Writing(s) at the Crossroads

The process-product interface

Edited by Georgeta Cislaru

Writing(s) at the Crossroads

Writing(s) at the Crossroads

The process-product interface

Edited by

Georgeta Cislaru

CLESTHIA, Université Sorbonne nouvelle Paris 3

Table of contents

Writing(s) at the crossroads

The process-product interface

Georgeta Cislaru
CLESTHIA, Université Sorbonne nouvelle Paris 3

Writing as a process and writing as a product are sometimes clearly distinguished, and sometimes confused. The two are studied in various domains: cognitive psychology, textual genetics, and anthropology are mainly – although not exclusively – interested in writing as a process, which they analyze from different perspectives; discourse analysis or text analysis, as well as corpus linguistics, are concerned with the product, i.e. the written text, or discourse, and its description or interpretation. One possible approach to grasping the process-product interface may be to confront disciplines and mix methodologies in order to determine to what extent the data that are provided by the study of the process are relevant to linguistic and interpretive approaches that focus on the functioning of the text as a product, and vice-versa.

Cross-disciplinary issues have been formulated in recent years from various theoretical standpoints (see Bazerman & Prior 2004; Torrance et al. 2012, for instance), and a number of works have examined the relationship between process and product, or attempted to establish clear connections between them. The pragmatics of writing developed in Perrin (2003a; 2013) is a perfect illustration of such a complex approach to writing. Perrin connects the explicit or supposed communicative intentions of the writer(s) with the writing strategies observed *in situ*:

> The central question of the pragmatics of writing […] is: what do people *want* to do when they write – and what do they *actually* do? Research attempting to explain natural, everyday writing tends to treat the writer as socially-bound, the written product as intertextually-bound, and the writing process as procedurally-bound. (Perrin 2003b, 825)

Although many researchers have formulated the question of the process-product interface, not enough efforts have been made to bring together tools and methodologies in order to promote constant dialogue between the different domains and formulate a twofold approach to the problem. The process-product interface may be analyzed through the language phenomena emerging both from the study of the process and from the study of the texts that are ultimately

produced. While different methods are applied to these phenomena by different disciplines, to what extent can they be considered to be working on the same objects? Textual genetics already assumes porosity between the concepts and methods of the study of process and product (Fuchs, Gresillon, Lebrave, Peytard & Rey-Debove 2003; Doquet-Lacoste 2004; Fenoglio & Adam 2009). It would be fruitful to observe the articulations around linguistic features at the crossroads of psycholinguistics and text or discourse analysis. How are description and interpretation (explanation) correlated in the different domains that concern themselves with the study of writing, and how are they opposed? This volume takes a linguistic perspective to writing against an interdisciplinary background, and seeks to highlight more than one point of articulation between writing as a process and writing as a product. The context of writing, the text, and the modules and complexity of the writing process itself are put under scrutiny in order to confront and blend methods, postulates, units of analysis, etc.

1.   At the crossroads between process and product

Writing is a communicative activity with its own specificities, associated to the media involved and to the particular temporality of the process. These specificities have been looked at by various studies from different perspectives. Strömqvist et al. (2006) emphasize the difference between speaking and writing, beyond the stylistic point of view. Both writing and interpreting a written text, they argued, require more empathy and imagination than the production-reception of oral texts (Brandt 1990; Strömqvist et al. 2006, 53). Kesseling (1992) examines the function of pauses in oral and written discourse, and tries to adapt the concept of prosodic structure used with oral productions to pauses during the writing process. Biber et al. (2004) compare linguistic data, and more specifically lexical bundles and idiomatic constructions, in oral and written texts. These specificities of written communication at different levels demonstrate the need to take an approach that is particularly tailored to the process and product of writing.

It was some time ago that the need for pedagogical applications of research on writing triggered the development of a two-sided conception of writing, as attested by numerous publications and methods in the hybrid teaching of writing, such as Donald D. Murray's "Teach Writing as a Process Not Product" (1972) or Gehle and Rollo (1977), which take the same line. Their companions to the writing process offer various linguistic clues to the relationship(s) between the text and the process of its composition.

Two fundamental questions seem to underpin the process-product articulation. The first is about interpretation, where the writing process is ascribed

an explanatory potential. The second is about constraints – which also have explanatory power, although their main role seems to be in helping to grasp the complexity of the activity of writing through its temporal, textual, memorial, visual, neurocognitive, etc., dimensions (cf. Berninger & Richards 2002; Olive et al. 2008; Olive & Passerault 2012; Plane et al. 2010). The study of the constraints that affect writing connects linguistics and psychology, in that mental activity is confronted with both the linguistic materials that are drawn on in the act of writing and the textuality of writing as a product. Sociolinguistics and anthropology are also involved, as shown in the chapter below on socio-anthropological approaches to writing (§ 4).

In terms of textuality, linguistic material, and pragmatic issues, the study of writing as a product is not separated from reflection on the process, even though the way the text is shaped more often than not remains "in the black box." The impact of social, cultural and institutional habits during the process of shaping and interpreting the text is detailed by Candlin and Hyland (1999), who studied writing practices from an interdisciplinary standpoint. The study of writing practices may also be considered as a cue to a better comprehension of texts (Bazerman & Prior 2004). Interpretive needs pushed Bhatia (1993) to propose a triple approach to text genres, from the point of view of linguistics, sociology and psycholinguistics, by examining the strategic choices of the writer, for example. Discourse analysis too is traditionally interested in the "conditions of production" (cf. Courtine 1981, for the French stream), although their effective relationship with the writing process remains unclear. Within the framework of Critical Discourse Analysis, Wodak (1992, 494 and 525) was interested in the relationship between text comprehension and text production, and argued that both are simultaneously cognitive and emotional processes. Two questions formulated by Wodak (1992, 494) are particularly relevant here: How does text comprehension (*Textverstehen*) differ from text production (*Textproduktion*)? And how does the understanding of a text (*Textverstehen*) differ from the interpretation of a text (*Textinterpretation*)?

It is evident that the writing process itself involves a certain anticipation of these questions, as the writer anticipates the later process of understanding and interpreting the resulting text. As such it is anchored in a double intertextuality (see also Bazerman & Prior 2004). Various constraints are thus at work in the writing process (Plane et al. 2010; Cislaru 2014).

In line with these findings, Grésillon and Perrin (this volume) argue that the collaborative study of writing in real-life situations – i.e. excluding laboratory data – is able to deal with the complexity of the writing act, which involves heterogeneous factors such as actors, media, environments, content, and textuality, as well as material, mental and social activity. The authors develop a cross-cutting

methodology that encompasses various types of constraints and processing situations, which constitutes a valuable advance in the study of writing interfaces.

One of the questions that emerge at this stage is the very definition of writing and of the act of writing (see also Delbreilh this volume), and its analytical force. In her paper, Plane (this volume) discusses the notion of the act of writing as a complex and heterogeneous activity, and highlights several aspects of the dynamics of writing both from the point of view of the process (temporality, writing operations) and from the point of view of the product (textuality). She notes the divergences between different approaches to the temporality of writing, and underlines their productivity in drawing out both the cognitive processes and the linguistic mechanisms involved in the writing process.

2.   Writing as a process: Fine-grained viewpoints on the act of writing

The study of the writing process implies putting all stages and levels of the production mechanisms under scrutiny (see also Van der Geest 1996): i.e. the situation (professional, personal, educational context) that hosts and generates the act of writing and all the related pragmatic issues, the writer's behavior and specific gestures (eye movements, pen or keystroke activity, etc.), the use of sources (books, Web sources, oral exchanges and discussions, etc.), the timespan involved and its relationship to previous stages and levels (length, pauses, fluency, etc.), the space which is exploited (pages, back-and-forth movements, etc.), the material produced and its qualities (linguistic description, operations involved, etc.). Until now, these aspects have been split among various disciplines, as mentioned above.

In terms of either a complex longitudinal path or a dynamic falling within temporal constraints (the real time of writing, for example) and revealing the way the text is shaped, the process involves various dimensions, such as:

– the context of production (social, cultural, historical, physical);
– the articulation of writing practices with socio-cognitive habits: domains such as literacy and the anthropology of writing contribute to the description of these practices and help to situate writing within contexts that require additional interpretation (Barton & Papen 2010a; Fraenkel 2007);
– the steps of the writing process, such as composition and revision (Fuchs et al. 1987; Fenoglio & Chanquoy 2007), and their temporality (van den Bergh & Rijlaarsdam 1996);
– the mental operations postulated by cognitive psychologists (Hayes & Flower 1980; see Alamargot & Chanquoy 2001, for a review);

–   the "real-time" process, recorded by keystroke logging programs (like Input-
    log, see Leijten & Van Waes 2006 and Leijten et al. this volume; ScriptLog, see
    Sullivan & Lindgren 2006; Leblay and Caporossi this volume, etc.). Spelman
    Miller and Sullivan (2006, 2) speak of a "writer- (rather than text-) based per-
    spective on writing."

The recent development of digital tools (see Van Waes & Mangen 2012) as well as
the development of cognitive psychology (Berninger 2012) have made possible
increasingly complex insights into the process of writing, and allowed the collec-
tion and analysis of quantitative and statistical data (Perrin & Wildi 2010). These
approaches are mainly based on psychology, the cognitive sciences and the com-
puter sciences. They open up access to what can otherwise look like the "black
box" of writing activity: memory (Olive et al. 2008), pauses (Kesseling 1992; Olive
et al. 2009), revision dynamics (Severinson-Eklundh & Kollberg 2001; Allal et al.
2004), etc. On the content side, Baaijen (2012) – among others – discusses the
creative dimensions of the writing process and the ways that writing influences
knowledge and ideas. On the formal side, some recent studies seek to articulate
linguistic description to real-time data (see Leijten et al. 2012; Macken et al. 2012).

Leijten, Van Waes and Van Horenbeeck (this volume) take the integration
of linguistic analysis with keystroke logging a step further. Their paper offers an
original point of view on the writing process; for the first time, it proposes to use
linguistic data from the writing process as a basis for neuropsychological tests to
diagnose Alzheimer's disease. They thus articulate the linguistic dimension of the
writing process to product data, and offer an interpretation of both in terms of
linguistic processing phenomena.

Galbraith and Baaijen (this volume) develop the concept of a dual process of
writing, which implies conflict between a system designed for constructing objects
and a system designed for action. Insofar as content is built throughout the writing
process, product and process are inseparable, with writing modelling the indi-
vidual's thoughts in accordance with culturally and socially shared norms. The
authors argue that text production is an active knowledge constitution process,
that the knowledge object, the final content of a written text, emerges during this
process, and that different memory systems are involved in writing processes in a
complex way.

From a different standpoint, Fenoglio (this volume) points out the elaboration
of theoretical thinking through the writing process, by following parallell linguistic
(lexical, syntactical) potentialities in the drafts of Saussure's and Benveniste's texts.
The broken linearity of the graphic traces clearly reflects a process of assembly.

Leblay and Caporossi (this volume) show that the study of the product alone
does not permit an understanding of the process and dynamics of writing. They

examine revision operations within the writing process and the representation of their spatial and temporal characteristics, based on real-time data from keystroke logging. The final visualization, supported by the mathematical theory of graphs, highlights the use of different linguistic strategies depending on the writer's skills.

Among the questions that may be formulated at this stage are those of what the context of writing and the writer's abilities can reveal about the text, and vice versa. There is a tight relationship between the simplicity of discourse constraints and the possibility of mastering them, which is assimilated to discursive competence (Hymes 1971; Maingueneau 1984, 51). Writing competence and performance can be connected to specific constraints, such as genre specificities (Cislaru and Lefeuvre this volume), professional writing skills (Brunner and Pordeus Ribeiro this volume; Doquet and Poudat this volume; Olive and Cislaru this volume), neuropsychological skills (Leijten et al. this volume), etc.

3.   From text to process, and back: What the text is

The process of writing involves a theory of discourse and meaning construction, as noted by Flower (1994, 5 et seq.). Can the features of discourse be disconnected from its practice? The answer is no. Indeed, as Stubbs (1997, 104) points out, "a text is seen as a series of traces left by the processes of production," and it is difficult to interpret the traces outside of the process. In line with this observation, Stubbs (1997, 110) proposes an ethnographic study of actual text production. "Writing as Text" is also the title of Part V of Bazerman's (2007) *Handbook of Research on Writing*, which clearly confirms the necessity of articulating the results of linguistic research on finished texts (see, for instance, Schleppegrell 2007) with data from multi-disciplinary research on writing as practice and process. On the other hand, Sanders & van Wijk (1996) propose to use text analysis to identify writing strategies. It is important to take into account the non-transparency of results and data on both sides, insofar as the different methodologies and aims at work shape the interpretation of the data.

The practices of discourse analysis and corpus linguistics, for instance, invite fine-grained text analysis, which takes into account various linguistic and structural parameters. The thorough methodologies applied in these areas suggest a need for caution with regard to certain speculative conclusions about language, production, and constraints.

Sinclair's (2004) book *Trust the Text: Language, Corpus and Discourse* is rather suggestive in this respect. Corpus-driven linguistics offers new insights into language structure, and even into the semantic-grammatical profile of linguistic units. It is indeed a privileged way to observe language in use and to obtain a global

view of discourse based on large corpora. The study of patterned data (cf. Sinclair 2004; Biber et al. 2004) reveals regularities and offers information about language structure and product(ion). The lexical/pattern grammar that has been developed in these studies raises some questions about the production process, insofar as its application in the domain of language acquisition inherently attests to a usage/ norm, but does not serve in itself to explain the underlying process.

> We should be open to what it [the text] may tell us. We should not impose our ideas on it, except perhaps just to get started. Until we see what the preliminary results are, we should apply only frameworks that are loose and flexible, in order to accommodate the new information that will come from the text. We should expect to encounter unusual phenomena; we should accept that a large part of our linguistic behaviour is subliminal, and that therefore we may find a lot of surprises. We should search for models that are especially appropriate to the study of text and discourse. (Sinclair 2004, 23)

The choice of the units of analysis is a crucial step in the study of writing interfaces. Tolchinsky et al. (1999) distinguish between generic (roughly, discourse types and genres, from a pragmatically oriented perspective) and microscopic approaches (taking into account specific linguistic features, following Biber's approach [2009; Biber et al. 2004], for instance). In corpus linguistics, macro- and micro-features are often associated (see Bednarek 2009). But the complexity of the writing process and the theoretical principles that frame approaches to process and product influence the interpretation of data. It is clear, for example, that the adoption of either the product or the process viewpoint determines the nature of the units of analysis, such as linguistic forms or temporal data.

Olive and Cislaru's paper (this volume) investigates the nature and interpretation of corpus linguistic data by comparing the linguistic structure of bursts of writing (which represent the routinized dimension of the writing process) and of repeated segments (considered in the literature as routine elements of the text as a product). The results highlight interesting differences between the two categories of data, and show that linguistic data that may be interpreted as routines when analyzing the product probably do not function as routines at the process level. The paper also offers hints on the nature of the gap between process and product and its impact on theoretical and methodological choices.

The notion of text itself is also subject to discussion. In our societies, texts are framed by various constraints of status, production, authorship, etc. From this point of view, a text cannot be reduced to a series of traces left on a medium. Fenoglio (this volume) sees *the* text – the result of a linearization process through manuscripts and drafts (see also de Beaugrande 1984 on the linearity of text composition) – as ended, closed, as soon as it is submitted to an editor, as soon as it passes from the private to the public sphere, etc. Mahrer et al. (this volume) argue

that, despite editorial constraints, a literary text bears its textuality long before it is published, and may continue to evolve beyond publication.

4.   Writing in context: A socio-anthropological approach

The process-product interface may be examined at a "macro" level that attempts to articulate the contextualized, ideological, interactional, and engaged aspects of writing (cf. Ivanič 1998; Pearce 2010 on the relationship between writing and identity construction; Hyland 2005, with the notion of metadiscourse) and the discourse itself. Written and oral texts have been scrutinized by disciplines such as discourse analysis and sociolinguistics (see Wodak 1992; Boutet 1997; Branca-Rosoff & Schneider 1994) from an interpretive perspective, in an attempt to uncover social, cultural and historical principles through text analysis. The study of the writing process in itself and a literacy viewpoint offer new insights into the relationship between language practice (more specifically writing activity) and other social practices.

The core interest of the anthropology of writing is "to examine the processes of production and use of texts" (Barton & Papen 2010b, 7), from the perspective of a strong relationship between writing and social practices:

> Examining written texts is essential for understanding how societies operate and are organized, how institutions communicate with the public, how work is being done, how individuals and social groups organize their lives and make sense of their experiences and how cultures in all their variations are produced and reproduced.   (Barton & Papen 2010b, 5)

Literacy studies are more generally socioculturally and anthropologically grounded. As recalled by Gee (2000, 189), New Literacy Studies (NLS), mainly grounded in the anthropology, "is based around the idea that reading, writing and meaning are always *situated* within specific social practices within specific Discourses," where the notion of context is seen from the viewpoint of its dynamics. A context-dependent approach could also be of help given that the conditions of production, which are embedded in a specific situation, are not identical to the conditions of reception, and a fine-grained definition of the context is thus required for a thorough interpretation of the studied phenomena.

