**Managing Academic Writing Conventions in French Masters Short Research Papers: Insights from Writing Process Data**

Georgeta Cislaru, Université Sorbonne Nouvelle & Institut Universitaire de France (georgeta.cislaru@sorbonne-nouvelle.fr)

Maya Sfeir, American University of Beirut & Clesthia Lab (Université Sorbonne Nouvelle) (ms213@aub.edu.lb)

It is now accepted that genre constraints shape language practices and production. Certain discourse genres, such as academic writing, are particularly subject to generic constraints, leading to the identification of several levels of **routines** in numerous works over the last 40 years (Bazerman, 1988; Hyland, 2000; Miller, 1984; Swales, 1990; Tardy, 2009). At the conceptual level, academic texts are also discourses that develop explicit dialogism, through cross-references and quotations. Mastering academic writing obviously requires a specific apprenticeship, but we may well ask how the presence of different routines, on the one hand, and the integration of other discourses, on the other, intervene in the production process. Do the former make writing more fluid, because they can potentially be automated? Do the latter constitute a stumbling block, because they integrate heterogeneity in a conventional way?

**A Process- and Product-Based Approach to Writing**

To address these questions, we present the early stages of our analysis of how French graduate students engage with academic conventions *in the process* *of* writing. Our research consequently combines a process-oriented approach, examining how writing is produced in real time, with a product-oriented approach, analyzing the textual drafts produced by writers (Bowen & Van Waes, 2020). Our project is thus largely grounded in the relatively new field of **writing process research**, which has been enhanced by new technologies/methods that facilitate the recording of writing in real time —think-aloud protocols, video recordings, and keystroke logging (Miller, 2006; see studies by Kaufer et al., 1986; Leijten & Van Waes, 2006, 2013; Matsuhashi, 1981). Our research mainly builds on **keystroke logging**, which we briefly define here as a method that records typing activity and mouse movements and clicks, and makes the captured data available for further analysis in the form of logging files (Leijten & Van Waes, 2013). While keystroke logging as a method is greatly used by cognitive scientists and psycholinguists (Lindgren et al., 2019), in our work, we integrate it into a more linguistic-discursive framework.

***Behavioral Observables***

Central to our linguistic-discursive framework from the process perspective are two notions that constitute behavioral units, that is activities or actions that writers perform during the writing process. The first notion, which is event-based (an observable salient action), is the act of **revision**. Revision reflects the complexity of writing and can be defined as “making any changes at any point in the writing process” (Fitzgeralds, 1987, p. 484). A number of taxonomies have been developed for analyzing revisions, taking into consideration the nature of the revision, its occurrence during the writing process, and its location in the text (Chanquoy, 2001; Faigley and Witte, 1981; Lindgren & Sullivan, 2006). The second notion is more formal (a measurable production activity) and consist of **bursts** **of written production**, often referred to as *bursts* for short. It stems from the empirical observation that writing production is made up of alternating periods of production and pauses, sharing writing time almost equally. A burst, as Chenoweth and Hayes (2003) defined it, “is a group of words produced by a writer for inclusion in a text that is bounded by breaks in the production process” (p. 103). In other words, bursts are strings of text produced within two **pauses**. In writing research, the common (though arbitrary and highly debatable) pause threshold is of a duration of two or more seconds (Van Waes & Leijten, 2015). This threshold is considered likely to eliminate purely mechanical pauses, leaving only those with cognitive value, contributing to planning, lexical selection, and revision. Pauses thus can provide powerful clues about cognitive processes of text production (Schilperoord, 1996, 2002). Hence, our project is partially informed by methods and principles for data collection and analysis from the field of (cognitive) writing process research.

