades. *British Journal of Educational Psychology*, *71*(1), 15-41. https://doi.org/10.1348/000709901158370

Chanquoy, L., Foulin, J. N., & Fayol, M. (1996). Writing in adults: A real time approach. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 36–43). Amsterdam University Press.

Cheng, Y. S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, *49*(3), 417–446. https://doi.org/10.1111/0023-8333.00095

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

Chenu, F., Pellegrino, F., Jisa, H., & Fayol, M. (2014). Interword and intraword pause threshold in the writing of texts by children and adolescents: A methodological approach. *Frontiers in Psychology*, 5, 182. https://doi.org/10.3389/fpsyg.2014.00182

Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., & Feng, H. (2019). Combined deployable keystroke logging and eyetracking for investigating L2 writing fluency. *Studies in Second Language Acquisition*, 41, 583–604. https://doi.org/10.1017/S027226311900007X

Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written text. *British Journal of Educational Psychology*, *74*, 551–564. https://doi.org/10.1348/0007099042376373

Christensen, C. A. (2009). The Critical role handwriting plays in the ability to produce high-quality written text. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development* (pp. 284–299). SAGE Publications. https://doi.org/10.4135/9780857021069.N20

Cislaru, G., Sitri, F., & Pugniиre-Saavedra, F. (2013). Figement et configuration textuelle: les segments de discours rйpйtйs dans les rapports éducatifs. *Text-structuring. Across the line of speech and writing variation, Louvain-la-Neuve, PU Louvain*, 165-183.

Conijn, R., Speltz, E. D., Zaanen, M. V., Waes, L. V., & Chukharev-Hudilainen, E. (2022). A product-and process-oriented tagset for revisions in writing. *Written Communication*, *39*(1), 97-128. https://doi.org/10.1177/074108832110521

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (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*(3), 278-302. https://doi.org/10.1177/07410883124511

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2, 223–235. https://doi.org/10.1093/applin/II.3.223

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415-443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., Cai, Z., & McNamara, D. S. (2012). Syntagmatic, paradigmatic, and automatic ngram approaches to assessing essay quality. In P. M. McCarthy & G. M. Young Eds.). *Proceedings of the 25th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 214-219). Menlo Park, CA: The AAAI Press.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35 (2), 115-135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., Weston, J., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28 (3), 282-311. https://doi.org/10.1177/0741088311410188

Dahl, Ц. (2004). *The growth and maintenance of linguistic complexity*. John Benjamins. https://doi.org/10.1075/slcs.71

De Smet, M. J. R., Leijten, M., & Van Waes, L. (2018). Exploring the process of reading during writing using eye tracking and keystroke logging. *Written Communication*, *13*, 411- 447. https://doi.org/10.1177/0741088318788070

Eklundh, K. S., & Kollberg, P. (2003). Emerging discourse structure: computer-assisted episode analysis as a window to global revision in university students' writing. *Journal of Pragmatics*, *35*(6), 869-891. https://doi.org/10.1016/S0378-2166(02)00123-6

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford University Press.

Feltgen, Q, & Lefeuvre, F. (2025). Clitic subjects as landmarks in the writing production process: A study based on a keylog-derived corpus of writing bursts. *Journal of Writing Research*, *16*(3), 433-460. https://doi.org/10.17239/jowr-2025.16.03.04

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, *57*(4), 481-506. https://doi.org/10.3102/00346543057004481

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

*Goodkind, A., & Rosenberg, A. (2015, June). Muddying the multiword expression waters: How cognitive demand affects multiword expression production. In *Proceedings of the 11th Workshop on Multiword Expressions* (pp. 87-95). https://doi.org/10.3115/v1/W15-0914

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, *36*(2), 193-202. https://doi.org/10.3758/BF03195564

Hall, S., Baaijen, V. M., & Galbraith, D. (2022). Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing*, 37, 329-357. https://doi.org/10.1007/s11145-022-10284-4

Halliday, M. K. A. (1985). *An introduction to functional grammar*. First Edition. London: Edward Arnold.