The sociocultural theory of writing (see Prior 2004, 2006) offers clues to understand the complexity of the interface, in the line of Vygotsky's thought:

> Sociocultural theory argues that activity is *situated* in concrete interactions that are simultaneously *improvised* locally and *mediated* by prefabricated, historically provided tools and practices, which range from machines, made objects, semiotic

means (e.g. languages, genres, iconographies), and institutions to structured environments, domesticated animals and plants and, indeed, people themselves. Mediated activity involves *externalization* (speech, writing, the manipulation and construction of objects and devices) and *co-action* (with other people, artifacts and elements of the social-material environment) as well as *internalization* (perception, learning). (Prior 2006, 55)

In his paper, Prior (this volume) gives a sound illustration of the sociocultural approach to writing, offering an exhaustive overview of the complexity of writing and the writing process. He underlines the dispersed, distributed and laminated chronotopic structure of the writing process, and relates these to its heterogeneous mediation by tools, practices, other people, etc. In this light, he formulates a new definition of writing: "Understood as a blend of texts, persons, activities, meditational means, and social formations/practices, writing appears as temporally and spatially stretched out trajectories rather than as punctual events in a narrow and isolated here-and-now."

In the framework of the anthropology of writing, Delbreilh (this volume) advocates taking into account the whole set of heterogeneous aspects of the writing process: linguistic, graphical, material, situational. All of these aspects may be understood as action-oriented, writing being a pragmatic act. The writing act (Fraenkel 2010) is thus seen as an essential part of collective action in literate societies.

Externalization and co-action imply the Other, as a potential reader of the text. Writing transforms discourse into a public good, a text-for-the-Other, its legibility being a condition of its existence and social perception. For instance, Cislaru and Lefeuvre (this volume) show how the private vs. public nature of writing determines linguistic choices in different discourse genres. Fenoglio (this volume) connects the linearization process that transforms an assembly of parallel data and forms into a linear textual construction with the constraints of the public sphere.

Brunner and Pordeus Ribeiro (this volume) connect linguistic production with exterior constraints that create a tension between subjective and objective stances. Text revision recorded during social workers' production of successive drafts of reports reveals the influence of the social institution, and shows that the contextualization of corpus data is a first step toward their interpretation. Finally, they show that not only the final text as a product, but the whole process of writing and editing is oriented toward maintaining the smooth functioning of the institution. They thus examine the interactions between text producers and their texts throughout the whole series of revision operations leading to the final text, in a way that confirms the character of (written) communication as "social engagement" (cf. Hyland 2005).

5. The volume

The chapters in this volume are united by their approach to examining writing at the interface. Each chapter offers methodological cues (see in particular Grésillon and Perrin this volume, Plane this volume, Prior this volume) for the joint study of both process and product. The actual indivisibility of process and product is the underlying theme of the volume, and it is particularly prominent in Galbraith and Baaijen (this volume). Although the individual chapters do not focus equally on process and product, as one or another dimension dominates in accordance with the theoretical framework adopted by the authors, each study confronts process and product in one way or another. Moreover, the diversity of approaches brings to light different features of the interface, depending on whether the point of departure is the product or the process. Some papers question the nature and the identity of the product in light of the dynamics of the process (see Mahrer et al. this volume; Cislaru and Lefeuvre this volume); others propose new descriptive categories which might better fit a unified approach to process and product (see Olive and Cislaru this volume).

The volume is structured in five sections.

Part One, "Some core questions about writing," points out various difficulties and proposes or confronts new methods in the study of writing.

Plane (this volume) presents a concise and enlightening history of questions about writing in its multiple facets, and more specifically from the product vs. process viewpoint in the field of teaching. She points out the difficulties for writers of grasping writing as both object and act(ivity), as well as the resulting product, the text, and notes how more traditional approaches to texts and writing have evolved towards chronometric methodologies with the support of technological advances. Ultimately, it is impossible to completely capture the dynamics of writing, but seeking to do so offers interesting insights into the writing process and linguistic functioning.

Grésillon and Perrin (this volume) advocate for a study of writing informed by applied linguistics, and thus conceived as "a joint activity of researchers, practitioners, and society at large." They bring together two methods developed in specific contexts: the *Idée suisse*, developed around Daniel Perrin in Zurich and that examines writing-in-context in the newsroom, and critical genetics, developed in Paris (see Hay 1996; Anokhina & Pétillon 2009) and illustrated by the study of the drafts and revisions of Heine's poem "Lebensfahrt."

Part Two, "Linguistic forms and choices at the interfaces," brings together three papers which offer different insights based on the same corpus data, the reports of social workers' on children, which are addressed to judges.

Cislaru and Lefeuvre (this volume) analyse the use of verbless sentences in the corpus of social workers' reports on the one hand, and in a corpus of diary entries with associated letters, on the other hand. They combine a longitudinal study of the drafts of the reports and a contrastive study of discourse genres in order to explore the impact of the private vs. public dimension of the constraints on the writing process.

Brunner and Pordeus Ribeiro (this volume) cross linguistics and a socio-genetic approach. They offer a detailed analysis of revision operations, and address the question of linguistic choices in writing. Alongside interesting linguistic observations illustrating discourse adaptability (cf. Verschueren & Brisard 2009, 40 et seq.) – in terms of intensity, for instance – this paper highlights the imbrication of writing and social practices or, more precisely, sheds light on the nature of writing as a social practice.

Olive and Cislaru (this volume) present a linguistic description of psycholinguistic units of analysis: bursts of production. This original approach, which combines linguistic description, textometry and cognitive psycholinguistics in an examination of real-time data, offers an interesting viewpoint on how writing data can be categorized and interpreted, and highlights a discrepancy between aspects of product and process that are respectively viewed as "routine."

Part Three, "Tracks and traces of the writing process," illustrates the archeology of writing through the approach of critical genetics. The first two papers look at the phenomenon of coexistence (of different versions, formulations, semiotic means…) and its importance in the interpretation of both process and product.

Fenoglio (this volume) examines scientific (Saussure, Benveniste) and literary texts and *avant-textes* (drafts), both manuscript and computer-typed. The author emphasizes the delinearization of the writing process, which is mainly visible in manuscripts, where parallel versions coexist. This visibility brings meaning construction close to the reader and offers the possibility of decoding from various standpoints.

In a slightly different framework, editorial genetics, Mahrer et al.'s paper (this volume) offers a fine-grained description of the phases that characterize literary text production and focuses on rewriting processes after initial publication. The authors perform a modular linguistic analysis supported by automatic analysis, going through several levels of analysis, from word to text, and including punctuation. The notion of "variation," which involves the coexistence of two objects (two versions of the same text, two equivalent words, etc.), is central to the paper.

Leblay and Caporossi (this volume) go beyond the recording and description of revision dynamics from an exclusively temporal point of view, and offer a legible spatio-temporal representation of different writing strategies according to

personal writing skills (see Lindgren et al. 2008 on writing skills development). Their concise semiotization of the writing process helps to draw out patterns in the complexity of revision. Moreover, it confirms that writing is a specific form of communication, anchored in specific conditions of production.

In Part Four, "Writing practices in context",

Prior (this volume) presents several case studies that illustrate the chronotopic lamination of writing activity. He argues that texts as a product of writing are non-autonomous, and shows that writing involves multiple semiotic resources. The notion of *semiotic remediation*, which argues for a dialogic approach to all semiotic practices and performances, and the notion of *literate activity*, a confluence of streams of activities (reading, thinking, talking, feeling, etc.), are meant to favor full accounts of writing.

Delbreilh (this volume) presents an original study of leaflets from a feminist protest march in 2011. Taking a complex approach to writing as an act and as part of collective action, the author observes all steps in the performativity of the distributed writings, from their making – composition of the text, layout, printing – to the strategies applied in handing it out, and argues that the stages of production and dissemination are part of the writing act.

Doquet and Poudat (this volume) carry out a fine-grained textometric analysis of inspection reports in primary school. They highlight the multiple entanglements between the laminated writing process (see also Prior this volume), genre and institutional constraints, and personal styles. They also emphasize the interdependence of lexical choices and the strength (but not the polarity) of assessments. A thorough analysis of lexical choices and distribution in the texts allows the authors to identify writing styles and categories of writers.

Part Five, "Cognitive insights through writing studies," articulates methods of writing research to core questions on human cognitive processes and capacities. Linguistic data, although approached differently in the different chapters, are central to each of these approaches, insofar as they offer access to various cognitive features and enable the assessment of cognitive competences.

Galbraith and Baaijen (this volume) interconnect writing processes, knowledge change, and text quality in order to examine the relationships between global planning processes (see also Torrance & Galbraith 1999), implicit text processing, and individual self-monitoring in writers. They discuss the way process organization (for instance, synthetically planned text production and other drafting strategies) determines the understanding and processing of ideas, and identify different types of writers: for example, high vs. low self-monitors, writers who prioritize either rhetorical or dispositional goals, and writers who develop their ideas under different writing conditions.

Leijten, Van Waes, and Van Horenbeeck (this volume) argue that the writing process may be a valuable observable, and could increase diagnostic accuracy for Alzheimer's disease if looked at through its cognitive and linguistic aspects. This approach enables the inter- and intrapersonal comparison of writing characteristics in demented and healthy elderly. The paper explores the keystroke logging approach, comparing several tools (Inputlog, Scriptlog, and Translog), and then details the linguistic analysis module of Inputlog, which includes a part-of-speech tagger, a lemmatizer, a chunker, a syllabifier, and also word frequency information. It demonstrates the importance and added value of taking into account the linguistic devices and mechanisms at work during the process of writing, as well as patterns of pauses between words, notably when testing the cognitive-linguistic capacities of demented and healthy individuals.

Based on empirical writing data, either real-time or archived drafts, this book aims to contribute to the development of an interpretive approach to writing and its dynamics. It gives an overview of the state of research on the process-product interface through a range of viewpoints on process, product, and the links between them. Collectively, its chapters explore the possibility of establishing a coherent path from the real-time dynamics of the writing process to the product anchored in its formal dimensions and its pragmatic functions. The result is a look at how results and concepts from different domains may support each other in the development of a mixed approach to the process-product interface.

Acknowledgements

References

Alamargot, Denis, and Lucile Chanquoy (Eds.). 2001. *Through the Models of Writing. Studies in Writing, Vol. 9*. Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-010-0804-4

Allal, Linda, Lucile Chanquoy, and Pierre Largy (Eds). 2004. *Revision: Cognitive and Instructional Processes* (Vol. 13). Dordrecht et al.: Kluwer Academic Publishers. DOI: 10.1007/978-94-007-1048-1

Anokhina, Olga, and Sabine Pétillon. 2009. *Critique génétique: concepts, méthodes, outils.* Paris: IMEC.

Baaijen, Veerle Marije. 2012. *The Development of Understanding Through Writing.* Groningen Dissertations in Linguistics 107.

Barton, David, and Uta Papen (Eds.). 2010a. *The Anthropology of Writing: Understanding Textually Mediated Worlds.* London: Continuum.

Barton, David and Uta Papen. 2010b. "What is the anthropology of writing?" In *The Anthropology of Writing: Understanding Textually Mediated Worlds*, ed. by David Barton, and Uta Papen, 4–32. London: Continuum.

Bazerman, Charles (Ed.). 2007. *Handbook of Research on Writing: History, Society, School, Individual, Text.* Mahwah, NJ: Erlbaum.

Bazerman, Charles, and Paul A. Prior (Eds.). 2004. *What Writing Does and How It Does It. An Introduction to Analyzing Texts and Textual Practices.* Mahwah, NJ/London: Lawrence Erlbaum. DOI: 10.1207/s15427625tcq1602_5

Beaugrande, Robert de. 1984. *Text Production: Toward a Science of Composition.* Norwood, N.J.: Ablex. DOI: 10.1017/s0047404500011520

Bednarek, Monika. 2009. "Language patterns and Attitude." *Functions of Language* 16(2): 165–192. DOI: 10.1075/fol.16.2.01bed

van den Bergh, Huub, and Gert Rijlaarsdam. 1996. "The dynamics of composing: modeling writing process data." In *The Science of Writing. Theories, Methods, Individual Differences and Applications*, ed. by C. Michael Levy, and Sarah Ransdell, 207–231. Mahwah: Erlbaum.

Berninger, Virginia W. (Ed.). 2012. *Past, Present and Future Contributions of Cognitive Writing Research to Cognitive Psychology*. New York/London: Psychology Press. DOI: 10.4324/9780203805312

Berninger, Virginia W., and Todd L. Richards. 2002. *Brain Literacy for Educators and Psychologists*. New York: Academic Press. DOI: 10.1016/b978-012092871-2/50006–1

Bhatia, Vijay K. 1993. *Analysing Genre. Language Use in Professional Settings*. London/New York: Longman. DOI: 10.1017/s0272263100013668

Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. "If you look at…: Lexical bundles in university teaching and textbooks." *Applied Linguistics* 25(3): 371–405. DOI: 10.1093/applin/25.3.371

Boutet, Josiane. 1997 [1994]. *Construire le sens*. Berne: Peter Lang. DOI: 10.1017/s0047404500019254

Branca-Rosoff, Sonia, and Nathalie Schneider. 1994. *L'écriture des citoyens. Une analyselinguistique de l'écrituredes peu lettrés pendant la période révolutionnaire*. Paris: Klincksieck. DOI: 10.1017/s0959269500002805

Brandt, Deborah. 1990. *Literacy as Involvement: The Acts of Writers, Readers and Texts*. Carbondale: Southern Illinois University Press.

Candlin, Christopher N., and Ken Hyland (Eds.). 1999. *Writing: Text, Processes, and Practices*. Harlow: Addison Wesley Longman.

Cislaru, Georgeta. 2014. "Contraintes linguistiques et contextuelles dans la production écrite." *Les Carnets du Cediscor* 12: 55–74.

Courtine, Jean-Jacques. 1981. "La notion de 'condition de production de discours.'" *Langaages* 15(62): 19–32. DOI: 10.3406/lgge.1981.1891

Doquet-Lacoste, Claire. 2004. "Genèse des textes, genèse des scripteurs: regards sur la langue dans l'écriture." In *Ecriture. Approches en sciences cognitives*, ed. by Annie Piolat, 233–250. Aix-en-Provence: Publication de l'Université de Provence.

Fenoglio, Irène, and Lucile Chanquoy (Eds.). 2007. *Avant le texte: les traces de l'élaboration textuelle*. Special issue of *Langue française* 155. DOI: 10.3917/lf.155.0003

Flower, Linda. 1994. *The Construction of Negotiated Meaning: A Social Cognitive Theory of Writing*. Carbondale, IL: Southern Illinois Press. DOI: 10.2307/358881

Fraenkel, Béatrice. 2007. "Actes d'écriture: Quand écrire c'est faire." *Langage & Société* 121–122: 101–112. DOI: 10.3917/ls.121.0101

Fraenkel, Béatrice. 2010. "Writing Acts: When Writing Is Doing." In *The Anthropology of Writing: Understanding Textually Mediated Worlds*, ed. by David Barton and Uta Papen, 33–43. London/New York: Continuum.

Fuchs, Catherine, Almuth Grésillon, Jean-Louis Lebrave, Jean Peytard, and Josette Rey-Debove. 2003[1982]. *La genèse du texte. Les modèles linguistiques*. Paris: CNRS Editions.

Galbraith, David. 2009. "Cognitive models of writing." *German as a Foreign Language* 2–3: 7–22.

Gee, James Paul. 2000. "The New Literacy Studies. From 'socially situated' to the work of the social." In *Situated Literacies. Reading and Writing in Context*, ed. by David Barton, Mary Hamilton, and Roz Ivanič, 180–196. London/New York: Routledge.

Gehle, Quentin L., and Duncan J. Rollo. 1977. *The Writing Process*. Toronto: Macmillan.

Hay, Louis. 1996. "History or Genesis?" *Yale French Studies* 89: 191–207. DOI: 10.2307/2930347

Hayes, John R., and Linda S. Flower. 1980. "Identifying the Organization of Writing Processes." In *Cognitive Process in Writing*, ed. by Lee Gregg, and Erwin R. Steinberg, 3–30. Hillsdale: Lawrence Erlbaum. DOI: 10.1177/002194368201900207

Hyland, Ken. 2005. *Metadiscourse. Exploring Interaction in Writing*. London/New York: Continuum. DOI: 10.1177/0957926508097560

Hymes, Dell. 1971. *On Communicative Competence.* Philadelphia: University of Pennsylvania Press.