***Linguistic Observables***

As mentioned above, the other dimension of our project involves **academic writing conventions.** Our understanding of these conventions is informed by the work of discourse studies theorists and scholars from different traditions, such as for instance, dialogism, genre analysis, and French and Anglo-Saxon discourse analysis. According to Bakhtin (1986/2010), when emerging in a specific social sphere, a genre generates its own discursive routines. In line with French discourse analyst Dominique Maingueneau’s (2002) categorization, academic writing corresponds to instituted genres, with a strong imposition of structural and linguistic formats.One of the particularities of academic writing is the high proportion of “déjà-là” (*already there*) (see Plane & Rondelli, 2017), or to say it another way, an important amount of writing is based on the retrieval of **reproducible sequences** from memory or other sources. For the purposes of our research, we retain two categories of reproducible sequences: on the one hand, genre-specific formulaic language (Wray, 2002), widely exploited to structure text sequences (Swales, 1990), and on the other hand, references and citations (Boch & Grossmann, 2001; Lucas, 2004) used to provide evidence or to appeal to authority.

Integrating one’s voice and opinion with the voice and opinions of others textually (Shi, 2004) is a genre characteristic of academic writing that implies mastering not only typographical marks like quotation marks, but also reformulation strategies, lexico-syntactic adjustments, etc. Writers often borrow “vocabulary, sentence structures, writing styles, organizational patterns, ideas and information” (Leki & Carson, 1997, p. 51) from source texts and integrate them into their own texts. A number of previous studies have examined how students integrate borrowed sequences (quotations, paraphrase, and summaries) into their writing, with a majority of studies focusing in copying and patchwork writing for L1 and L2 writing (Campbell, 1990; Pecorari, 2004; Shi, 2004) and on issues of text ownership (Pennycook, 1996) and intertextuality (Chandrasoma et al., 2004).

However, the focus of our study is not the nature of borrowed sequences (although we take them into account), but the strategies that writers use to integrate them within the writing process and the eventual difficulties they encounter through the writing process in doing so. Borrowed sequences should be considered as part of the dialogical discourse crafting that writers engage in when they write (Bakhtin, 1981). To further examine this dialogical discourse crafting, we draw on the work of a number of French discourse analysts (cf. Authier-Revuz, 1984; Bres, 2005; Ducrot, 1984) who emphasized the dialogic nature of discourse and the interplay between the speaker’s voice and the voice of Other during communication. Authier-Revuz’s (1984) notion of enunciative heterogeneity is particularly relevant for our analysis, as will be briefly discussed in our data analysis section. Drawing on Bakhtinian dialogism and Lacan’s conception of the subject of language, Authier-Revuz (1984/2014) has argued that constitutively, in the subject, in their discourse, there is the Other, whereby discourse is inherently heterogeneous. From a linguistic point of view, this raises the question of how to manage language heterogeneities when producing a routine discursive genre (see Grossmann 2022 on the use of specific markers like *selon*=according to).

**Formulaic language** or repeated sequences—also referred to as n-grams, multiword expressions, lexical bundles, phraseological units (for example, *on the other hand, on the one hand,* etc.)—contribute various rhetorical purposes (Hyland 2008a, 2008b) in academic writing. Lexical bundles are one type of repeated sequences and can be defined as “recurring sequences of three or more words . . . that commonly go together in natural discourse” (Biber et al., 1999, p. 990). These are often identified through a computer program, are characterized by frequency, and differ from idioms and fixed lexical expressions. Overall, lexical bundles have been described as the building blocks of discourse. Lexical bundles canalso“signal competent participation in a given community of users” (Hyland, 2008a, p. 42). They serve as indicators of writers’ acculturation into their discourse community (Cortes, 2006).

While a substantial body of research on lexical bundles and multiword expressions has been conducted on English, studies on French have been carried out by Tutin (2010) (also see Legallois & Tutin, 2013; Tutin & Falaise, 2013). More recently, a number of studies have begun examining multi-word units at the product-process interface. Gilquin (2024), for instance, has examined the processing of multiword units in L2 English data and developed a schematic that takes into consideration pause placement in bursts. To do so, she extracted multi-word units from students’ final texts and then examined how they were produced as bursts in the process data. The methodology developed by Gilquin (2024) is important for our study, as we aim to examine borrowed sequences in the form of multi-word expressions in the final texts and then examine their production in real time in the process data.

In light of the two definitions of “déjà-là” or reproducible sequences presented above, our project seeks to answer the following two questions that reflect the two dimensions of sequences we have outlined above and that are linked to the broader questions we presented in our introduction: What strategies are used by French Master students in the process of drafting and revising short research projects to manage structural and dialogic academic conventions? To which extent is there a clear dividing line between more or less personal formulations and borrowed sequences?