Hasund, I. K., & Hasselgerd, H. (2022). Writer/reader visibility in young learner writing: A study of the TRAWL corpus of lower secondary school texts. *Journal of Writing Research*, *13*(3), 447-472. ttps://doi.org/10.17239/jowr-2022.13.03.04

Hayes, J.R., & Chenoweth, N.A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, 23(2), 135–141. https://doi.org/10.1177/0741088306286283

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S.Rosenberg (Ed.), *Advances in applied psycholinguistics*, *2*, (pp.176-240). Cambridge University Press.

Horwithz, E. K., Horwithz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Hyland, K. (2009). Teaching and Researching Writing. *Teaching and Researching Writing, Second Edition*, 263. https://doi.org/10.4324/9781315833729

Johansson, R., Wengelin, E., Johansson, V., & Holmqvist, K. (2010). Looking at the keyboard or the monitor: Relationship with text production processes. *Reading and Writing*, *23*(7), 835–851. https://doi.org/10.1007/s11145-009-9189-3

*Kerz, E., Pruneri, F., Wiechmann, D., Qiao, Y., & Strubel, M. (2020). Understanding the dynamics of second language writing through keystroke logging and complexity contours. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 182-188).

Kaufer, D.S., Hayes, J.R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English*, 20, 121- 140. https://doi.org/10.58680/rte198615612

Keuleers, E., & Balota, D. A. (2015). Mega-studies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. In *Quarterly Journal of Experimental Psychology*, 68 (8). https://doi.org/10.1080/17470218.2015.1051065

Kruse, M. (2024). Problem-solving activity during the foreign language writing process: A proposal for categorisation and visualisation of source use and a new take on fluency in multilingual writing. *Journal of Writing Research*, *16*(1), 129-161. https://doi.org/10.17239/jowr-2024.16.01.05

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49 (4), 757-786. https://doi.org/10.1002/ tesq.194

Lafon, P., & Salem, A. (1983). L'inventaire des segments répétés d'un texte. *Mots. Les langages du politique*, *6*(1), 161-177. https://doi.org/10.3406/mots.1983.1101

Lardilleux, A., Yvon, F., & Lepage, Y. (2013). Generalizing sampling-based multilingual alignment. *Machine Translation*, *27*, 1-23. https://doi.org/10.1007/s10590-012-9126-0

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358-392. https://doi.org/10.1177/0741088313491692

*Leijten, M., Waes, L. V., & Horenbeeck, E. V. (2015). Analyzing writing process data: A linguistic perspective. In G. Cislaru (Ed.), *Writing(s) at the crossroads: The process/product interface* (pp. 277– 302). John Benjamins Publishing Company. https://doi.org/10.1075/z.194.14lei

*Leijten, M., Van Horenbeeck, E., & Van Waes, L. (2019). Analysing keystroke logging data from a linguistic perspective. In *Observing writing* (pp. 71-95). Brill. https://doi.org/10.1163/9789004392526_005

Lindgren, E., & Sullivan, K. P. H. (2006). Analysing online revision. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging: Methods and applications* (pp. 157–188). Oxford: Elsevier. https://doi.org/10.1163/9780080460932_010

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45 (1), 36-62. https://doi.org/ 10.5054/tq.2011.240859

* Mahlow, C., Ulasik, M. A., & Tuggener, D. (2024). Extraction of transforming sequences and sentence histories from writing process data: A first step towards linguistic modeling of writing. *Reading and Writing*, *37*, 443–482. https://doi.org/10.1007/s11145-021-10234-6

Mangen, A., & Balsvik, L. (2016). Pen or keyboard in beginning writing instruction? Some perspectives from embodied cognition. *Trends in Neuroscience and Education*, *5*(3), 99–106. https://doi.org/10.1016/j.tine.2016.06.003

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

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication*, 27 (1), 57-86. https://doi.org/10.1177/0741088309351547

Miaschi, A., Brunato, D., & Dell'Orletta, F. (2021). A NLP-based stylometric approach for tracking the evolution of L1 written language competence. *Journal of Writing Research*, *13*(1), 71-105. https://doi.org/10.17239/jowr-2021.13.01.03