Ivanič, Roz. 1998. *Writing and Identity. The Discoursal Construction of Identity in Academic Writing*. Amsterdam: John Benjamins. DOI: 10.1075/swll.5

Keseling, Gisbert. 1992. "Pause and intonation contours in written and oral discourse." In *Cooperating with Written Texts. The Pragmatics and Comprehension of Written Texts*, ed. by Dieter Stein, 31–66. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110881196.31

Leijten, Mariëlle, and Luuk Van Waes. 2006. "Inputlog: New Perspectives on the Logging of On-Line Writing." In *Studies in Writing: Vol. 18. Computer Key-Stroke Logging and Writing: Methods and Applications*, ed. by Kirk P.H. Sullivan, and Eva Lindgren, 73–94. Oxford: Elsevier.

Leijten, Mariëlle, Lieve Macken, Véronique Hoste, Erik Van Horenbeeck, and Luuk Van Waes. 2012. "From Character to Word Level: Enabling the Linguistic Analyses of Inputlog Process Data." In *Linguistic and Cognitive Aspects of Document Creation and Document Engineering. Proceedings of the Second Workshop on Computational Linguistics and Writing* (Avignon, CLW 2012), ed. by Michael Piotrowski, Cerstin Mahlow, and Robert Dale, 1–8. ACL: Stroudsburg.

Lindgren, Eva, Mariëlle Leijten, and Luuk Van Waes. 2008. "Writing development and knowledge crafting." *Conference Writing Research Across Borders*, Santa-Barbara.
DOI: 10.1075/wll.14.2.02lin

Macken, Lieve, Véronique Hoste, Mariëlle Leijten, and Luuk Van Waes. 2012. "From keystrokes to annotated process data: Enriching the output of Inputlog with linguistic information." *Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC'12), 2224–2229. Istanbul: Turkey.

Maingueneau, Dominique. 1984. *Genèses du discours*. Liège: Mardaga.

Murray, Donald D. 1972. "Teach language as a process not product." *The Leaflet*, November 1972 11–14. ⟨http://www.willamette.edu/gse/owp/docs/TeachWritingasaProcessNotProduct.pdf⟩

Olive, Thierry, Ronald T. Kellogg, and Annie Piolat. 2008. "Verbal, visual and spatial working memory demands during text composition." *Applied Psycholinguistics* 29: 669–687. DOI: 10.1017/s0142716408080284

Olive, Thierry, Rui A. Alves, and São Luís Castro. 2009. "Cognitive processes in writing during pauses and execution periods." *European Journal of Cognitive Psychology* 21: 758–785. DOI: 10.1080/09541440802079850

Olive, Thierry, and Jean-Michel Passerault. 2012. "The visuospatial dimension of writing." *Written Communication* 29: 326–344. DOI: 10.1177/0741088312451111

Pearce, Lynn. 2010. "'Writing' and 'region' in the 21st century: Epistemological reflections on regionally-located art and literature in the wake of the digital revolution." *European Journal of Cultural Studies* 13: 27–41. DOI: 10.1177/1367549409352550

Perrin, Daniel (Ed.). 2003a. *Pragmatics of Writing*. Special issue of *Journal of Pragmatics* 35(6).

Perrin, Daniel. 2003b. "Towards a Pragmatics of Writing." *Journal of Pragmatics* 35(6): 825–828. DOI: 10.1016/s0378-2166(02)00120-0

Perrin, Daniel. 2013. *The Linguistics of Newswriting*. Amsterdam, New York et al.: John Benjamins.

Perrin, Daniel, and Marc Wildi. 2010. "Statistical modeling of writing processes." In *The Traditions of Writing Research*, ed. by Charles Bazerman, 378–393. New York: Routledge.

Plane, Sylvie, Thierry Olive, and Denis Alamargot (Eds.). 2010. *Traitement des contraintes de la production d'écrits: aspects linguistiques et psycholinguistiques*. Special issue of *Langages* 177. DOI: 10.3917/lang.177.0003

Prior, Paul. 2004. "Tracing Process: How Texts Come Into Being." In *What Writing Does and How It Does It. An Introduction to Analyzing Texts and Textual Practices*, ed. by Charles Bazerman, and Paul A. Prior, 167–200. Mahwah, NJ/London: Lawrence Erlbaum. DOI: 10.1207/s15427625tcq1602_5

Prior, Paul. 2006. "A Sociocultural theory of writing." In *Handbook of Writing Research*, ed. by Charles A. MacArthur, Steve Graham, and Jill Fitzgerald, 54–66. New York: Guilford. DOI: 10.1111/j.1467-873x.2008.00423.x

Sanders, Ted, and Carel Van Wijk. 1996. "Text analysis as a research tool. How hierarchical text structure contributes to the understanding of conceptual processes in writing." In *The Science of Writing. Theories, Methods, Individual Differences and Applications*, ed. by Michael C. Levy, and Sarah Ransdell, 251–270. Mahwah: Erlbaum.

Schleppegrell, Mary J. 2007. "Grammar, the Sentence, and Traditions of Linguistic Analysis." In *Handbook of Research on Writing: History, Society, School, Individual, Text,* ed. by Charles Bazerman, 673–691. Mahwah, NJ: Erlbaum.

Severinson-Eklundh, Kerstin, and Py Kollberg. 2001. "Studying writers' revision patterns with s-notation analysis." In *Contemporary Tools and Techniques for Studying Writing*, ed. by Thierry Olive, and Michael C. Levy, 89–104. Dordrecht/Boston/London: Kluwer Academic Publishers. DOI: 10.1007/978-94-010-0468-8_5

Sinclair, John. 2004. *Trust the Text. Language, Corpus and Discourse*. London/New York: Routledge. DOI: 10.4324/9780203594070

Spelman Miller, Kristyan, and Kirk P.H. Sullivan. 2006. "Keystroke logging: an introduction." In *Computer Keystroke Logging: Methods and Applications*, ed. by Kirk P. H. Sullivan, and Eva Lindgren, 1–9. Oxford: Elsevier.

Stein, Dieter. 1992. "Cooperating with written texts: Towards a science of the text. Issues and overview." In *Cooperating with Written Texts. The Pragmatics and Comprehension of Written Texts*, ed. by Dieter Stein, 1–12. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110881196.1

Strömqvist, Sven, Kenneth Holmqvist, Victoria Johansson, Henrik Karlsson, and Åsa Wengelin. 2006. "What keystroke logging can reveal about writing." In *Computer Keystroke Logging: Methods and Applications*, ed. by Kirk P. H. Sullivan, and Eva Lindgren, 45–71. Oxford: Elsevier.

Stubbs, Michael. 1997. "Critical comments on Critical Discourse Analysis (CDA)." In *Evolving Models of Language*, ed. by Ann Ryan, and Alison Wray, 100–116. Clevedon: Multilingual Matters.

Sullivan, Kirk P.H., and Eva Lindgren (Eds). 2006. *Computer Key-Stroke Logging and Writing* (*Studies in Writing, Methods and Applications, vol. 18*). Oxford: Elsevier.

Tolchinsky, Liliana, Joan Perera, and Noemi Argerich. 1999. "Units of analysis in two approaches to text analysis." In *Writing and Learning to Write at the Dawn of the 21st Century. 1998 Writing Conference Proceedings*, ed. by Marie-Françoise Crété, and Eric Espéret, 339–344. Poitiers: LaCo-CNRS, Maison des Sciences de l'Homme et de la Société, Université de Poitiers.

Torrance, Mark, Denis Alamargot, Montserrat Castelló, Franck Ganier, Otto Kruse, Anne Mangen, Liliana Tolchinsky, and Luuk Van Waes (Eds.). 2012. *Learning to Write Effectively: Current Trends in European Research. Studies in Writing, vol. 25*. Brill.

Torrance, Mark, and David Galbraith (Eds). 1999. *Knowing What to Write: Conceptual Processes in Text Production*. Amsterdam: Amsterdam University Press. DOI: 10.5117/9789053563076

Van der Geest, Thea. 1996. "Studying 'real life' writing processes. A proposal and an example." In *The Science of Writing. Theories, Methods, Individual Differences, and Applications*, ed. by C. Michael Levy, and Sarah Ransdell, 309–322. Mahwah: Erlbaum.

Van Waes, Luuk, and Anne Mangen. 2012. "Introduction. Tools for studying and supporting writing. Technological advances in writing research." In *Learning to Write Effectively. Current Trends in European Research*, ed. by Mark Torrance, Denis Alamargot, Montserrat Castello, Franck Ganier, Otto Kruse, Anne Mangen, Liliana Tolchinsky, and Luuk Van Waes, 345–351. Bingley: Emerald.

Verschueren, Jeff, and Frank Brisard. 2009. "Adaptability." In *Key Notions for Pragmatics*, ed. by Jeff Verschueren and Jan-Ola Östman, 28–47. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hoph.1.02ver

Vygotsky, Lev S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Ed. by Michael Cole, Vera John-Steiner, Sylvia Scribner, and Ellen Souberman. Cambridge, MA: Harvard University Press. DOI: 10.1017/s0047404500008071

Wodak, Ruth. 1992. "Strategies in text production and text comprehension: A new perspective." In *Cooperating with Written Texts. The Pragmatics and Comprehension of Written Texts*, ed. by Dieter Stein, 493–528. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110881196.493

Some core questions about writing

Some problems encountered in the description and analysis of the dynamics of writing

Sylvie Plane

Université Paris-Sorbonne, EA 4509- STIH

Research dealing with the dynamics of writing is carried out in different theoretical fields and consequently resorts to different research methodologies. The aim of this paper is to study which definitions of the act of writing underlie the research reports which all intend to describe its dynamics.

Two aspects of the dynamics of writing will be studied in order to see how they are organized into systems that configure the description of the act of writing in diverging ways: (i) temporality itself with an emphasis on the consequences of distortion or focus caused by the size of the time span or the granularity chosen; (ii) the notion of writing seen as a partly observable and partly unobservable activity that leads the researchers to reconstruct it in ways that differ with the indices they have selected.

Keywords: writing; temporality; act of writing; methodology

1. Consensus and divergences surrounding the idea of the dynamics of writing

Whereas the analysis of the comprehension of texts, their transmission, and their functioning has a venerable tradition of research that lasted throughout antiquity from Aristotle to Cassiodorus, was perpetuated by scholastic teaching, and was renewed by stylistics and different currents of contemporary linguistics, research on written production is relatively recent. It began because of ruptures that occurred in the 20th century: the rupture with philology under the influence of genetic criticism[1] that rejected any teleological vision of literary creation (Hay 1979), the rupture with subjective representations of the act of writing under the impulse of cognitive psychology and psycholinguistics whose theoretical and

---

1. "Genetic criticism" is a critical method that originated in France ("*critique génétique*"), which focuses on the sources used in a work as well as the writer's manuscripts.

methodological demands were not in accordance with the impressionist conceptions that were predominant at the time.

According to Matsuhashi (1982), 1970 was the turning point when the focus shifted from product to production. Various 20th-century works contributed to initiating this transition. The modeling of text production carried out by Hayes & Flower in 1980, at a time when questions about the activity of producing a text were becoming more crystallized, is considered to constitute an important reference point. This was followed soon thereafter by adjustments that would complement it (Flower, Hayes, Carey, Schriver, and Stratman 1986) and by partial or global modelizations that resembled it.

This change of focus was facilitated by three factors: (i) the integration of some linguistic issues into more general questions in the area of cognitive psychology regarding creation and act, (ii) the focus of multiple currents of linguistic research on problems of a textual or discursive order (in particular Halliday & Hasan 1976 *passim*; Kintsch & van Dijk 1978) and the related development of pragmatic approaches that would fuel reflection about the handling of language data; (iii) the impulse from writing didactics that would define needs in terms of analytical tools, as Perl's requisite demonstrated in 1979. Indeed, to a significant extent, the first works regarding the procedural dimensions of writing were conducted from a didactic perspective, beginning with those of Murray (1972), Sommers (1980) or, if we go back further, those of Tressler (1912, cited by Roussey 1999) or even of Albalat in 1903. The concern of these authors was to respond to the needs that they noticed among their students. With Bereiter & Scardamalia (1987, *passim*) and in France with Charolles (1986) and Garcia-Debanc (1986), we move from a simple didactic perspective to a genuinely didactic treatment[2] of the production of writing.

While the act of writing is viewed as a focal point for linguists as well as psychologists and educators, the approaches employed to describe this object stem from research traditions that are quite distinct from each other. Thus, they differ in terms of the concepts that structure them and the methods that they use. Certainly they agree that it is best to view writing with respect to its dynamics, thus supposing that one is interested in the way in which the act of writing takes place in time. However, the analyses of the temporality of writing that they put forward may not be entirely superimposed. Indeed, the temporality of writing is subject to very different interpretations depending on the point of view assumed, the time span taken into consideration, granularity, and the type of indices selected. More

---

2.   In France, "didactics" is a domain of research which deals with the theory of teaching and learning in relation to the content of the various academic disciplines.

precisely, one's conception of the act of writing is closely related to the way in which its temporality is examined. As a result, besides agreement on the fact that writing is a dynamic phenomenon, many differences remain among the conceptions of its temporality and necessarily among the points of reference that are chosen. We will examine these differences not in an effort to reduce them, but rather because it seems that the questions they raise make them epistemologically fruitful.

2.   Observing the act of composition: From ideal temporality to reconstructed temporality

2.1   From the ideal image of writing toward the search for observable elements

The attention given to the temporal dimension of writing was first of all linked with a prescriptive goal: determining the order of operations that allowed the writer to attain the best written results. The scholarly tradition, inspired by classic rhetoric, thus codified a list of steps that the writers were supposed to follow to best complete the writing assignment given to them. This ideal order of steps proceeded from a rational analysis of the act of writing that was not based on empirical observations, but reconstituted *in abstracto* the writing process and identified all the moments where the writer must take a decision. This step-by-step conception of writing resulted not only in pedagogical prescriptions, but it was also subject to theorizations and formalizations. Among these we may cite the model elaborated by Rhoman and Weckle (1975), on which Sommers (1980) composed a virulent critique. We can additionally make reference to the model of Britton et al. (1975, 29) that proposed a schema containing three steps *conception/incubation/production*, which was inspired by Jakobson's communication schema (1960).

The inversion of the perspective began when tools became available to observe the actual activity of the writer or, more precisely, to identify the physical manifestations and the written marks that could be interpreted as the indices of this activity. The very first tools utilized in these types of investigations were born as a result of new and even innovative uses of tools that existed beforehand but that until then had served other objectives and in a different context. Thus, the technique of introspection that had been in common use for a long time in the paradigms of non-behavioral psychology was again made use of in certain research protocols; additionally, the procedures of manuscriptology, until that time used to establish the *lectio melior*, were again used to examine written productions. Following this, new needs became clear, and specific tools were created to supplement or complement those that were already available. These tools are not simple technical

additions: these are either purely conceptual tools – this is the case, for example, for the linguistic categorizations used in the analysis of written marks – or tools that combine a technical instrument and a system of interpretation (Alamargot et al. 2007; Leblay & Caporossi 2014). In any case, as with all technological progress, technical advances were always inseparable from conceptual developments. In contrast to common belief, this link between technology and conceptualization is reciprocal: the arrival of a new question leads to the need for material means to address it; but in return, the principle of affordance means that the provision of a new tool will suggest new uses and thus lead to the emergence of new ways of using them (Plane 2014). This is why tools, methodologies, and representations related to the act of writing evolved together. Therefore, two modes of investigation, one founded on empirical research and the other inclined toward formalization or the exploration of hypotheses, interacted with each other.

2.2   From the analysis of observable elements toward the reconstruction of cognitive activity

Regarding the methods used in the field of cognitive psychology, we may note along with Olive (2002, 2010) that the research has a relatively stable repertoire of investigative techniques that allow verbal production to be studied in real-time. These protocols define a specific task that the writer must accomplish under controlled conditions and which, depending on the needs of the study, may be perturbed or placed in competition with one or more other tasks. The temporal progression, the linguistic forms produced, and potentially movements as well are recorded and provide much data. These protocols have two characteristics in common:

–   as for practicality, not only are the parameters that determine the execution of the task defined with precision, but also the nature of the linguistic product to be provided is even controlled in such a way that the observation may address an isolated and well-targeted element. The linguistic product may even be totally fixed in advance – it may be, for example, a task to copy words or part of a text – thus allowing the observation to focus on other components of the task.
–   the target in the prescribed task is the use of a particular cognitive process or, most frequently, of a component of this process. The linguistic dimension of the product is only of interest to the extent that it provides information about the way in which this process is performed.