**Methods and Data**

Our analysis focuses primarily on empirical data. Our data were recorded using two keystroke-logging software programs: Inputlog (Leijten & Van Waes, 2013) and Scriptlog (Strömqvist & Karlsson, 2002). The data is part of a larger initiative, the Pro-Text (Processus de Textualisation) Corpus, a corpus of final texts and keystroke logs written in French in different genres (analysis papers, narratives, social reports, and translations), by different writers (university students, primary and secondary school students, dyslexic students, social workers, and translation students), and in different situations (natural and experimental) (N = 422, tokens=206,665). The global corpus consists of six-sub-corpora (Academic, Adult, Children, Dyslexia, Professional, and Translation) and was partially enriched with three types of annotation: (1) part-of-speech, lemma, and syntactic information; (2) burst annotation, whereby bursts were categorized into three types following Cislaru et al. (2024): P-bursts, or bursts of writing production; RB-bursts, or immediate revision bursts; and R-bursts or revision bursts (see glossary); and (3) rewriting operations like adding, deleting, replacing, and displacing linguistic contents (Miletić et al., 2022). Both product data in the form of final texts and process data in the form of keystroke logs are available for analysis.

For the present study we exploit the Academic Sub-corpus (thereafter Academic). This sub-corpus consists of 26 long papers written by first-year Master’s students enrolled in an introductory course on discourse analysis at the Université Sorbonne Nouvelle (Paris 3). The assignment prompt required students to pastiche a text and analyze the implied processes. The students composed the texts in a nearly-natural situation. Inputlog or Scriptlog software were used directly on their personal laptops, over as many writing sessions as necessary for the completion of the paper. Data was collected in [2012-2017], prior the COVID pandemic, and prior to the widespread adoption of Generative AI (GenAI) tools like ChatGPT as writing aids.

**Data Analysis**

Taking into account the data collection process, whereby students were asked to copy-paste their previous drafts into Word after each writing session in order to secure their production, we expect to find some inconsistencies in the process data. To address this issue, we will select a handful of essays for a fine-grained analysis. The analysis will be based on the qualitative identification of academic-genre prefabs and borrowed sequences in the final texts and their retro-analysis throughout the writing process in the keystroke logging files. For this last point, we investigate whether these sequences were produced in one go, as bursts of writing (Chenoweth & Hayes 2001, 2003), or whether they were subject to revisions and, if so, of what type: immediate or post-editing, reformulations or deletions, displacements, additions.

Below an example of writing and rewriting operations aiming at formulating the pragmatic purpose of the analyzed discourse genre. Each line corresponds to a burst of writing (we added the mention [pause] at the end of each). The bursts are listed in chronological order of their production. Deletions of immediately or previously produced sequences and insertion on the left side of new sequences show the spatial non-linearity of the production, in opposition to the temporal linearity. This spatio-temporal data enables us to observe how formulaic sequences are updated during the writing process.

***Example 1***:

[pause] Ce genre a donc une visée [pause]

pratique [pause]

{*deleting ‘a donc une visée’*}

vise donc une finalité [pause]

car son but e [pause]

{*deleting ‘n but e’*}

n but est [pause]

{*inserted in the continuity of the second burst : c'est-à-dire la vente de l'objet sujet de l'annonce.*}

***Or, represented more schematically:***

Ce genre ~~a donc une visée~~ [pause] pratique [pause] 

 vise donc une finalité [pause]

car ~~son but e~~[pause] n but est [pause]

 c'est-à-dire la vente de l'objet sujet de l'annonce.

Another example below highlights the separation of enunciatively heterogeneous sequences by pauses: the student produces the reference to the quoted author, opens the quotation marks, then pauses before producing the quotation extract - the pause could be considered to be due to verification of the source. The student then corrects the last word of the quoted sequence. They proceed in the same way with the second quoted author, pausing after the introductory verb “prefers.” The student hesitates over this lexical choice, deletes it, and then reproduces the same verb with the quoted element of speech.