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18* (143). https://doi.org/10.1186/s12874-018-0611-x

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin and Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.), *Formulaic language (Volume 2): Acquisition, loss, psychological reality, and functional explanations* (pp. 375-385). John Benjamins. https://doi.org/10.1075/tsl.83.07ohl

Olive, T. (2010). Methods, techniques, and tools for the on-line study of the writing process. In N. Mertens (Ed.), *Writing: processes, tools and techniques* (pp.1-18). Nova Science Publishers, Inc.

*Olive, T., & Cislaru, G. (2015). Linguistic forms at the process-product interface: Analysing the linguistic content of bursts of production crossroads: The process-product interface (pp. 99–123). John Benjamins Publishing Company. https://doi.org/10.1075/z.194.06oli

Olive, T., Kellogg, R. T., & Piolat, A. (2002). The triple task technique for studying the process of writing. In T. Olive & M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp.31-59). Springer.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657–660. https://doi.org/10.1177/1745691612462588

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492–518. doi:10.1093/applin/24.4.492.

Papi, M., Vasylets, O. & Ahmaddian, M. (2022). Individual differences in L2 writing. In M. Papi & S. Li (Eds.), *Handbook of second language acquisition and individual differences* (pp. 381-397). Routledge. https://doi.org/10.4324/9781003270546-30

Paul, N., Lin, T.-J., Ha, S. Y., Chen, J., & Newell, G. E. (2021). The role of achievement goal orientations in the relationships between high school students' anxiety, self-efficacy, and perceived use of revision strategies in argumentative writing. *Journal of Writing Research*, *12*(3), 657-684. https://doi.org/10.17239/jowr-2021.12.03.05

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217– 230. https://doi.org/10.1016/j.jslw.2013.02.003

Planton, S., Longcamp, M., Péran, P., Démonet, J. F., & Jucla, M. (2017). How specialized are writing-specific brain regions? An fMRI study of writing, drawing and oral spelling. *Cortex*, *88*, 66–80. https://doi.org/10.1016/j.cortex.2016.11.018

Révész, A., Michel, M., & Lee, M. (2017). *Investigating IELTS Academic Writing Task 2: Relationships between cognitive writing processes, text quality, and working memory*. IELTS Research Report. informit.840774005843694

Roca de Larios, J., Marнn, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, *51*(3), 497–538. https://doi.org/10.1111/0023-8333.00163

Sala-Bubaré, A., Castelly, M., & Rijlaarsdam, G. (2021). Writing processes as situated regulation processes: A context-based approach to doctoral writing. *Journal of Writing Research*, *13*(1), 1-30. https://doi.org/10.17239/jowr-2021.13.01.01

Schilperoord, J. (2001). 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). Springer. https://doi.org/10.1007/978-94-010-0468-8_4

Strıbel, M., Kerz, E., Wiechmann, D., and Neumann, S. (2016). Cocogen-complexity contour generator: Automatic assessment of linguistic complexity using a sliding window technique. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity* (CL4LC), (pp.23–31).

Stroebe, W. (2016). Are most published social psychological findings false? *Journal of Experimental Social Psychology*, *66*, 134–144. https://doi.org/10.1016/j.jesp.2015.09.017

Sumner, E., Connelly, V., & Barnett, A. L. (2013). Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution. *Reading and Writing*, *26*, 991–1008. http://dx.doi.org/10.1007/s11145-012-9403-6

Tian, Y., Kim, M., & Crossley, S. (2024a). Making sense of L2 written argumentation with keystroke logging. *Journal of Writing Research*, *15*(3), 435-461. https://doi.org/10.17239/jowr-2024.15.03.01

Ulasik, M.A., Mahlow,C., & Piotrowski, M. (2025). Sentence-centric modeling of the writing process. *Journal of Writing Research*, *16*(3), 497-532. https://doi.org/10.17239/jowr-2025.16.03.05

Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting writing process feedback in the classroom: Using keystroke logging data to reflect on writing processes. *Journal of Writing Research.12*(1), 109-140. https://doi.org/10.17239/jowr-2020.12.01.05

Van Drie, J., Van Driel, J., & Van Weijen, D. (2021). Developing students' writing in History: Effects of a teacher-designed domain-specific writing instruction. *Journal of Writing Research*, *13*(2), 201-229. https://doi.org/10.17239/jowr-2021.13.02.01

Van Waes, L., Leijten, M., Roeser, J. Olive, T., & Grabowski, J. (2021). Measuring and assessing typing skills in writing research. *Journal of Writing Research*, *13*(1), 107-153 https://doi.org/10.17239/jowr-2021.13.01.04

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, 829–853. https://doi.org/10.1111/lang.12228

Vasylets, O. (2017). *Task-modality effects: A study of task complexity effects in speech and writing*. Unpublished Doctoral Dissertation, University of Barcelona, Barcelona, Spain.

Vasylets, O., Gilabert, R., & Manchyn, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning*, *67*(2), 394-430. https://doi.org/10.1111/lang.12228

Vasylets, O., & Marнн, J. (2021). The role of working memory and proficiency in second language writing. *Journal of Second Language Writing*, *52*, 100786. https://doi.org/10.1016/j.jslw.2020.100786

Vasylets, O., & Marin, J. (2022). Pen-and-paper versus computer-mediated written modality as a new dimension of task complexity, *Languages*, *7*(3), 195. https://doi.org/10.3390/languages7030195

Vasylets, O., Mellado, M.D., & Plonsky, L. (2022). The role of working memory and language aptitude in digital versus pen-and-paper writing, *Studies in Second Language Learning and Teaching*, *12* (4), 721-743. https://doi.org/10.14746/ssllt.2022.12.4.9

Wengelin, A. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging: Methods and applications* (pp. 107–130). Elsevier https://doi.org/10.1163/9780080460932_008

Wengelin, E., Frid, J., Johansson, R., & Johansson, V. (2019). Combining keystroke logging with other methods: Towards an experimental environment for writing process research. In *Observing writing* (pp. 30-49). Brill.

Wengelin, E., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, *41*(2), 337–351. https://doi.org/10.3758/BRM.41.2.337

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf

Wirtz, J. (2025). Written performance indicators from experienced creative writers. *Journal of Writing Research*, *16*(3), 533-554. https://doi.org/10.17239/jowr-2025.16.03.06

Appendix: Studies which explored linguistic and behavioral alignment in writing

Study	Design	Aim	Participants	Procedure/task	Data analysis	Main results
Olive & Cislaru (2015) book chapter	Pluridisciplinary corpus-analysis study	Explore if bursts & repeated segments share common structural characteristics	*N*=6 L1 French adults (social workers)	Reports on children at risk	*Morphosyntactic & semantic description of bursts & repeated segments *Pause analysis *Inputlog (Leijten & van Waes, 2013) *NLP tool (Lardilleux et al. 2013)	*Less than 3% of bursts & repeated segments converged *The longest bursts consisted of complete syntactic structures, but they were rare *High proportion of noun-phrases & syntactically unsatured strings in both bursts & repeated segments
Leijten et al. (2015) book chapter	Explorative quantitative case study	Explore the potential of keystroke & linguistic analysis in the investigation of cognitive characteristics of people with Alzheimer's disease	*N*= 2 L1 Dutch: healthy female (81 y.old), female with dementia (79 y. old)	*N*=2 picture description tasks in the lab	*Pause analysis *Revisions *Product/process ratio *Linguistic text analysis *Inputlog	Participant with dementia: longer pauses and more pausing overall; more pauses within words; longer pauses at production of nouns, verbs and pronouns; rare use of adjectives; less regular use of connectives; word-by-word production