These experimental protocols are associated with online devices that can record very subtle signs, such as pauses or ocular movements, and thus provide the material for

a chronometric analysis. The collected information is considered as providing indices of the compositional act or, more precisely, indices of the way in which different operations are carried out, thus making it possible to analyze, for example, the way that written or morphographical tasks are performed in the production of writing.

As they pertain to epistemology, these works present a particularity that is not found in other domains where composition is studied: homogeneity of the theoretical frame on which this research is based. This homogeneity should not be interpreted as an effect of the communal cohesion of the researchers in this area. Indeed, this is not simply the result of sharing knowledge and methods via professional channels (i.e. journals, conferences, etc.), rather it is a real consensus about the research issues: following the publications of Kellogg (1996) and Hayes (1996), the nodal point of all of these works rest on the idea that we will only be able to understand (and, if possible, improve) the cognitive-linguistic functioning involved in the production of writing if we are able to account for the functioning of memory and its limits (Kellogg, Olive & Piolat 2007). As a result, experiments were conducted with the goal of testing modelizations of the mnesic system, particularly the modelization developed by Baddeley (1986), in order to refine them. This is done by shedding light on the competition between the procedures (Fayol & Pacton 2006) or by examining a particular dimension of the production of writing (for example, the visual-spatial dimension (Olive et al. 2010) or by examining the performance of a population that has certain particularities (Bourdin 1999; Richards, Berninger & Fayol 2012; McCutchen 2011).

2.3   From the identification of traces toward the articulation of new questions

In the field of linguistics, it is customary of epistemology to exploit confrontation and divergence to make conceptual progress rather than to seek consensus; this means that the theories used to analyze linguistic data will not ever be unequivocally unified. However, the methodologies for collecting corpora have been stabilized, and the list of the available descriptors has been well established.

In this context, the writing – and, more broadly speaking, all of the signs written during the creation of these writings – have been studied not because they allowed a reconstruction of a cognitive mechanism, but because they provide information about the functioning of language itself or about the use of language by a particular writer-subject.

This is why the same written sign can be the subject of different investigations by psychologists and by linguists. For the one group, it provides information about a process, which is to say about a temporal sequence of operations and the states achieved, whether or not the accomplishment of these operations

supposes an act of decision on the part of the writer (automatic processes/controlled processes, to take up the classic distinction made by Schneider & Shiffrin 1977); for the others, it is examined for the information that it provides about a linguistic operation or a linguistic mechanism. We may take the act of deletion as a concrete example that illustrates this difference in perspective. Indeed, deletions or stopping during production are interpreted in the same way by psychologists and linguists as signs indicating a difficulty experienced by the writer. However, for the psychologist, the examination of these phenomena, the conditions in which they are produced, and their frequency may serve, for example, to verify whether certain operations are performed sequentially or simultaneously. By contrast, for the linguist, each deletion must be examined within its textual and/or unique discursive context. Depending on the semantic or morphosyntactic properties of each crossed-out segment and its context, the linguist will see in the deletion the index of a change in point of view (Fenoglio 2012), the indication of an internal dialogism (Boré 2013), the signs of an intentionality that requires interpretation (Pétillon 2006), or more broadly the sign of the non-coincidence of the discourse with itself, to take up the formula provided by Authier-Revuz (1995, 235).

Certainly the divergences on the nature of the subject-writer that we are interested in – a cognitive subject whose language activity lends itself to quantification *vs.* a unique individual who manipulates a complex object with indefinite contours, i.e. language, in an idiosyncratic way – in no way prevents pluridisciplinary research, for example, as is shown by the works that address the management of orthography (Bourdin, Cogis & Foulin 2010) or coherence (Chuy & Rondelli 2010). However, they indicate that a parameter of the linguistic activity that the two approaches take into account, specifically the temporality of writing, may be viewed in radically different ways. Therefore, it is the very notion of the dynamics of writing that is brought into question.

3.   Toward a plural definition of the dynamics of writing

3.1   Atemporality/sequentiality/continuity

We suggested above that one of the traits characteristic of the dynamics of writing is that it takes place within time. However, Bronckart (1996, 112) demonstrates that linguistic activity can be described by removing its temporality. In fact, we may take account of a process, and thus of a dynamic, by examining it only from the perspective of the forces set in motion, the equilibriums and the disequilibriums that are established or removed. In this case, it is an approach that could

be qualified as vectorial by analogy with the works of physicists who examine phenomena by looking at opposing forces.

This atemporal approach to the compositional dynamics – which is not, by the way, the approach adopted by Bronckart – contradicts the approach used by cognitive psychologists who rely on the observation of the temporal progression of the production of writing to model the cognitive architecture, and it also contradicts the approach taken by linguistics who endeavor to reconstitute the genesis of a text. However, it has the merit of attracting attention to a neglected fact by means of the chronometric analyses of the verbal production, i.e. the cognitive processes at work in the production of writing and the linguistic operations do not have an identical relationship with time. The cognitive processes are considered to develop in a definite and measurable time span. We may define them and analyze the temporal waning of the written production (Chanquoy, Foulin, & Fayol 1990; Olive 2011). On the other hand, the linguistic mechanisms, such as mechanisms of anchoring or aspectualization (Adam 2011, 174) are continuous, and the moment that they are produced cannot be isolated. Certainly when examining the production of a child, we easily identify the moment where this mechanism breaks down when a dysfunction appears. This happens, for example, when a written account in the third person suddenly changes and becomes an account written in the first person, most frequently following the insertion of a dialogue that breaks with the linearity of the narrative. However, aside from these exceptional cases, the maintenance, or dissimulation, of an enunciative agency, whether homogenous or heterogeneous, is nevertheless a continuous phenomenon, even if the initial choice of this agency can be localized (Rabatel 2012).

In the examination of the temporality of writing, we must take into account not only the succession or the recursion of the phenomena, but also the fact that certain phenomena are continuous and others are discrete, to take up a distinction belonging to the mathematicians.

3.2   Time of the writer/time of the text

We are putting forward an additional factor to consider in the analysis of the dynamics of writing: the state of the text on which the writer is working (Plane, Alamargot & Lebrave 2010). The text that the writer composes is a an object in a state of change since it is in the process of being elaborated, and the instability of its state is thus a noteworthy element in the act of writing.

The act of producing a text requires the writer to adapt linguistic segments in such a way that their assembly creates "text". To satisfy this condition, this assembly must possess a certain number of properties. Some of these are extrinsic or transitional, that is to say they are more or less dependent on the context in which

the text is received. For example, this is the case for significance, which relies on the capacity of the reader to interpret the text and to construct a representation of its referential content. Others are intrinsic, and thus independent of the actualization that a virtual reader would carry out; these intrinsic properties should be the writer's primary concern (Lundquist 1999; Plane 2006). Among these properties, the most notable are cohesion, which means that the meaning is passed beyond the segments that carry it, and coherence, related to the fact that the text takes an object for itself to represent, and then constructs this representation. However, these properties are only observable in the completed text. While the writer is composing the text, the completed text is of course not yet available; the writer only has a vague, colorless image that will become more precise as the writer finds the words and expressions that complete the rough draft of the text which is still in a state of change. Over the course of writing, the author does not have the necessary means to judge the coherence of the entirety of the text that he or she is writing. This is because the text does not pre-exist its actualization, and this actualization only comes through the process of composition. Even worse, when the text is long, the writer does not work based on the text that is before his or her eyes, but based on the memory of what he or she has composed. This phenomenon explains the referential incoherencies in a certain number of well-written literary works. Thus, despite the great attention given by Flaubert to his text, the color of Madame Bovary's eyes changes over the course of the eponymous novel. In other words, to return to the definition given above, writing is not only adapting linguistic segments among themselves such that their assembly creates a text; rather it is adjusting linguistic segments among themselves and also doing so with a more or less accurate memory of what was already written and in anticipation of the elements that will materialize later (Plane, Rondelli & Vénérin 2014). Indeed, this device evolves during the compositional process since the text develops incrementally and undergoes variations, thus adding a factor that makes the dynamics of writing even more complex.

3.3   Observable time, inaccessible time

The majority of the works that address the dynamics of writing precisely define the time span that is of interest to them. In the works that seek chronometric information, the time span taken into consideration is either determined in advance by the experimental protocol (Leblay 2007; Leblay & Caporossi 2014, 13) or delimited by the recording device: we may consider that writing begins at the moment where the recording starts and that it finishes as soon as the recording stops. In this case, it is most frequently the writer that determines the temporal limits by starting up the computer that is equipped with a means of recording and then by terminating the work session (Leitjen, Van Waes, Schriver & Hayes 2014).

But the analyst may also be interested in a larger time span and include the various moments of preparation leading up to the actual composition process. In this case, the analyst will create a genetic file[3] that contains all of the documents showing that the author – in general, this is a writer – has begun his or her writing project (Grésillon 1994, 107). This file is therefore composed of material elements that allow the analyst to reconstruct the steps taken in a process that can be very long.

Nevertheless, in both cases, the online collection of data and the creation of a genetic file, an important part of the writing process, evades the analyst. The production of a text does not only take place when the author is writing, and not even when documents and notes are being collected to be used for a composition at a later time. It is also made up of the accumulation of discursive fragments, the acquisition of linguistic experience, the product of which crystallizes during composition. The texts that one has read, the speech that one has heard, as well as the discourses that one has produced are the material, even the rough drafts of future texts. Therefore, one part of the composition takes place before the act commonly referred to as composition. When examining the production of a writer, we may very often see how each book is a re-writing of an older book. The same is true for assignments done by pupils and students; each assignment is more or less a re-writing of another assignment. This shows a part of the writing process that takes place before the observable period of composition. Additionally, once the period is added to the final sentence, there is no guarantee that the act of writing will stop: the writer may continue to mentally work on the text that he or she has finished and put away. We could also say that the text continues to work on its author.

4.   Conclusion: Epistemological investigations to pursue

The question of the dynamics of writing is particularly complex because it involves concepts that are difficult to define: time, intention, writing. Here we will pass over the difficulties associated with defining writing since the volume to which this chapter belongs is fully dedicated to these problems. Intention, or rather intentionality, is a subject of debate as much for linguists, as shown by Pétillon (2006), as for philosophers because it brings into play a representation of the subject that is only of interest in certain research paradigms. Time is an immaterial and experiential phenomenon. It is intrinsically continuous and indivisible, but for convenience we cut it into small, measurable units. In other words, beyond the methodological and theoretical difficulties that were mentioned above, we are constrained to the

---

3.   In French, such a file is referred to as a "*dossier génétique*". This file contains all of the documents relating to the research and preparation of a writing project (e.g. researched papers and documents, notes, drafts, etc.). The genetic file is part of genetic criticism (mentioned above).

use of imperfect artifacts just in order to consider the idea of the dynamics of writing. The desire to define the dynamics of writing is thus, in a sense, a vain attempt. However, it is also a fruitful means to study the act of writing.

References

Adam, Jean-Michel. 2011. *La linguistique textuelle: introduction à l'analyse textuelle des discours*. Paris: Armand Colin.

Alamargot, Denis, Christophe Dansac, David Chesnet, and Michel Fayol. 2007. "Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production." In *Writing and Cognition: Research and Applications*, ed. by Mark Torrance, Luuk van Waes and David Galbraith, 13–29. Amsterdam, Elsevier.

Albalat, Antoine. 1903. *Le travail du style enseigné par les corrections manuscrites des grands écrivains*. Paris: Armand Colin.

Authier-Revuz, Jacqueline. 1995. *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidence du dire*. Paris: Larousse. DOI: 10.3406/litt.1996.2426

Baddeley, Alan. D. 1986. *Working Memory*. New York: Oxford University Press. DOI: 10.1017/s0033291700025228

Bereiter Carl, and Marlene Scardamalia. 1987. *The Psychology of Written Composition*. Hillsdale, N.J.: Lawrence Erlbaum. DOI: 10.2307/327270

Boré, Catherine. 2013. "L'énonciation dialogique au principe de la créativité du sujet scolaire." In *L'école, l'écriture et la création. Études françaises et brésiliennes,* ed. by Catherine Boré and Eduardo Calil, 55–74. Louvain-La-Neuve: L'Harmattan.

Bourdin, Béatrice. 1999. "Mémoire de travail et production langagière: comparaison de l'oral et de l'écrit chez les adultes et les enfants." *L'année psychologique* 99(1): 123–148. DOI: 10.3406/psy.1999.28551

Bourdin, Béatrice, Danièle Cogis, and Jean-Noël Foulin. 2010. "Influence des traitements graphomoteurs et orthographiques sur la production de textes écrits: perspective pluridisciplinaire." *Langages* 177: 57–82. DOI: 10.3917/lang.177.0057

Britton, James, Antony Burgess, Nancy Martin, Alex Mc Leod, and Harold Rosen. 1975. *The Development of Writing Abilities* (*11–18*). London: Macmillan Education and Welfare.

Bronckart, Jean-Paul. 1996. *Activité langagière, textes et discours. Pour un interactionnisme discursif*. Lausanne: Delachaux et Niestlé.

Chanquoy, Lucile, Jean-Noël Foulin, and Michel Fayol. 1990. "Temporal management of short text writing by children and adults." *European Bulletin of Cognitive Psychology* 10: 513–540.

Charolles, Michel. 1986. "L'analyse des processus rédactionnels: aspects linguistiques, psychologiques et didactiques." *Pratiques* 49: 3–21.

Chuy, Maria, and Fabienne Rondelli. 2010. "Traitement des contraintes linguistiques et cognitives dans construction de la cohérence textuelle." *Langages* 177: 83–112. DOI: 10.3917/lang.177.0083

Fayol, Michel, and Sébastien Pacton. 2006. "L'accord du participe passé: entre compétition de procédures et récupération en mémoire." *Langue française* 151: 59–73. DOI: 10.3406/lfr.2006.6774

Fenoglio, Irène. 2012. "L'abyssal *in vivo* de l'écrire. Mise en abyme scripturale chez Pascal Quignard." In *Écritures du XXIᵉ siècle. Genèses* in vivo, ed. by Claudia Amigo Pino, Irène Fenoglio and Veronica Galindez-Jorge, 117–146. Paris: Le manuscrit.

Flower, Linda, John R. Hayes, Linda Carey, Karen Schriver, and James Stratman. 1986. "Detection, Diagnosis, and the Strategies of Revision." *College Composition and Communication* 37(1): 16–55. DOI: 10.2307/357381

Garcia-Debanc, Claudine. 1986. "L'Intérêt des modèles du processus rédactionnel pour une pédagogie de l'écriture." *Pratiques* 49: 23–49.

Grésillon, Almuth. 1994. *Éléments de critique génétique. Lire les manuscrits modernes*. Paris: Puf. DOI: 10.1515/arbi.2000.18.1.1

Halliday, Michael A. K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman. DOI: 10.2307/412795

Hay, Louis. 1979. "La critique génétique. Origine et perspectives." In *Essais de Critique Génétique*, ed. by Louis Hay, 227–236. Paris: Flammarion.

Hayes, John R. 1996. "A new framework for understanding cognition and affect in writing." In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, ed. by Michael C. Levy, and Sarah Ransdell, 1–28. Mahwah: Lawrence Erlbaum Associates.

Hayes, John. R., and Linda S. Flower. 1980. "Identifying the organization of writing processes." In *Cognitive Processes in Writing*, ed. by Lee W. Gregg, and Erwin R. Steinberg, 3–30. Hillsdale, NJ: Lawrence Erlbaum. DOI: 10.1177/002194368201900207

Jakobson, Roman. 1960 (trad. 1963). *Essais de linguistique générale*. Paris: Seuil.

Kellogg, Ronald T. 1996. "A model of working memory in writing." In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, ed. by Michael C. Levy, and Sarah Ransdell, 57–71. Mahwah: Lawrence Erlbaum Associates.

Kellogg, Ronald T., Thierry Olive, and Annie Piolat. 2007. "Verbal and visual working memory in written sentence production." In *Writing and Cognition: Research and Applications*, ed. by Gert Rijlaarsdam, Mark Torrance, Luuk van Waes, and David Galbraith (Studies in Writing 20), 1–12. Amsterdam: Elsevier.

Kintsch, Walter, and Teun van Dijk. 1978. "Toward a model of Text comprehension and production." *Psychological Review* 85(5): 363–394. DOI: 10.1037/0033-295x.85.5.363

Leblay, Christophe. 2007. "L'avant-texte comme texte sur le vif. Analyse génétique d'opérations d'écriture en temps réel." *Langue française* 155: 101–113. DOI: 10.3917/lf.155.0101

Leblay, Christophe, and Gilles Caporossi (Eds.). 2014. *Temps de l'écriture. Enregistrements et représentations*. Louvain-La-Neuve: L'Harmattan.