***Example 2***:

JM Abram penche pour «  [pause]

prescriptif ou in~~jonctif » et~~ [pause]

 structif », et [pause]

Longacre lui ~~préfère~~ [pause]

 préfère le terme "procédural"\*. [pause]

**Implications for Writing Studies Research**

By relying on empirical data in the form of product (final texts) and process data (keystroke logging), our project will, on a broad level, provide insight into how French writers manage academic writing conventions at a micro-level in real time. We expect to gain insights into how writers structure their texts, engage with other texts, and integrate other voices in their writing *during the process* of drafting. The data will also provide insight into the revision practices of these writers, as we expect to find a lot of individual differences across our data. Overall, our project can lead to pedagogical application for the teaching of academic writing in France and elsewhere.

**Institutional Description**

Université Sorbonne Nouvelle is a French public university specializing in the fields of Letters, Languages, Culture and the Humanities, with over 15,000 students (including some 5,000 international students). The university is situated in Paris and is reputed to be one of the best in its field, with a strong research dimension. The university has earned distinction for its focus on written discourse analysis. Teaching is conducted in French, apart from the language faculties.

All students benefit from at least one course in academic methodology in the first year of study, and a course in research methodology, including modules on academic writing, in the first year of the Master’s program. Over the last fifteen years, the question of writing and literacy has become increasingly important, both in terms of training and research. The collection of data corresponds with the beginning of this dynamic. It is also linked to one of the authors’ publication of a guide on academic writing emanating from the Master’s course in academic writing (Cislaru et al., 2021 [2009]).

The data were collected in an introductory seminar on discourse analysis that students take “by choice”, i.e., they are given a list of seminars from which they are required to choose a certain number of courses to take, according to their specialty. In the field of Language Sciences, where the data was collected, the university has the largest number of specialists and students in France. Data collection conditions were totally unprecedented. To put the students at ease, they received training in keylogging software, with modules at the start of each seminar. The primary objective was to obtain ecological recorded writing data, and more specifically multi-session data, enabling us to observe the complexity of the writing process beyond the modelling of a 20-30 line text (most often encountered in work on recorded writing). The second objective, linked to the first, was to observe the process of putting text into writing, while taking into account the various constraints that make academic writing so specific.

**Key Theorists and Theoretical Groundings**

**Theory of enunciation:** Although enunciation is little known outside of France and Latin America (Angermuller et al., 2014), it remains a fundamental concept for French discourse analysis. The concept of enunciation can be traced back to French linguist Émile Benvensite (1970), who defines it as putting language to work through an individual act of use. This definition places the subject at the center of language practices, and questions both the construction of the subject through the means of language, and the trace left by her/his use of certain forms (like deictics, modalities, reported speech markers). Scholars such as Ducrot (1984) and Authier-Revuz (1984) have highlighted the polyphonous nature of enunciation; this approach provides a wide range of linguistic tools to cope with dialogism.

**Dialogism:** Our work falls within the perimeter of the Bakhtinian theory of dialogism (Bakhtine, 1977), taking for granted that in any discourse there is no exclusive “I speak” at all - it always speaks elsewhere and before. In the French-speaking world, this principle has been considered from the angle of enunciative heterogeneity, and enriched by special attention to metalinguistic uses (Authier-Revuz, 1984, 2014, 2020). This dimension has been widely applied to media, literary, and didactic discourse, but less so to French-language academic discourse (Boch & Grossmann, 2001; Boch & Rinck, 2010; Rinck & Boch, 2012; Rinck, 2013; Steriu, 2021; Steriu & Vlad, 202).

**Prefabs**: For several decades, it has been widely recognized in the linguistic literature that almost half of language communication is made up of prefabricated sequences, which are re-actualized as they are in use and are not the result of ad hoc combinatorics (Mel’čuk, 2003; Sinclair, 1991; Wray, 2002). Approaches based on corpus analysis and generic specificities have led to the identification of sequences according to their capacity to preserve an empty space to accommodate a diversity of forms, like the lexical bundles of Biber et al. (1999), or their recurrence in a corpus identified via textometric analysis, such as repeated segments (Lafon & Salem, 1983). One of the first authors to observe the use of this type of sequence in scientific articles was Fonagy (1982). However, already in 1974, the writer Georges Perec pastiched a scientific article, first in English then in French, relying on the exploitation of linguistic clichés.