Goodkind & Rosenberg (2015) conference proceedings	Experimental study	Explore the relationship between multiword untis & pauses as mediated by cognitive task complexity	*N*=189 L1 English students (42% female)	*N*=10-12 writing tasks responding to prompts of varying cognitive load; no time limit; in the lab	*Pause analysis *Automatic tagging of multiword units *OpenNLP (Baldridge, 2002) *QUERTY keyboard	*Pauses within an multiword unit can vary significantly, depending upon the cognitive demands of the task within which they were produced
Rйvйsz et al. (2017) research report	Experimental study	*Relate L2 test-takers`cognitive processes & online behaviours to the quality of the texts *Explore working memory effects	*N*=30 L1 Chinese university learners of L2 English	IELST Academic Writing task in a lab	*Speed fluency, pauses, revisions, bursts *Lexical, syntactic, discourse complexity *Inputlog *Automated software for linguistic analysis *Eye-tracking, stimulated recall	*More frequent pausing predicted lower IELTS scores & less sophisticated lexis *More frequent pausing between sentences predicted lower syntactic complexity (greater strucutral similiraity & shorter t-units) *More higher-order revisions predicted more sophisticated lexis.
Medimorec & Risko (2017) journal article	Experimental study	Explore the relation between pauses & lexical textual features while controlling for	*N* =101 university students; fluent English speakers	Narrative & argumentative tasks; timed writing (50 min)	*Pause analysis *Lexical sophistication, lexical diversity, sentence length *Inputlog *Coh-metrix (Graesser et al.	*Increased pause rates at word & sentence boundaries related to decreased word frequency & increased sentence length

		transcription fluency and genre			2004)	respectively *More pausing in argumentative texts
Bowen (2019) journal article	Explorative quantitative case study	Examine linguistic content of dynamically unfolding revisions	*N*=2 L1 English university students (females; age 19)	*N*=4 argumentative essays	*Revisions *Thematic text analysis *Linguistic text analysis (clause, phrase, word, morpheme) *Inputlog	Noun group was the most commonly revised structure
Leijten et al. (2019) book chapter	Quasi-experimental exploratory study	Explore the possibilities of linguistic analysis of keystroke logging data in L1 & L2 writing	*N*=18 L1 Dutch master students mean age 22; B2 level of L2 English	Knowledge-telling task in L1 Dutch & L2 English; 2 min for planning, timed writing in the lab with Inputlog	*Pause analysis *S-notation *Revisions *Number & length of bursts *Fluency measures *Part of speech tagging, word length & frequency * Inputlog	*Longer *p*-bursts, longer texts & more fluent production in L1 Dutch * Longest pauses before nouns & conjunctions; shorter before adjectives *Higher variation in pause length before different word categories in L2 English

Kerz et al. (2020) conference proceedings	Corpus-analysis study	Relate behavioral writing data to indices of syntactic & lexical complexity of L2 texts	*N*= 512 L1 German university students; upper intermediate-advanced L2 English learners	*N*=3,454 texts of reflective writing (learning journals)	*Fluency measures *Pauses *Revisions *Syntactic complexity (e.g.,clauses per sentence), lexical diversity & sophistications (syntactic complexity) *Etherpad (www.etherpad.org) *Sliding window approach (Strυbel et al. 2016)	*Higher syntactic complexity in more fluent L2 writing production *Sentence production time affected by syntactic complexity & lexical diversity *Increased cognitive effort associated with higher lingьнstic complexity and drp`s in fluency
Mahlow et al. (2024) journal article	Small-scale exploratative study	Linguistic modeling of writing process	*N*=7 subjects (4 male, age range 34-62), L1 German, academic background	*Retelling task based on 2-minute video, in L1 German (writing time: approx. 10 minutes; mean text length: 196 words)	*Text history *Sentence history *Combined raw keystroke logging data with NLP tools to supplement the analysis with morphosyntactic & syntactic annotations *Text History Extraction Tool (THEtool) *Scriptlog (Strυmqvist & Karlsson, 2002)	*Complex pattern of revision behaviors (e.g., writers can switch production mode even mid-word, do not necessarily choose complete syntactic constituents as the targets for revision, etc.) *Modeling of written text production from the perspective of linguistic structures should be done beyond the word level