Leitjen, Mariëlle, Luuk Van Waes, Karen Schriver, John R. Hayes. 2014. "Writing in the workplace: constructing documents using multiple digital sources." *Journal of Writing Research* 5(3): 285–337.

Lundquist, Lita. 1999. "Le Factum Textus: fait de grammaire, fait de linguistique ou fait de cognition?" *Langue française* 121: 56–75. DOI: 10.3406/lfr.1999.6279

Matsuhashi, Ann. 1982. "Exploration in the real-time production of written discourse. "In *What Writers know. The Language Process and Structure of Written Discourse*, ed. by Martin Nystrand, 269–290. New York: Academic Press. DOI: 10.1017/s0047404500010095

McCutchen, Deborah. 2011. "From novice to expert: implications of language skills and writing-relevant knowledge for memory during the development of writing skill." *Journal of Writing Research* 3(1): 51–68.

Murray, Donald. M. 1972. "Teach writing as a process not product." *The Leaflet*: 11–14.

Olive, Thierry. 2002. "La gestion en temps réel de la production verbale: méthodes et données." In *Traité des Sciences Cognitives: Production du langage,* ed. by Michel Fayol, 131–147. Paris: Hermès.

Olive, Thierry. 2010. "Methods, tools and techniques for the on-line study of the writing process." In *Writing: Processes, Tools and Techniques*, ed. by Nathan L. Mertens, 1–18. New York: Nova Publishers.

Olive, Thierry. 2011. "L'analyse du décours temporel de la production écrite." In *Sciences du langage et nouvelles technologies,* ed. by Thierry Ponchon, and Isabelle Laborde-Milaa, 61–68. Limoges: Lambert-Lucas.

Olive, Thierry, Jean-Michel Passerault, Jean-Louis Lebrave, and Nathalie Le Bigot. 2010. "La dimension visuo-spatiale de la production de textes: approches de psychologie cognitive et de critique génétique." *Langages* 177: 29–55. DOI: 10.3917/lang.177.0029

Perl, Sondra. 1979. "The Composing Processes of Unskilled College Writers." *Research in the Teaching of English* 13(4): 317–336.

Petillon, Sabine. 2006. "Intentionnalité et révision à l'épreuve du manuscrit d'écrivain. Genèse de *Génitrix* de Mauriac." *Langages* 164: 26–42. DOI: 10.3917/lang.164.0026

Plane, Sylvie. 2006. "Singularités et constantes de la production d'écrit – L'écriture comme traitement de contraintes." In *Didactique de l'écrit. La construction des savoirs et le sujet-écrivant*, ed. by Jacqueline Laffont-Terranova, and Didier Colin, 33–54. Namur: Presses Universitaires de Namur.

Plane, Sylvie. 2014. "Questions posées par les rapports complexes entre l'instrument et son utilisateur. L'exemple des outils d'écriture." In *Enseigner les langues aux enfants en contexte scolaire. Diversité des approches et outils d'enseignement*, ed. by Malory Leclère, and Jean-Paul Narcy-Combes, 87–103. Paris: Riveneuve.

Plane, Sylvie, Denis Alamargot, and Jean-Louis Lebrave. 2010. "Temporalité de l'écriture et rôle du texte produit dans l'activité rédactionnelle." *Langages* 177: 7–28. DOI: 10.3917/lang.177.0007

Plane, Sylvie, Fabienne Rondelli, and Christine Vénérin. 2014. "Variations, fidélité, infidélité: l'écriture et la réécriture de discours rapportés par de jeunes scripteurs." In *Le discours rapporté: approches linguistiques et perspectives didactiques*, ed. by Cécile Desoutter, and Caroline Mellet, 215–232. Berne: Peter Lang.

Rabatel, Alain. 2012. "Positions, positionnements et postures de l'énonciateur." *Travaux Neuchatelois de Linguistique* 56: 23–42. DOI: 10.11606/issn.2236–4242.v26i2p159-183

Richards, Todd, Virginia Berninger, and Michel Fayol. 2012. "The writing brain of normal child writers and children with writing disabilities: Generating ideas and transcribing them through the orthographic loop." In *Writing: A Mosaic of Perspectives and Views*, ed. by Elena Grigorenko, Elisa Mambrino, and David Preiss, 85–105. New York: Psychology Press.

Rohman, Gordon, and Albert O. Wlecke. 1975. "Pre-writing: the construction and application of models for concept formation in writing". Cooperative Research Project 2174. U.S. Office of Education. In *The Development of Writing Abilities*, ed. by James Britton, Antony Burgess, Nancy Martin, Alex McLeod, and Harold Rosen, 11–18. London: Macmillan Education.

Roussey, Jean-Yves. 1999. *Le contrôle de la rédaction de textes. Perspective cognitive*. Dossier d'habilitation à diriger des recherches. Université d'Aix.

Schneider, Walter, and Richard M. Shiffrin. 1977. "Controlled and automatic human information processing: 1. Detection, search, and attention." *Psychological Review* 84: 1–66. DOI: 10.1037/0033-295x.84.1.1

Sommers, Nancy. 1980. "Revisions Strategies of Student Writers and Experienced Adult Writers." *College Composition and Communication* 31(4): 378–388. DOI: 10.2307/356588

Tressler, Jacob C. 1912. "The efficiency of student correction of composition." *English Journal* 1: 405–411. DOI: 10.2307/801436

Methodology

Investigating real-life writing processes*

Almuth Grésillon & Daniel Perrin

ITEM-ENS, France / Zurich University of Applied Sciences,
Department of Applied Linguistics

Doing writing research in real-life settings means investigating individual, collaborative, and organizational writing and text production in complex and dynamic contexts. Methodological questions need to be clarified, such as: which method fits which problem and how should and can various methods complement each other? – In our paper, we start from two methodologically complementary approaches of doing research into real-life writing processes (part 1). Based on these approaches, we then explain why collecting data represents a key challenge in the history of writing research (part 2). Finally, we outline a typology of state-of-the-art methods in writing research that can be combined to face this methodological challenge (part 3).

Keywords: Progression analysis; Text Genetics; Idée suisse; writing; methodology

1. Lebensfahrt, Idée suisse, and *AL-informed writing research*

Throughout this chapter, we use two methodologically complementary approaches to research into writing processes to illustrate what we mean by methodology and methods of writing research that is informed by Applied Linguistics (AL):

– In the Lebensfahrt case, the genesis of Heinrich Heine's four-stanza poem "Lebensfahrt" (1843) is analyzed in depth as an individual author's genuine writing process (Grésillon 1987; Grésillon 2014). Of course, the German poet

---

* This article draws on existing publications by the authors. Paragraphs and formulations have been reproduced from the following papers without explicit cross-references: Perrin, Daniel (2013). *The Linguistics of Newswriting*. Amsterdam, New York et al.: John Benjamins. Grésillon, Almuth & Perrin, Daniel, (2014). Methodology. From speaking about writing to tracking text production. In Daniel Perrin and Eva-Maria Jakobs (Eds.), *Handbook of Writing and Text Production*. New York et al.: De Gruyter.

worked at an analogue workplace, using pen and paper. Moreover, and in contrast to other writers, he avoided all kinds of written metadiscourse, such as comments in side notes. Finally, it seems that no correspondence with peers and publishers has been preserved. Thus, the main data sources are material traces in manuscripts and biographical data. The analysis of this data is oriented towards a better understanding of a literary text's reconstructed genesis – and of (literary) writing in general.

– In the IDÉE SUISSE project, in contrast, the analysis focuses on the interplay of language policy, norms, and practice in the newsrooms of an entire public service media organization. Using logging and screen recording software, journalists' collaborative writing activities were recorded. In addition, editorial conferences and negotiations with peers such as video editors and cameramen were videotaped. Finally, writers, media managers, and policy makers were interviewed and policy documents were analyzed, following the principles of Progression Analysis (see below, part 3.2). The research project aimed to understand and develop the broadcaster's competence to fulfill its public mandate.

The poet's and the journalists' writing both represent relevant cases for applied linguistics (AL). As a "user-friendly linguistics" (Wei 2007, 117), AL has always been oriented towards practice with a twofold goal: understanding and improving language use. From a production perspective, it deals with the reflection and optimization of speaking and writing for certain communicative tasks and domains, including language learning or workplace communication (e.g. Cicourel 2003; Alatis, Hamilton & Tan 2002; Candlin 2003). AL can investigate the repertoires of strategies and practices that individuals or language communities use when they make linguistic decisions (e.g. Cook 2003, 125; Zhong & Newhagen 2009) in discussions or writing processes. Then, these repertoires can be expanded through knowledge transformation processes, e.g. in training, coaching, and organizational development.

In the present chapter, we thus conceive AL-informed writing research as a joint activity of researchers, practitioners, and society at large. They collaborate to investigate (i) individual or collaborative writing (ii) as material, mental, and social activity (iii) in analogue or digital environments, (iv) ex post or in situ, (v) in order to understand and improve it. This understanding of writing research has consequences for the methodological design of research projects.

i. Investigating individual or collaborative writing: Depending on the research object and underlying key concepts such as authorship (Schindler & Wolfe 2014), AL-informed writing research investigates the activity of subjects of varied complexities. They range from individuals to peer groups and entire organizations in complex contexts. Suitable methods enable researchers to

capture and analyze the corresponding activities. Eye tracking, for example, can capture pupil movements in highly computerized settings. They are interpreted as shifts of the focus of attention by individual human text processors. Analyzing a novelist's remarks at the margin of a manuscript can point towards her or his individual decisions and activities. In contrast, comparing versions of an organization's editorial guidelines over time reveals the big picture of their evolving explicit quality discourse.

ii. Investigating writing as mental, material, and social activity: Writing takes place within and between people, as well as at their physical interface (see also Prior this volume). Analyzing inner, mental activities related to writing in natural contexts requires indirect methods and procedures, such as retrospective verbal protocols (e.g. Camps 2003; Ericsson & Simon 1984; Greene & Higgins 1994; Smagorinsky 1994). Some material activities of handwriting, however, leave directly accessible, manifest traces in manuscripts, which offer insights into the material activity of long past writing processes. In in-situ approaches, these material activities can be captured in real-time, using video recording (e.g. Van Waes & Mangen 2012) or keystroke logging (e.g. Flinn 1987; Van Waes & Van Herreweghe 1995; Spelman Miller 2006a; Strömqvist, Holmqvist, Johansson, Karlsson & Wengelin 2006). Social aspects of writing, however, such as balancing workflows and editorial quality discourse in organizations, call for methods such as network analyses or dynamic modeling that capture the complexity of writing on macro levels too.

iii. Investigating writing in analogue or digital workplaces: In computerized environments, most material text production activities such as archive research or editing is performed at computers. The same computers can be used by researchers to automatically collect data, for example about pausing times between linguistic units. As these data are available in digital formats, they can be analyzed using algorithms. In contrast, writing with pen and paper does not leave digital imprints for analyses. Some traces of the writing process, such as crossed-out words, are directly visible in handwritten texts; others, such as pauses, are not. Thus, capturing handwriting calls for meticulous manuscript collection or rather intrusive methods like videotaping or observing – and the related interpretative analysis.

iv. Investigating writing ex post or in situ: Knowing in advance that one would like to investigate a particular kind of contemporary writing processes puts researchers in a comparably comfortable position: They can develop a methodology and choose methods and recording procedures that capture as many as possible of the relevant aspects of text production in appropriate depth and breadth. Then, the field can be prepared according to the research question, e.g. by building trust within the organization and installing logging software. In contrast, researchers who analyze a 19th century novelist's writing process

have to limit themselves to available traces from a field that was not designed to support research.

v.   Investigating writing in order to understand and improve it: Whoever analyzes the genesis of a poem written in 1843 will not be motivated by the idea of helping the author ameliorate his or her writing process. Thus, methods applied focus on diagnosis. Of course, the knowledge generated can be applied later to teach young poets in data-based courses of creative writing, but usually such applications are not part of theoretically-driven research. In contrast, in transdisciplinary action research, interventions are considered crucial procedures. Writing at specific workplaces is investigated mainly in order to improve it, for example by elaborating the writers' repertoires of text production practices.

2.   The double black box: A brief history of investigating writing in the field

In the next paragraphs, we use the Lebensfahrt and the Idée suisse case to illustrate why collecting data has represented a key challenge in the history of AL-informed writing research. We then formulate quality criteria for selecting methods that help gather relevant information about real-life writing. Guided by these criteria, we will develop a typology of the field's state-of-the-art methods (3).

Investigating writing confronts researchers with a problem we term the *double black box*. First, written language is mostly presented as a finalized product, detached from all traces of genesis such as insertions and deletions. The written (the product) aims at overcoming the writing (the process). Whereas conversations overtly evolve over time, naturally observable for both participants and researchers, writing processes are performed as "back stage" (Goffman 1959) activities, hidden away from the addressees. In consequence, they are hardly accessible for researchers. Regarding this first, outer black box, analyses of writing processes inherently differ from conversation analyses.

However, the metaphor of the double black box points at a second, inner box, too. Once researchers manage to shed light on the backstage processes of writing, what they capture is material activity only, just as with turn-taking and repair in conversation analysis. In individual writing, the material activity is limited to physical behavior – the activity of processing signs on screens and papers in co-adaptive contexts. In collaborative writing, material activities include writers' negotiations about the task to be solved and its context. Analyses of the mental and social levels of writing processes have to precisely draw on evidence from these material activities as the main source of natural data.

Regarding this second, inner black box, writing research finds itself confronted with the methodological core problems of all AL-research: Language use

allows for a highly differentiated, yet indirect view of mental and social structures and processes. In ex-post approaches of writing research, researchers may indirectly access a long dead poet's mind and context by analyzing a corpus of manuscripts with changes and meta-communicative comments – and by interpreting a piece writers wrote about their own writing. In-situ approaches, in contrast, allow researchers to query writers about their writing. After finalizing a text version, writers can be asked to view recordings of their text production processes and to comment on these activities. However, the access to the mind remains indirect, based on subjects' own explanations about what they have in mind and are able and willing to share with the researchers.

So it is mainly the first black box that stopped researchers for a long time from investigating writing processes in natural contexts. Linguistics first focused on written language from a product perspective, later it described conversations as processes, and only then rediscovered written language from a process perspective. But writing is usually still investigated from a product perspective, without "empirical ethnographic considerations" (Widdowson 2000, 22). In the programmatic introduction to their collection of early approaches to writing processes, Gerd Antos and Hans Peter Krings assumed that the "analysis of text emergence, including drafts, versions, and revisions, is an approach which basically is feasible and worthwhile for non-literary texts too. […] Empirical analyses of text geneses would be an important contribution for a clearly linguistically motivated text theory" (Antos 1989, 36, translated from German; see also Krings 1992).

What Krings and Antos had in mind[1] when they – albeit indirectly – referred to process analyses of literary texts is Genetic Criticism (e.g. Grésillon 1994; Grésillon 2008a; Grésillon & Lebrave 2008; Hay 2002; de Biasi 2011; Ferrer 2011; Lebrave 1987; Lebrave 1992). In this research framework, the object of analysis is the literary manuscript, with "the trace of a dynamic of the text in the making". The methods applied "reveal the body and the course of writing in order to construct a series of hypotheses on the operation of writing" (Grésillon 1997, 106). Empirical evidence and plausibility of interpretations complement each other when, based on material traces, writing processes are reconstructed ex-post, with archeological accuracy, in order to better understand the final product and, most importantly, the writing process itself in the light of its mental, material, and social emergence.

A key driver of Genetic Criticism (e.g. Grésillon & Mervant-Roux 2010) is the acquisition of Heinrich Heine's manuscripts by the Paris National Library in 1966. In 1968, a research group was commissioned to analyze these manuscripts. First of all, an appropriate method had to be developed – Genetic Criticism. It allows

---

1.   Personal communication between Gerd Antos and Daniel Perrin, Zurich, 12 September 2008.

researchers to reconstruct the genesis of literature based on preserved traces of the writing process. Depending on the author, these traces can include notes and excerpts from sources, such as dictionaries in the case of Francis Ponge (Grésillon 2008c) or historical and geographic sources in the case of Flaubert (Grésillon 2008b; Grésillon, Lebrave & Fuchs 1991). Other examples of traces are drafts, outlines, plans, first versions, revised versions, final versions, first editions, and revised editions. Beside these autographs, auto-biographic and biographic material can be collected, such as correspondence, interviews, diaries, and third persons' reports referring to the genesis of a text. The entire collection is termed *genetic dossier* or *avant-texte*. The method of genetic criticism draws, inter alia, on concepts of modern linguistics (Grésillon & Lebrave 2008). By and by, it has been broadened to be applied to non-literary texts and non-verbal works of art (Grésillon 1994).