**Constraints in academic writing:** The identification and linguistic characterization of moves in academic writing has given rise to the development of an entire approach to teaching these writing practices (Swales, 1990). Research targeting prefabs, and more particularly lexical bundles and their exploitation in the different discursive sequences of academic writings (moves), was subsequently developed, whether for English (Hyland 2008) or for French (Tran, 2014).

**Writing process research**: Research on the writing process falls more generally within the framework of the writing model of Hayes and Flower (1980) and Flower and Hayes (1980), which has a strong communicative anchor, and integrates several process modules such as translation (of thought into language forms), planning, revision. Kellogg’s (1994) psycho-cognitive model is also commonly used. Linguistic approaches to the process are relatively recent (around twenty years), with the notable exception of work in genetic criticism which has studied manuscripts since the 1970s. A linguistic study of written production units, bursts, was proposed in 2018 by Cislaru and Olive. Authors like Gilquin (2020) have been interested in prefabs and their behavior/actualization throughout the writing process.

**Glossary**

**Conventions (discourse, writing):** Conventions are defined at the socio-communicative level (they may, for example, be part of cultural norms or more institutionalized norms), and constitute more or less tacit agreements on language practices, discourse structures, etc., that are binding on all parties in a linguistic community.

**Constraints (writing):** Constraints can be rooted in conventions, but can also refer to specific (contextual, material) conditions of language production (see Cislaru, 2014). In writing, for example, material constraints relate to the medium used (handwritten or typescript), the spatial distribution of the text and the possibilities for restructuring in line with language constraints (lexical, syntactic and pragmatic rules, for example), motor and typing skills. Cognitive constraints depend on the brain’s ability to process information and to retrieve from working or long-term memory the forms and content needed to configure the discourse.

**Routines (discourse, writing):** Routines are procedural savoir-faire (Langacker, 1987) consisting in the ability to choose forms and formats corresponding to specific speech acts or textual organization patterns (Legallois, 2022). Thus, reproducible sequences such as prefabricated phrases like *our objective is to* and patterns for introducing cross-references or quotations such as *according to* + are part of the routines of academic discourse.

**Keystroke logging:** Keystroke logging is a research method that enables the observation of digital writing processes (Leijten & Van Waes, 2013). Keystroke logging programs record (log) and time stamp keyboard and mouse activity so as to enable the reconstruction, description, and analysis of the text production process. The data is often available for analysis in the form of keystroke log files. Inputlog (Leijten & Van Waes, 2013) and Scriptlog (Strömqvist & Karlsson, 2002) are two keystroke logging programs.

**Bursts of writing production:** The term *bursts of written production* or simply *bursts* can be traced back to Chenoweth and Hayes (2001, 2003). A burst, as Chenoweth and Hayes (2003) defined it, “is a group of words produced by a writer for inclusion in a text that is bounded by breaks in the production process” (p. 103). In other words, bursts are the building blocks of texts (Kaufer et al., 1986).

**Pauses and revision events:** Pauses and revision events are central to writing process research as they offer important behavioral, linguistic, and cognitive clues about the writing process, such as planning and decision-making (Van Waes & Leijten, 2015). The duration and location of pauses, and the nature of revisions and the point in the writing process during which they happen, among other criteria, have been the object of analysis (Lindgren & Sullivan, 2006; Van Waes & Leijten, 2015).

**Acknowledgements**: This research was funded by the French National Research Agency (ANR), N° ANR-18-CE23-0024-01 “Pro-TEXT”.

**References**

Angermuller, J., Maingueneau, D., & Wodak, R. (Eds.). (2014). *The discourse studies reader: Main current in theory and analysis*. John Benjamins.

Authier-Revuz, J. (1984). Hétérogénéité(s) énonciative(s). *Langages*, *19*(73), 98-111. Doi: 10.3406/lgge.1984.1167

Authier-Revuz, J. (2014). Enunciative heterogeneity. In J. Angermuller, D. Maingueneau, & R. Wodak (Eds.), The discourse studies reader: Main current in theory and analysis (pp. 155-165). Netherlands: John Benjamins. (Original work published 1984).

Authier-Revuz, J. (2020). *La représentation du discours autre: Principes pour une description*. De Gruyter.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays*. University of Texas Press.

Bakhtin, M. M. (2010). *Speech genres and other late essays.* University of Texas press. (Original work published 1986)

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. University of Wisconsin Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Pearson Education Limited.

Boch, F., Grossmann, F. (2001). De l’usage des citations dans le discours théorique : des constats aux propositions didactiques. *LIDIL - Revue de linguistique et de didactique des langues* 24, 91-112.

Boch, F., Rinck, F. (2010). Pour une approche énonciative de l'écrit scientifique. Introduction. *LIDIL - Revue de linguistique et de didactique des langues* 41, 5-14.

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. DOI: 10.1177/0741088320916508

Bres, J. 2005. Savoir de quoi on parle : dialogue, dialogal, dialogique ; dialogisme, polyphonie…. In J. Bres et al. (Eds.), *Dialogisme, polyphonie: Approches linguistiques* (pp. 47-62). De Boeck.

Chandrasoma, R., Thompson, C., & Pennycook, A. (2004). Beyond plagiarism: Transgressive and nontransgressive intertextuality. *Journal of Language, Identity & Education, 3*(3), 171–193. https://doi.org/10.1207/s15327701jlie0303\_1

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*, 15–41.

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/10.1177/0741088301018001](https://doi.org/10.1177/0741088301018001004)

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, *20*(1), 99-118.

Cislaru, G. (2014). Contraintes linguistiques et contextuelles dans la production écrite. *Les Carnets du Cediscor* 12, 55-74.

Cislaru, G., Claudel, C., & Vlad, M. (2021). L'écrit universitaire en pratique. De Boeck Supérieur.

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.

Cislaru, G., & Olive, T. (2018). *Le processus de textualisation: Analyse des unités linguistiques de performance écrite*. De Boeck.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*, 397-423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17(4), 391–406.

Ducrot, O. (1984). *Le dire et le dit.* Les Éditions de Minuit.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*, 400–414.

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

Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing : Making plans and juggling constraints. In L. W. Gregg, & E. R. Steinberg (Eds). *Cognitive processes in writing* (pp. 31-50). Lawrence Erlbaum.

Fónagy, I. (1982). *Situation et signification*. John Benjamins.

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*(51), 1-22. Doi:<https://doi.org/10.3390/languages9020051>

Grossmann, F. (2022). Marquage évidentiel et autres fonctions discursives : le cas des routines initiées par *selon, suivant* et *d’après*. In M. Luodonpää, F. Grossmann, & A. Tutin (éds) *Les routines discursives dans le discours scientifique oral et écrit* (95-117). UGA Editions.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Lawrence Erlbaum.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.

Hyland, K. (2008a). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21.

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

Kellogg, R. T. (1994). *The psychology of writing*. Oxford University Press.

Lafon, P., & Salem, A. (1983). L’inventaire des segments répétés d’un texte. *Mots,* *6*, 161-177. Doi: 10.3406/mots.1983.1101

Langacker, R. (1987). *Foundations of Cognitive Grammar (Vol. I). Theoretical Prerequisites*. Stanford UP.

Legallois, D. (2022). Routines, routinisation : à quels unités et processus a-t-on affaire exactement ? In M. Luodonpää, F. Grossmann, & A. Tutin (éds) *Les routines discursives dans le discours scientifique oral et écrit* (19-35). UGA Editions.

Legallois, D., & Tutin, A. (2013). Présentation: Vers une extension du domaine de la phraséologie. *Langages*, *189*(1), 3-25.

Leijten, M., & Van Waes, L. (2006). Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment. In K.P H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging and writing* (pp. 73-93). Brill.

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>

Leki, I., & Carson, J. (1997). “Completely different worlds”: EAP and the writing experiences of ESL students in university courses. *TESOL quarterly*, *31*(1), 39-69.

Lindgren, E., & Sullivan K.P.H. (2006). Analyzing on-line revision. In K.P.H. Sullivan, & E. Lindgren. (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 157–188). Elsevier.