In the LEBENSFAHRT case, where only a few manuscript pages escaped from various kinds of cleansing, a reconstruction of the production process has to focus on analyzing the traces on the preserved manuscripts – and on interpreting them, at a macro level, in their biographical, socio-historical, and political context. Taking into account the slight change in handwriting and a larger line space and indent after the third stanza, it looks as if the fourth stanza on this oldest preserved manuscript, from 1843, was written down later than the first three ones. This hypothesis is strongly supported by the fact that, in 1933, the newspaper "Neue Zürcher Zeitung" printed an even older, three-stanza version of this poem. What the newspaper referred to was a manuscript which must then have been destroyed in Nazi book-burning. Genetic Criticism, therefore, combined methods of manuscript analysis with methods of media analysis in order to reconstruct and understand the genesis of the poem.

What Krings and Antos demanded explicitly, however, started to be realized with key logging and notational systems. They facilitate the step-by-step analysis of the dynamics of (digital) writing. In the 1990s, a Swedish research group around Kerstin Severinson Eklundh and Py Kollberg developed a research tool combining both: The text editor J-Edit allowed for the automatic keylogging of writing processes at computers, whereas the analysis software Trace-it transformed the logging data into S-notation. This notational system allows for human in-depth analyses of writers' revision activity (e.g. Severinson-Eklundh & Sjöholm 1991; Severinson-Eklundh & Kollberg 1996; Kollberg & Severinson-Eklundh 2002). Originally developed for laboratory experiments, the software and notation principles were further developed towards the research framework of Progression Analysis (Perrin 2003).

Analyzing what journalists do in a television newsroom, as in the IDÉE SUISSE case, requires methods that capture collaboration in multimodal text production. Biographical interviews and video recordings of news conferences can reveal that,

for example, a highly experienced journalist was told to report on demonstrations in Lebanon, a region he frequently travels in. Recordings of keyboard and screen activities show that he watched a lot of footage, then started to write and soon changed the formulation "voie express" into "voie tranquille", referring to a ferry normally called "voie express" by the locals (see also Prior this volume on the notion of confluence of activities). It is only the use of cue-based retrospective verbal protocols, however, that allows for an empirically-based reconstruction of the journalist's intentions: He wanted to use "voie tranquille" as a leitmotif to foreground the tranquility of the demonstrations in a region western media tend to show as full of violence.

As the two complementary examples of Genetic Criticism and Progression Analysis have foreshadowed and will show in more detail throughout the chapter, investigating writing from a dynamics perspective, be it ex post or in situ, has required and still requires methodological finesse and innovation. Methods that shed light on the backstage processes have to be developed, applied – and evaluated. This is where the question of methodological accuracy arises. If we consider <u>research methods</u> to be theoretically-based procedures for clarifying and answering research questions (e.g. Litosselity 2009), applying them results in a certain reliability and validity of the findings: Methods can be reliable (or not) in providing (or not) valid information about the object of research.

<u>Reliability</u> means that the same answers are obtained when someone else repeats an investigation following the same procedure. It requires precise data collection and analysis based on carefully considered, transparent rules. In AL-informed writing research, such rules have to explain, for example, how the temporality of writing processes is consistently and transparently transformed into spatial representations for analyses. This is why most traditions of writing research classify the dynamics of text production using systems of operations at various levels. Progression Analysis, for example, regards insertions and deletions of text bits as the basic linguistic operations in writing, allowing for variegated sequential combinations such as deleting a word in one place and inserting it somewhere else, whereas Genetic Criticism differentiates between insertion, deletion, substitution, and relocation.

<u>Validity</u> means that similar answers are obtained when the same research question is investigated with various procedures. It can be achieved by procedures that capture the relevant features of a problem, rather than treating it in a methodically convenient but simplistic way. If, for example, the question concerns which form a stretch of language such as "voie express" will take in later texts, it is sufficient to compare the original utterance from the source text with later reproductions. However, if the reasons why an author recontextualizes utterances are to be captured, then cognitive aspects have to be considered too. Finally, if the author

is seen as embedded in a social context, then social aspects and interconnections have to be included.

In the Lebensfahrt case, a comparison of a first conserved version with a later manuscript shows that the German author initially wrote down three stanzas describing a boat trip in Germany as a lovely, romantic experience (albeit one that ended badly), followed by an escape to "beautiful France". Soon afterwards, he expanded the poem adding a fourth stanza, rewriting the first one, and changing key terms throughout the poem (Fig. 1). This resulted in his far more dramatic narration of, and metaphorical reflection on, shipwrecking twice, first at home in Germany, now in foreign and threatening France. Researchers' contextual knowledge about political tensions surrounding the author allows for an evident and valid interpretation of the writing process as taking the author from an initially romantic to a highly political poem.
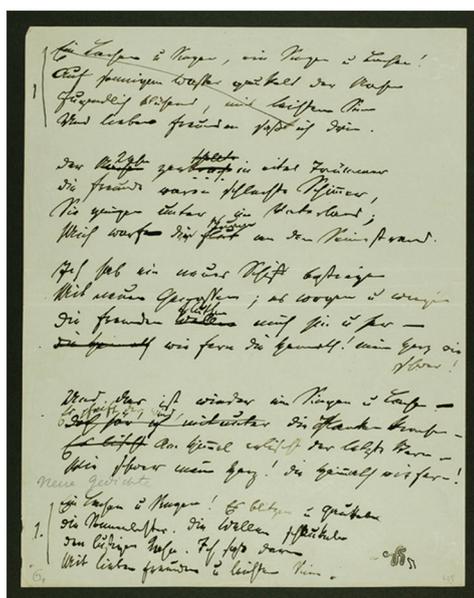
Figure 1. Lebensfahrt Manuscript (left) and transcription (right) showing the insertion of a fourth stanza, the replacement for the first one, and subsequent changes to the second and the third (Grésillon 1987; Grésillon 2014)

In the Idée suisse case, the data analysis allows for the conclusion that the journalist prepared himself for emergence in order to solve a problem he considered difficult. Moreover, the analysis shows how he did so: by telling an attractive, interesting story about peaceful demonstrations. Remaining open to the unexpected, he first exposed himself to a flood of footage, where he stumbled over

the picture of the slowly moving ferry. Later, while writing, the idea emerged to change the ferry's ordinary term/marking, "voie express", into "voie tranquille" – and to use it as a leitmotif: According to the journalist, this term "reverberates in the minds of the demonstrators" (Perrin 2013, 23). In this and similar cases, Progression Analysis shows itself capable of supporting methodologically valid reconstructions of the interplay of routine and emergence in writing (Perrin 2012).

In order to meet the criteria of reliability and of multi-faceted validity, AL-informed writing research aims at combining methods that are rooted in complementary paradigms: linear causality vs. dynamic complexity, verbalization vs. observation, single case studies vs. large samples.

–   Causality vs. complexity: Experiments in the laboratory allow for strict control over isolated parameters related to an object under investigation (e.g. Levy & Ransdell 1996; Whithaus, Harrison & Midyette 2008). The cognitive loads of a specific, isolated writing task for example can be investigated in an experiment where pausing times between linguistic units are measured and interpreted as depending on the state of the mental text processor (e.g. Keseling 1992; Spelman Miller 2006b). On the other hand, laboratory experiments may result in inadequate reduction when the interplay with contexts is too complex to be modeled in linear causal relations (e.g. Sullivan & Porter 1993). When conceiving writing as an activity that is situated in dynamic and complex real-life contexts (e.g. Bracewell 2003; MacMillan 2012; Schneider 2002; Van der Geest 1996), researchers tend to opt for ex-post reconstructions of traces or in-situ ethnographic field studies (e.g. Chin 1994a; Lillis 2008) and dynamic modeling (e.g. Van den Bergh & Rijlaarsdam 1996; Perrin & Wildi 2010) rather than linear experiments. Monocausal relations in real-life writing are limited to non-complex problems such as having access or not to a specific printed source text or data file when writing.

–   Questions vs. observation: Questionnaires can easily be evaluated, and indepth interviews and verbal protocols allow researchers to access mental reflections. On the other hand, such verbal data are closely related to the self-awareness of the people under investigation. In contrast, observation directly captures people's actual activities, but leaves it to the researchers to interpret why those observed do what they do (e.g. Chin 1994a; Cottle 1998). Researchers who are interested in both, the writers' views and their activities, tend to combine verbal and observational methods and data for multi-perspective insights (e.g. Tashakkori & Teddlie 2003; Woolley 2009; Wolfe 2005). As could be shown with Progression Analysis in projects similar to IDÉE SUISSE, journalistic practices of inventing quotes could only be identified by confronting

the journalists under investigation with recordings of their material writing activities on screen.

–   Single case study vs. statistical representation. Mathematically composed, broad samples allow for statistically evident generalizations. On the other hand, the breadth of a data collection limits its depth since research resources tend to be limited; a plethora of cases cannot be analyzed as profoundly as a few well-selected ones (e.g. Abbott, Amtmann & Munson 2006; Schultz 2006). When aiming at in-depth insights into contextualized and therefore complex real-life writing activities, researchers usually decide against purely statistical representation in favor of in-depth analyses of exemplary cases (e.g. Bisaillon 2007; Schultz 2006) such as Lebensfahrt. In the Idée suisse research project, in-depth case studies are combined using qualitative generalization strategies such as grounded theory and theoretical sampling. They allow researchers to develop theoretically reflected mid-range theories by systematically generalizing from a small number of well-selected and thoroughly analyzed cases.

Such general methodological decisions lay the groundwork for selecting and, if necessary, triangulating specific methods of and for writing research (Part 3).

3.   State-of-the-art toolkit: Four complementary types of methods

In this chapter, we outline a typology of four methodological perspectives in AL-informed writing research. Throughout contemporary projects, the methods applied provide empirical evidence of material, cognitive, social, or socio-cognitive aspects of writing. Respective state-of-the-art methods focus, for example, on material differences between text versions (3.1), individuals' writing strategies (3.2), variation of practices within and across organizations' writing (3.3), and communities' metadiscourse reflecting their written communication (3.4).

3.1   The material focus: Tracking intertextual chains with version analysis

First and foremost, linguistics investigates stretches of language in context (e.g. McCarthy 2001, 115). From this material perspective, AL-informed writing research emphasizes the intertextual nature of writing: new texts and text versions are created and differ from earlier ones. Material changes to the linguistic products are captured with version analyses. By <u>version analysis</u>, we understand the method of collecting and analyzing data in order to reconstruct the changes that linguistic features undergo in intertextual chains. The methods and procedures applied originate in comparative text analysis.

Prototype version analyses trace linguistic products (e.g. Sanders & Van Wijk 1996) and elaborate on the changes in text features from version to version, be it at

one single production site or across a series of sites. In the framework of Genetic Criticism for example, close-to-final versions of literary writing were compared (Mahrer 2006) and play writing was tracked from the initial draft to the authors' notes on first performances (Grésillon & Mervant-Roux 2010). In projects similar to IDÉE SUISSE, a quote from a politician's original utterance was traced throughout the intertextual chain of correspondents, local and global news agencies, broadcasters, and the follow-up discourse in social media (Perrin 2011). Other media linguistic studies draw on version analyses to reveal how texts change throughout the intertextual chains (e.g. Van Dijk 1988; Bell 1991, 56 ff.; Luginbühl, Baumberger, Schwab & Burger 2002; Robinson 2009; Lams 2011).

The very minimal variant of version analysis limits the empirical access to one single version, with implicit or explicit reference to other versions that were not explicitly analyzed (e.g. Ekström 2001). This variant of version analysis is widespread in the framework of Critical Discourse Analysis (Van Dijk 2001; see also critiques by Stubbs 1997 or Widdowson 2000).

Another frequent, yet empirically denser, variant of the version analysis focuses on changes performed at one single production site. In the LEBENSFAHRT case, for example, most empirically accessible changes are documented on the so-called "Arbeitshandschrift H1" manuscript (Fig. 1), whereas earlier versions are only indirectly preserved through the 1933 newspaper article (2), and later versions add minor changes only. Similarly, the "voie tranquille" analysis in the IDÉE SUISSE project draws on data from one single site, the TÉLÉJOURNAL newsroom. There, the news piece emerges in four states: drafting, main writing session, cutting session, and speaking in the booth.

Of course it could be argued that the journalist's office, the cutting room, and the speaking booth are different production sites within one media production plant. They differ for example in terms of technical tools (hard- and software facilities for editing text, video, and spoken language), social environments (cutters as collaborators) and dominant activity (spoken vs. written text reproduction). Taken to the limit, the discussion shows that the context of writing keeps on changing: colleagues may call, send messages, or show up and add information that modifies the task; new source texts appear on the screen; and, most of all, the text produced so far, with its power to trigger thoughts when re-read by the writer, is altered through every single insertion and deletion (e.g. Chin 1994b; Hanauer & Englander 2013; Prior this volume).

This fine-grained understanding of constantly changing contexts points towards a shift of focus from the stabilized version to the dynamics of writing processes. Comparing various versions of texts is sufficient to gain empirical evidence of material text changes. However, in itself, it provides hardly any data on the context of material activity. In order to develop such knowledge, additional methodological approaches are required. They focus, for example, on whether the writers

were conscious of their actions (3.2); whether the practices are typical of certain text production institutions (3.3); or how the practices and related norms are negotiated in organizations (3.4).

3.2   The mental focus: Identifying writing strategies with progression analysis

From a cognitive perspective, AL-informed writing research emphasizes individuals' language-related decisions in writing processes. What exactly do individual authors do when they produce their texts? What are they trying to do, and why do they do it the way they do? Such mental reflections of material changes are captured with Progression Analyses. By <u>Progression Analysis</u>, we understand the multimethod approach of collecting and analyzing data in natural contexts in order to reconstruct text production processes as a cognitively reflected activity in context.

Progression Analysis combines ethnographic observation, interviews, computer logging, and cue-based retrospective verbalizations to gather linguistic and contextual data. The approach was developed to investigate newswriting (e.g. Perrin 2003; Sleurs, Jacobs & Van Waes 2003; Van Hout & Jacobs 2008) and later transferred to other application fields of writing research, such as children's writing processes (e.g. Gnach, Wiesner, Bertschi-Kaufmann & Perrin 2007) and translation (e.g. Ehrensberger-Dow & Perrin 2009). With Progression Analysis, data are obtained and related on three levels.

– Before writing begins, Progression Analysis determines through interviews and observations what the writing situation is (e.g. Quandt 2008). Important factors include the writing task, the writers' professional socialization and experience, and economic, institutional, and technological influences on the workplaces and workflows. In the IDÉE SUISSE project, data on the self-perception of the journalists investigated were obtained in semi-standardized interviews about their psychobiography, primarily in terms of their writing and professional experience, and their work situation. In addition, participatory and video observations were made about the various kinds of collaboration at the workplace.
– During writing, Progression Analysis records every keystroke and writing movement in the emerging text with keylogging (e.g. Flinn 1987; Lindgren & Sullivan 2006; Spelman Miller 2006a) and screenshot recording programs (e.g. Degenhardt 2006; Silva 2012) that run in the background behind the text editors that the writers usually use, for instance behind the user interfaces of news editing systems. The recording can follow the writing process over

several workstations and does not influence the performance of the editing system. From a technical point of view, it does not influence the writers' performance either, since it operates automatically and without changing the user interfaces of the editing software. Nevertheless, knowing about the recording alters writers' behavior, with decreasing effect over time. This is why, in projects such as IDÉE SUISSE, the first four weeks of data are excluded from analyses.

– After the writing is over, Progression Analysis records what the writers say about their activities. Preferably immediately after completing the writing process, writers view on the screen how their texts came into being. While doing so, they continuously comment on what they did when writing and why they did it. An audio recording is made of these cue-based retrospective verbal protocols (RVP). This level of Progression Analysis opens a window onto the mind of the writer. The question is what can be recognized through this window: certainly not the sum of all (and only) the considerations that the author actually made, but rather the considerations that an author could have made in principle (e.g. Camps 2003; Ericsson & Simon 1993; Hansen 2006; Levy, Marek & Lea 1996; Smagorinsky 2001). The RVP is transcribed and then encoded as the author's verbalization of aspects of his or her language awareness, writing strategies, and conscious writing practices. As doing an RVP strongly influences writers' awareness, this level of Progression Analysis is normally limited to one RVP per writer, at the end of the investigation.

In sum, Progression Analysis allows researchers to consider all the revisions to the text as well as all of the electronic resources accessed during the production process; to trace the development of the emerging text; and, finally, to reconstruct collaboration at workplaces from different perspectives. The main focus of Progression Analysis, however, is the individual's cognitive and manifest processes of writing. Social structures such as organizational routines and editorial policies are reconstructed through the perspectives of the individual agents involved, the writers under investigation. If entire organizations are to be investigated with respect to how they produce their texts as a social activity, then Progression Analysis has to be extended by another two methods: variation analysis (3.3) and metadiscourse analysis (3.4).

3.3   The social focus: Revealing audience design with variation analysis

From a social perspective, AL-informed writing research focuses on how social groups such as editorial teams customize their linguistic products for their target audiences. Which linguistic means, for example which gradient of normativity and

formality, does an organization choose for which addressees? Such social language use is captured with variation analyses. By <u>variation analysis</u>, we understand the method of collecting and analyzing text data to reconstruct the special features of the language of a certain discourse community. The basis for comparing versions is discourse analysis.

Variation analyses investigate the type and frequency of typical features of certain language users' productions in certain communication situations such as writing for a specific audience. What variation analysis discerns is the differences between the language used and the related practices in one situation type from that of the same users in another (e.g. Koller 2004) or from the language and practices of other users in similar situations (e.g. Fang 1991; Werlen 2000). In the IDÉE SUISSE project, variation analyses can reveal whether language properties of the newscast TAGESSCHAU and the newsmagazine 10 VOR 10, competing in the same German television program of the Swiss public broadcaster, differ according to their program profiles.

Such broadly-based variation analysis is able to show the special features of the language used by specific groups of writers. However, what the method gains in width, it loses in depth. Why a community prefers to formulate its texts in a certain way and not another cannot be captured by variation analysis, which, similar to version analysis, neglects access to mental aspects of writing. It would be possible to regain some of that depth using a procedure that examines not only the text products, but also the institutionalized discourses connected with them – the comments of the community about its joint efforts (3.4).

3.4   The socio-cognitive focus: Investigating language policing with metadiscourse analysis

From a socio-cognitive perspective, AL-informed writing research focuses on text producers' collaboration and metadiscourse (e.g. Mey 2005), such as correspondence between authors, quality control discourse at editorial conferences, and negotiations between journalists, photographers, and text designers. What do the various stakeholders think about their communicational offers? How do they evaluate their activity in relation to policies – and how do they reconstruct and alter those policies? Such socio-cognitive aspects of language use are captured by metadiscourse analysis, by which we understand the method of collecting and analyzing data in order to reconstruct the socially- and individually-anchored (language) awareness in a discourse community. The basis for analyzing the metadiscourse of text production is conversation and discourse analysis.

Metadiscourse analyses investigate spoken and written communication about language and language use. This includes metaphors used when talking about writing (e.g. Gravengaard 2012; Levin & Wagner 2006), explicit planning or criticism

of communication measures (e.g. Peterson 2001), the clarification of misunderstandings and conversational repair (e.g. Häusermann 2007), and follow-up communication by audiences (e.g. Klemm 2000). In all these cases, the participants' utterances show how their own or others' communicational efforts and offers have been perceived, received, understood, and evaluated. The analysis demonstrates how rules of language use are explicitly negotiated and applied in a community.

Genetic Criticism for example draws on metadiscourse analyses where writers' manuscript side notes about their own writing are taken into account. Some examples: In the case of Proust's "Cahiers", side notes refer to the writing activity itself (e.g. Herschberg Pierrot 1994). Flaubert's correspondence provides variegated information about the emergence of his texts (e.g. Grésillon et al. 1991). The drafts of Zola's novels are full of metalinguistic comments about what has been written so far and what still is to be written: "Tout cela me paraît bon" – "This seems all good to me"; "Quand il s'aperçoit que sa femme le trompe, il faut une scène d'une violence inouïe" – "When he realizes that his wife is unfaithful to him, it takes a scene of outrageous violence"; "Pourtant, cela est à régler, lorsque j'aurai tous les éléments" – "Nevertheless, this is to be put straight, when I have all the elements" (e.g. Grésillon 2002).

Due to a computer crash, the journalist writing about the demonstrations in Lebanon lacks the time to discuss his news piece with the cutter. In other case stories from the IDÉE SUISSE project, cutters challenge the journalists' ethics and esthetics or appear as representatives of a critical audience. On a macro level of the project, interviews and document analyses reveal policy makers' and media managers' contradictory evaluation of and expectations towards the broadcasters' – and the journalists' – ability to fulfill the public mandate of promoting public understanding. Whereas media policy makers expect the Swiss national broadcasting company to foster public discourse through stimulating contributions, media managers tend to consider this public mandate to be unrealistic (Perrin 2011, 8).

Thus, the focus of metadiscourse analysis scales up from negotiations about emerging texts at writers' workplaces, to organizational quality control discourse and related discussions in audiences and society at large. Integrating metadiscourse analyses extends the reach of writing research from a single author's micro activity to societal macro structures. However, for empirical evidence of writers' actual behavior, metadiscourse analysis must be combined with progression analyses (3.2) or, in more coarse-grained studies, at least with version analyses (3.3).

4.   Conclusion

In sum, by applying and combining methods of the four types, researchers investigate real-life writing from product and process perspectives, as cognitive and

social activity, and on micro and macro levels. In contrast, analyzing only text products, as often practiced in empirical approaches to written language, risks falling short of explaining writing in its variegated dynamics and purposes, as a playful, epistemic, and communicative activity in complex contexts. However, applying, let alone combining, innovative methods in multi-perspective real-life writing research causes methodological problems which can be carefully addressed – albeit not completely solved yet.

From the four perspectives combined, research is as good as its methods are. What basically applies for all academic work is, within AL-informed research, particularly true for the field of writing and text production. For decades, most analyses of written communication drew on text products only, neglecting procedural insights due to methodological constraints. With digital writing environments, things started changing. However, the non-digital aspects and the backstage activities of real-life writing are still hard to capture in their dynamics and complexity, requiring archeological approaches such as Genetic Criticism and leaving researchers with assumptions – albeit empirically grounded – about mental and societal structures and processes.

References

Abbott, Robert D., Dagmar Amtmann, and Jeff Munson. 2006. "Statistical analysis for field experiments and longitudinal data in writing research." In *Handbook of Writing Research*, ed. by Charles A. MacArthur, Steve Graham, and Jill Fitzgerald, 374–385. New York: Guilford.

Alatis, James E., Heidi E. Hamilton, and Ai-Hui Tan (Eds.). 2002. *Linguistics, Language and the Professions. Education, Journalism, Law, Medicine, and Technology*. Washington D.C.: Georgetown University Press. DOI: 10.1017/s0047404504241057

Antos, Gerd. 1989. "Textproduktion. Ein einführender Überblick." In *Textproduktion. Ein interdisziplinärer Forschungsüberblick*, ed. by Gerd Antos, and Hans Peter Krings, 5–57. Tübingen: Niemeyer. DOI: 10.1515/9783110962109

Bell, Allan. 1991. *The Language of News Media*. Oxford: Blackwell.
DOI: 10.1017/s004740450001695x

Bisaillon, Jocelyne. 2007. "Professional editing strategies used by six editors." *Written Communication* 24(4): 295–322. DOI: 10.1177/0741088307305977

Bracewell, Robert J. 2003. "Tasks, ensembles, and activity. Linkages between text production and situation of use in the workplace." *Written Communication* 20(4): 511–559.
DOI: 10.1177/0741088303260691

Camps, Joaquim. 2003. "Concurrent and retrospective verbal reports as tools to better understand the role of attention in second language tasks." *International Journal of Applied Linguistics* 13(2): 201–221. DOI: 10.1111/1473–4192.00044

Candlin, Sally. 2003. "Issues arising when the professional workplace is the site of applied linguistic research." *Applied Linguistics* 24(3): 386–394. DOI: 10.1093/applin/24.3.386

Chin, Elaine. 1994a. "Ethnographic interviews and writing research. A critical examination of the methodology." In *Speaking About Writing. Reflections on Research Methodology*, ed. by Peter Smagorinsky, 247–272. Thousand Oaks et al.: Sage. DOI: 10.1017/s027226310001514x

Chin, Elaine. 1994b. "Redefining "context" in research on writing." *Written Communication* 11(4): 445–482. DOI: 10.1177/0741088394011004002

Cicourel, Aaron Victor. 2003. "On contextualizing applied linguistic research in the workplace." *Applied Linguistics* 24(3): 360–373. DOI: 10.1093/applin/24.3.360

Cook, Guy. 2003. *Applied linguistics*. Oxford: Oxford University Press. DOI: 10.1017/s0272263197223053

Cottle, Simon. 1998. "Participant observation: Researching news production." In *Mass Communication Research Methods*, ed. by Andres Hansen, Simon Cottle, Chris Newbold, and Ralph Negrine, 25–65. New York: New York University Press.

de Biasi, Pierre-Marc. 2011. *Génétique des textes*. Paris: CNRS Editions.

Degenhardt, Marion. 2006. CAMTASIA and CATMOVIE. "Two digital tools for observing, documenting and analysing writing processes of university students." In *Writing and Digital media*, ed. by Luuk Van Waes, Mariëlle Leijten, and Christine M. Neuwirth, 180–186. Amsterdam: Elsevier. DOI: 10.1163/9781849508209_002

Ehrensberger-Dow, Maureen, and Daniel Perrin. 2009. "Capturing translation processes to access metalinguistic awareness." *Across Languages and Cultures* 10(2): 275–288. DOI: 10.1556/acr.10.2009.2.6

Ekström, Mats. 2001. "Politicians interviewed on television news." *Discourse & Society* 12(5): 563–584. DOI: 10.1177/0957926501012005001

Ericsson, Ken A., and Herbert A. Simon. 1984. *Protocol Analysis. Verbal Reports as Data*. Cambridge: MIT Press. DOI: 10.1097/00005053-198511000-00011

Ericsson, Ken A., and Herbert A. Simon. 1993. *Protocol Analysis. Verbal Reports as Data*. (Rev. ed.). Cambridge, MA: MIT Press. DOI: 10.1097/00005053-198511000-00011

Fang, Irving. 1991. *Writing Style Differences in Newspaper, Radio and Television News*. Minnesota: Center for Interdisciplinary Studies of Writing, University of Minnesota.

Ferrer, Daniel. 2011. *Logiques du brouillon. Modèles pour une critique génétique*. Paris: Seuil.

Flinn, Jane Z. 1987. "Case studies of revision aided by keystroke recording and replaying software." *Computers and Composition* 5(1): 31–44. DOI: 10.1016/s8755-4615(87)80013–0

Gnach, Aleksandra, Esther Wiesner, Andrea Bertschi-Kaufmann, and Daniel Perrin. 2007. "Children's writing processes when using computers. Insights based on combining analyses of product and process." *Research in Comparative and International Education* 2(1): 13–28. DOI: 10.2304/rcie.2007.2.1.13

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor. DOI: 10.1525/aa.1962.64.1.02a00260

Gravengaard, Gitte. 2012. "The metaphors journalists live by. Journalists' conceptualisation of newswork." *Journalism* 2012(13): 1064–1082. DOI: 10.1177/1464884911433251

Greene, Stuart, and Lorraine Higgins. 1994. ""Once upon a time": The use of retrospective accounts in building theory in composition." In *Speaking About Writing. Reflections on Research Methodology*, ed. by Peter Smagorinsky, 115–140. Thousand Oaks et al.: Sage.

Grésillon, Almuth. 1987. "Wege und Irrwege. Zu Heines Gedicht "Lebensfahrt"." In *Literarische Schreibprozesse [Zeitschrift für Literaturwissenschaft und Linguistik]*, ed. by Brigitte Schlieben-Lange, and Almuth Grésillon, 84–103. Göttingen: Vandenhoeck & Ruprecht.

Grésillon, Almuth. 1994. *Eléments de critique génétique. Lire les manuscrits modernes*. Paris: PUF. DOI: 10.1515/arbi.2000.18.1.1

Grésillon, Almuth. 1997. "Slow: work in progress." *Word & Image* 13(2): 106–123. DOI: 10.1080/02666286.1997.10434276

Grésillon, Almuth. 2002. "Langage de l'ébauche. Parole intérieure extériorisée." *Langages* 147: 19–38. DOI: 10.3406/lgge.2002.2410

Grésillon, Almuth (Ed.). 2008a. *La Mise en œuvre. Itinéraires génétiques*. Paris: CNRS Editions.

Grésillon, Almuth. 2008b. "Lire pour écrire." In *La Mise en œuvre. Itinéraires génétiques*, ed. by Almuth Grésillon, 67–81. Paris: CNRS Editions.

Grésillon, Almuth. 2008c. "Ponge: «L'Ardoise» dans tous ses états." In *La Mise en œuvre. Itinéraires génétiques*, ed. by Almuth Grésillon, 207–241. Paris: CNRS Editions.

Grésillon, Almuth. 2014. "Erfahrungen mit Textgenese, Critique génétique und Interpretation." In *Text – Material – Medium. Zur Relevanz editorischer Dokumentationen für die literaturwissenschaftliche Interpretation*, ed. by Wolfgang Lukas, Rüdiger Nutt-Kofoth, and Madleen Podewski, 66–79. Berlin et al.: De Gruyter. DOI: 10.1515/9783110364408

Grésillon, Almuth, and Jean-Louis Lebrave. 2008. "Linguistique et génétique des textes. Un décalogue." *Le Français moderne* [Numéro spécial publié à l'occasion de son 75e anniversaire]: 37–49.

Grésillon, Almuth, Jean-Louis Lebrave, and Catherine Fuchs. 1991. "«Ruminer Hérodias». Du cognitif-visuel au verbal-textuel." In *L'écriture et ses doubles. Genèse et variation textuelle*, ed. by Daniel Ferrer, and Jean-Louis Lebrave, 27–109. Paris: CNRS Editions.

Grésillon, Almuth, and Marie-Madeleine Mervant-Roux. 2010. "Marguerite Duras/Claude Régy. «L'amante anglaise». Genèse d'une écriture, gésine d'un théâtre." In *Genèses théâtrales*, ed. by Almuth Grésillon, Marie-Madeleine Mervant-Roux, and Dominique Budor, 211–232. Paris: CNRS Editions.

Hanauer, David Ian, and Karen Englander. 2013. *Scientific Writing in a Second Language*. Anderson: Parlor Press.

Hansen, Gyde. 2006. "Retrospection methods in translator training and translation research." *Journal of Specialised Translation* 5: 2–40.

Häusermann, Jürg. 2007. "Zugespieltes Material. Der O-Ton und seine Interpretation." In *Original/Ton. Zur Mediengeschichte des O-Tons*, ed. by Harun Maye, Cornelius Reiber, and Nikolaus Wegmann, 25–50. Konstanz: UVK.

Hay, Louis. 2002. *La littérature des écrivains. Questions de critique génétique*. Paris: Corti.

Herschberg Pierrot, Anne. 1994. "Les notes de Proust." *Genesis* 6: 61–78. DOI: 10.4000/genesis.125

Keseling, Gisbert. 1992. "Pause and intonation contours in written and oral discourse." In *Cooperating With Written Texts. The Pragmatics and Comprehension of Written Texts*, ed. by Dieter Stein, 31–66. Berlin: De Gruyter. DOI: 10.1515/9783110881196.31

Klemm, Michael. 2000. *Zuschauerkommunikation. Formen und Funktionen der alltäglichen kommunikativen Fernsehaneignung*. Frankfurt am Main et al.: Lang.

Kollberg, Py, and Kerstin Severinson-Eklundh. 2002. "Studying writers' revising patterns with S-notation analysis." In *Contemporary Tools and Techniques for Studying Writing*, ed. by C. Michael Levy, and Thierry Olive, 89–104. Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-010-0468-8_5

Koller, Veronika. 2004. "Businesswomen and war metaphors: "Possessive, jealous and pugnacious"?" *Journal of Sociolinguistics* 4(1): 3–22. DOI: 10.1111/j.1467–9841.2004.00249.x

Krings, Hans Peter. 1992. "Schwarze Spuren auf weißem Grund. Fragen, Methoden und Ergebnisse der empirischen Schreibprozeßforschung im Überblick." In *Textproduktion. Neue*

Wege der Forschung, ed. by Hans Peter Krings, and Gerd Antos, 45–110. Trier: Wissenschaftlicher Verlag.

Lams, Lutgard. 2011. "Newspapers' narratives based on wire stories. Facsimiles of input?" *Journal of Pragmatics* 43(7): 1853–1864. DOI: 10.1016/j.pragma.2010.09.021

Lebrave, Jean-Louis. 1987. "Lecture et analyse des brouillons." *Langages* 69: 11–23. DOI: 10.3406/lgge.1983.1139

Lebrave, Jean-Louis. 1992. "La critique génétique. Une discipline nouvelle ou un avatar moderne de la philologie?" *Genesis* 1: 33–72.