Lindgren, E., Knospe, Y., & Sullivan, K. P. (2019). Researching writing with observational logging tools from 2006 to the present. In E. Lindgren & K. P. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 1-29). Brill.

Lucas, N. (2004). La citation et l'appel à référence bibliographique dans les articles académiques In J. M. López-Muñoz S. Marnette, L. Rosier, (eds.) *Le discours rapporté dans tous ses états : Question de frontières* (pp. 419-427). Paris, l’Harmattan.

Maingueneau, Dominique. (2002). Analysis of an academic genre. *Discourse Studies* 4: 319 - 341.

Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English 15*(2), 113-134.<https://www.jstor.org/stable/40170920>

Mel’čuk, I. (2003). Collocations, définition, rôle et utilité. In F. Grossmann & A. Tutin (Eds), *Les collocations : Analyse et traitement* (pp. 23-32). Amsterdam.

Miletić, A., Benzitoun, C., Cislaru, G., & Herrera-Yanez, S. (2022, June). Pro-text: An annotated corpus of keystroke logs. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 1732-1739.

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing, 12*(4), 317–345.<https://doi.org/10.1016/j.jslw.2003.08.004>

Pennycook, A. (1996). Borrowing others’ words: Text, ownership, memory, and plagiarism. *TESOL Quarterly, 30*(2), 201–230.<https://doi.org/10.2307/3588141>

Plane, S., & Rondelli, F. “Le déjà-là dans l’écriture: Quel substrat pour quels (ré)emplois?.” *Pratiques* [En ligne] *173-174*. DOI: https://doi.org/10.4000/pratiques.3254

Rinck, F. (2013). Comment gérer les sources dans un article de recherche? Un scénario didactique pour les apprentis-chercheurs. C. Desoutter; C. Mellet. *Le discours rapporté : approches linguistiques et perspectives didactiques*, Peter Lang, 181-196, 2013.

Rinck, F., Boch, F. (2012). Enunciative Strategies and Expertise Levels in Academic Writing: How do Writers Manage Point of View and Sources?. M. Castello, C. Donahue. *University Writing: Selves and Texts in Academic Societies*, Emerald Group Publishing, 111-128.

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Rodopi.

Schilperoord, J. (2002). On the cognitive status of pauses in discourse production. *Contemporary tools and techniques for studying writing*, *10,* 61-87.

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication, 21*(2), 171–200.<https://doi.org/10.1177/0741088303262846>

Sinclair, J. (1991). *Corpus, concordance, and collocation*. Oxford University Press.

Steriu, L. (2021). Les modes d'articulation entre discours d'autrui et discours propre dans l'écriture du mémoire de master, Universitatea Ovidius, Constanta, Romania.

Steriu, L., & Vlad, M. (2024). La reprise du discours d’autrui dans l’écriture du mémoire de master en français langue étrangère. *Action didactique*, *6*(2), 71-90. https://asjp.cerist.dz/en/article/239581

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge UniversityPress.

Strömqvist, S., & Karlsson, H. (2002). *ScriptLog for Windows: User’s manual. Technical Report*. University of Lund: Department of Linguistics, and University College of Stavanger: Centre for Reading Research.

Tardy, C. M. (2009). *Building genre knowledge*. Parlor Press.

Tran, T. T. H. (2014). *Description de la phraséologie transdisciplinaire des écrits scientifiques et réflexions didactiques pour l'enseignement à des étudiants non-natifs: application aux marqueurs discursifs* (Doctoral dissertation, Université de Grenoble).

Tutin, A. (2010). Showing phraseology in context: Onomasiological access to lexico-grammatical patterns in corpora of French scientific writings. In S. Granger & M. Paquot (2010), *Lexicography in the 21st Century: New Applications, New Challenges* (pp. 313-324)*.* Presses Universitaires de Louvain.

Tutin, A., & Falaise, A. (2013). Multiword expressions in scientific discourse: A corpus-driven database. *Actes de eLex 2013*.

Van Waes, L. & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition* *38*, 79-95.<https://doi.org/10.1016/j.compcom.2015.09.012>

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge University Press.