Levin, Tamar, and Till Wagner. 2006. "In their own words: Understanding student conceptions of writing through their spontaneous metaphors in the science classroom." *Instructional Science* 34(3): 227–278. DOI: 10.1007/s11251-005-6929-x

Levy, C. Michael, Pamela J. Marek, and Joseph Lea. 1996. "Concurrent and retrospective protocols in writing research." In *Theories, Models and Methodology in Writing Research*, ed. by Gert Rijlaarsdam, Huub Van den Bergh, and Michael Couzijn, 542–556. Amsterdam: Amsterdam University Press. DOI: 10.5117/9789053561973

Levy, C. Michael, and Sarah Ransdell. 1996. "Writing signatures." In *The Science of Writing. Theories, Methods, Individual Differences and Applications*, ed. by C. Michael Levy, and Sarah Ransdell, 127–148. Mahwah: Erlbaum. DOI: 10.1037/e537272012-375

Lillis, Theresa. 2008. "Ethnography as method, methodology, and "deep theorizing". Closing the gap between text and context in academic writing research." *Written Communication* 25(3): 353–388. DOI: 10.1177/0741088308319229

Lindgren, Eva, and Kirk Sullivan. 2006. "Analysing online revision." In *Computer Keystroke Logging and Writing. Methods and Applications*, ed. by Kirk Sullivan, and Eva Lindgren, 157–188. Amsterdam: Elsevier.

Litosselity, Lia (Ed.). 2009. *Research Methods in Linguistics*. London: Continuum.

Luginbühl, Martin, Thomas Baumberger, Kathrine Schwab, and Harald Burger. 2002. *Medientexte zwischen Autor und Publikum. Intertextualität in Presse, Radio und Fernsehen*. Zürich: Seismo.

MacMillan, Stuart. 2012. "The promise of ecological inquiry in writing research." *TechnicalCommunication Quarterly* 21(4): 346–361. DOI: 10.1080/10572252.2012.674873

Mahrer, Rudolf. 2006. "Quand Ramuz remonte le courant. Réécriture des romans en vue des œuvres complètes." *Genesis* 27: 59–72.

McCarthy, Michael. 2001. *Issues in Applied Linguistics*. Cambridge: Cambridge University Press. DOI: 10.1017/s1360674307002535

Mey, Jacob L. 2005. "Discourse and metadiscourse." *Journal of Pragmatics* 37(9): 1323–1324. DOI: 10.1016/j.pragma.2005.05.006

Perrin, Daniel. 2003. "Progression analysis (PA). Investigating writing strategies at the workplace." *Journal of Pragmatics* 35(6): 907–921. DOI: 10.1016/s0378-2166(02)00125-x

Perrin, Daniel. 2011. *Medienlinguistik* (2 ed.). Konstanz: UVK. DOI: 10.1007/s11616-006-0264-3

Perrin, Daniel. 2012. "La voie tranquille. Routine und Emergenz in Formulierungsprozessen als service public." In *Schreib- und Textroutinen. Theorie, Erwerb und didaktisch-mediale Modellierung*, ed. by Helmuth Feilke, and Katrin Lehnen, 215–239. Frankfurt am Main et al.: Lang.

Perrin, Daniel. 2013. *The Linguistics of Newswriting*. Amsterdam, New York et al.: John Benjamins.

Perrin, Daniel, and Marc Wildi. 2010. "Statistical modeling of writing processes." In *Traditions of Writing Research*, ed. by Charles Bazerman, 378–393. New York: Routledge.

Peterson, Mark Allen. 2001. "Getting to the story. Unwriteable discourse and interpretive practice in American journalism." *Anthropological Quarterly* 74(4): 201–211. DOI: 10.1353/anq.2001.0038

Quandt, Thorsten. 2008. "News tuning and content management. An observation study of old and new routines in German online newsrooms." In *Making Online News. The Ethnography of New Media Production*, ed. by Chris Paterson, and David Domingo, 77–97. New York: Peter Lang.

Robinson, Sue. 2009. "A chronicle of chaos. Tracking the news story of hurricane Katrina from The Times-Picayune to its website." *Journalism* 10(4): 431–450. DOI: 10.1177/1464884909104949

Sanders, Ted, and Carel Van Wijk. 1996. "Text analysis as a research tool. How hierarchical text structure contributes to the understanding of conceptual processes in writing." In *The Science of Writing. Theories, Methods, Individual Differences and Applications*, ed. by C. Michael Levy, and Sarah Ransdell, 251–270. Mahwah: Erlbaum.

Schindler, Kirsten, and Joanna Wolfe. 2014. "Beyond single authors. Organizational text production as collaborative writing." In *Handbook of Writing and Text Production*, Vol. 10, ed. by Daniel Perrin, and Eva-Maria Jakobs, 159–173. New York et al.: De Gruyter. DOI: 10.1515/9783110220674.159

Schneider, Barbara. 2002. "Theorizing structure and agency in workplace writing. An ethnomethodological approach." *Journal of Business and Technical Communication* 16(2): 170–195. DOI: 10.1177/1050651902016002002

Schultz, Katherine. 2006. "Qualitative research on writing." In *Handbook of Writing Research*, ed. by Charles A. MacArthur, Steve Graham, and Jill Fitzgerald, 357–373. New York: Guilford. DOI: 10.1111/j.1467-873x.2008.00423.x

Severinson-Eklundh, Kerstin, and Py Kollberg. 1996. "Computer tools for tracing the writing process. From keystroke records to S-notation." In *Current Research in Writing. Theories, Models and Methodology*, ed. by Gert Rijlaarsdam, Huub Van den Bergh, and Michael Couzijn, 526–541. Amsterdam: Amsterdam University Press. DOI: 10.5117/9789053561973

Severinson-Eklundh, Kerstin, and Carin Sjöholm. 1991. "Writing with a computer: a longitudinal study of writers of technical documents." *Man-Machine Studies* 35: 723–749. DOI: 10.1016/s0020-7373(05)80185–3

Silva, Mary Lourdes. 2012. "Camtasia in the classroom. Student attitudes and preferences for video commentary or Microsoft Word comments during the revision process." *Computers and Composition* 29(1): 1–22. DOI: 10.1016/j.compcom.2011.12.001

Sleurs, Kim, Jacobs, Geert, and Luuk Van Waes. 2003. "Constructing press releases, constructing quotations: A case study." *Journal of Sociolinguistics* 7(2): 135–275. DOI: 10.1111/1467–9481.00219

Smagorinsky, Peter. 2001. "Rethinking protocol analysis from a cultural perspective." *Annual Review of Applied Linguistics* 21: 233–245. DOI: 10.1017/s0267190501000149

Smagorinsky, Peter (Ed.). 1994. *Speaking About Writing. Reflections on Research Methodology*. Thousand Oaks et al.: Sage. DOI: 10.1017/s027226310001514x

Spelman Miller, Kristvan. 2006a. "Keystroke logging. An introduction." In *Computer Keystroke Logging and Writing. Methods and Applications*, ed. by Kirk Sullivan, and Eva Lindgren, 1–9. Amsterdam: Elsevier.

Spelman Miller, Kristvan. 2006b. "The pausological study of written language production." In *Computer Keystroke Logging and Writing. Methods and Applications*, ed. by Kirk Sullivan, and Eva Lindgren, 11–30. Amsterdam: Elsevier.

Strömqvist, Sven, Kenneth Holmqvist, Victoria Johansson, Henrik Karlsson, and Åsa Wengelin. 2006. "What key-logging can reveal about writing." In *Computer Keystroke Logging and Writing. Methods and Applications*, ed. by Kirk Sullivan, and Eva Lindgren, 45–72. Amsterdam: Elsevier.

Stubbs, Michael. 1997. "Critical comments on Critical Discourse Analysis (CDA)." In *Evolving Models of Language*, ed. by Ann Ryan, and Alison Wray, 100–116. Clevedon: Multilingual Matters.

Sullivan, Patricia, and James E. Porter. 1993. "On theory, practice, and method. Toward a heuristic research methodology for professional writing." In *Writing in the Workplace. New Research Perspectives*, ed. by Rachel Spilka, 220–237. Carbondale et al.: Southern Illinois University Press.

Tashakkori, Abbas, and Charles Teddlie. 2003. "Major issues and controversies in the use of mixed methods in social and behavioral research." In *Handbook of Mixed Methods in Social & Behavioral Research*, ed. by Abbas Tashakkori, and Charles Teddlie, 3–50. Thousand Oaks: Sage. DOI: 10.1002/9781118133880.hop202015

Van den Bergh, Huub, and Gert Rijlaarsdam. 1996. "The dynamics of composing. Modeling writing process data." In *The Science of Writing. Theories, Methods, Individual Differences and Applications*, ed. by C. Michael Levy, and Sarah Ransdell, 207–232. Mahwah: Erlbaum.

Van der Geest, Thea. 1996. "Studying "real life" writing processes. A proposal and an example." In *The Science of Writing. Theories, Methods, Individual Differences and Applications*, ed. by C. Michael Levy, and Sarah Ransdell, 309–322. Mahwah: Erlbaum.

Van Dijk, Teun A. 1988. *News Analysis. Case Studies of International and National News in the Press*. Hillsdale/London: Erlbaum.

Van Dijk, Teun A. 2001. "Multidisciplinary CDA. A plea for diversity." In *Methods of Critical Discourse Analysis*, ed. by Ruth Wodak, and Michael Meyer, 95–120. Los Angeles: Sage. DOI: 10.4135/9780857028020.n5

Van Hout, Tom, and Geert Jacobs. 2008. "News production theory and practice. Fieldwork notes on power, interaction and agency." *Pragmatics* 18(1): 59–86.

Van Waes, Luuk, and Anne Mangen. 2012. "Introduction. Tools for studying and supporting writing. Technological advances in writing research." In *Learning to Write Effectively. Current Trends in European Research*, ed. by Mark Torrance, Denis Alamargot, Montserrat Castello, Franck Ganier, Otto Kruse, Anne Mangen, Liliana Tolchinsky, and Luuk Van Waes, 345–351. Bingley: Emerald.

Van Waes, Luuk, and Liesbet Van Herreweghe. 1995. "Computerprotokolle in der Schreibprozessforschung. Der Gebrauch von Keytrap als Beobachtungsinstrument." In *Wissenschaftliche Textproduktion. Mit und ohne Computer*, ed. by Eva-Maria Jakobs, Dagmar Knorr, and Sylvie Molitor-Lübbert, 35–51. Frankfurt am Main et al.: Lang.

Wei, Li. 2007. "A user-friendly linguistics." *International Journal of Applied Linguistics* 17(1): 117–119. DOI: 10.1111/j.1473–4192.2007.00136.x

Werlen, Iwar. 2000. ""Zum Schluss das Wetter". Struktur und Variation von Wetterberichten im Rundfunk." In *Sprachsplitter und Sprachspiele*, ed. by Jürg Niederhauser, and Stanislaw Szlek, 155–170. Bern: Lang.

Whithaus, Carl, Scott B. Harrison, and Jeb Midyette. 2008. "Keyboarding compared with handwriting on a high-stakes writing assessment. Student choice of composing medium, raters' perceptions, and text quality." *Assessing Writing* 13(1): 4–25. DOI: 10.1016/j.asw.2008.03.001

Widdowson, Henry G. 2000. "On the limitations of linguistics applied." *Applied Linguistics* 21(1): 3–35. DOI: 10.1093/applin/21.1.3

Wolfe, Joanna. 2005. "Gesture and collaborative planning. A case study of a student writing group." *Written Communication* 22(3): 298–332. DOI: 10.1177/0741088305278108

Woolley, Claire M. 2009. "Meeting the mixed methods challenge of integration in a sociological study of structure and agency." *Journal of Mixed Methods Research* 3(1): 7–25. DOI: 10.1177/1558689808325774

Zhong, Bu, and John E. Newhagen. 2009. "How journalists think while they write. A transcultural model of news decision making." *Journal of Communication* 59(3): 587–608. DOI: 10.1111/j.1460–2466.2009.01439.x

Linguistic forms and choices at the interfaces

The *instrumental* use of verbless sentences in writing and rewriting

A longitudinal and genre-contrasted point of view

Georgeta Cislaru & Florence Lefeuvre
CLESTHIA, University Sorbonne nouvelle Paris 3

We sought to determine the role of verbless sentences across the writing process. We compared two corpora: a longitudinal corpus of social reports and a transversal corpus of diary entries and letters. An examination of successive draft reports revealed that verbless sentences were present in incipient versions, but subsequently grew into full sentences or even paragraphs. They served as reminders - condensed syntactic and semantic content intended to be expanded into full text. Similarly, verbless sentences contained in the diary entries were swiftly worked up into verbal sentences in letters relating the same events as the diary. We hypothesize the existence of a pragmatic constraint differentiating between self-addressed texts (incipient versions, personal diaries) and other-addressed texts (final reports, letters).

Keywords: verbless sentences; rewriting; social reports; diary; letters

1. Introduction

We conducted a study to determine the role and behavior of verbless sentences (VS) in hybrid corpora that allowed us to undertake a thorough observation of the rewriting process and the text as a finished product. VS are generally studied in languages where verbless copular clauses are the norm, such as Arabic, Hebrew, Hungarian and Russian (see Eid 1991; Hengeveld 1992;[1] Nordlinger & Sadler 2006). In languages like French or English, where standard copular clauses are verbal, the existence of VS such as:

---

1. According to Hengeveld (1992, 26–27), nonverbal predication can be expressed by means of either a copular construction or a nominal sentence. We did not take copular constructions into account here. Our nonverbal clauses seemed to be closer to nominal sentences, as per Hengeveld's definition.

(1)   a.   *Beautiful dress*!
      b.   *What for, this book*?
      c.   *Hey, guys, on the road*! (Lefeuvre 1999)

is subject to major problems of categorization. Linguists tend to treat them as ellipses, or as mangled versions of verbal structures (see Merchant 2004). For the purposes of our research, however, we regarded these syntactic constructions as fully-fledged VS. A VS (see Lefeuvre 1999) is a direct speech act (sentence type: declarative, interrogative, imperative) that contains a nonverbal predicate. The nonverbal predicate can take negation and other modalities (e.g. *Not a sound, Not fair*). A VS may characterize (a), localize (b, c), assert the existence of something (d) or express an event (e):

      d.   *On the left, a table*
      e.   *Yesterday, walk on the riverside*

VS are compact constructions with a high potential from an informative and expressive viewpoint. As stated by Kotjevskaja-Tamm (1993, 266), nominalizations "may be used for at least two opposing purposes – to compress texts and to elaborate them" (see also Jespersen 1924). When considering heterogeneous corpora such as social reports on children at risk, and diary entries and letters, we may find numerous VS in both the draft reports and the diary entries that are specifically developed in the final reports and in letters whose content is derived from the diary entries. What, then, is the role of VS in text construction, through process to product? And how can we explain the parallel observed between social reports, diary entries and letters? We argue that VS play a complex, instrumental role in writing and rewriting, functioning as semantic and cognitive textual cues (see also Schleppegrell 2007).

2.   Verbless sentences and discourse genre constraints

As noted by Benveniste (1966) and discussed further by Lefeuvre (1999), VS are relatively frequent in written communication. Benveniste (1966) showed that VS are often used in direct speech expressing universal truths, although they can be found in a variety of discourse genres. Thus, verbless sentences are used in literature (Behr 2011; Delorme 2004; Lefeuvre 2015) and academic discourse (Nichols 1988), as well as media discourse (Lefeuvre 2004, 2007; Grinshpun 2011) for a variety of rhetorical, stylistic and textual purposes.

It was Jespersen (1924, 137–139), followed by Kotjevskaja-Tamm (1993, 266), who first signaled the text compressing uses of nominalizations, which can be used as VS to express an event:

(2) a. *Mais tout de même, les chômeurs ne sont-ils pas stigmatisés? <u>Grand éclat de rire général</u>: "Mais non: c'est plutôt nous, les travailleurs, qui passons pour des cons!" Sûrement exagéré. Mais pas très rassurant.* (*Le Nouvel Observateur*, 6–12 February 2014 – media discourse)

But, all the same, the unemployed aren't they stigmatized? Laughter all round: "No! it's us, the employed, who look like damned fools!" Surely an exaggeration. But not very comforting.

b. *"[…] D'ailleurs, les meilleurs s'arrangent pour ne pas trop savoir ce qui s'est passé. La danse finie, Bach de nouveau."*
<u>*Sonnerie du téléphone*</u>. *Il décrocha l'appareil* […].
(Cohen, *Belle du Seigneur* – literary discourse)

"[…] Moreover, the best ones manage to turn a blind eye. Once the dance over, Bach again."
Ring of the phone. He picked up the receiver […].

They are even more frequent in some types of discourse, such as stage directions and all kinds of notes:

(3) *Texte de L. non encore parvenu* (note in the margin of a script)
L.'s script not come yet

as well as in short forms of discourse, such as text messages:

(4) *Gâteau pas encore pris*
Cake not picked up yet

Private diaries constitute a discourse genre that favors and easily accommodates verbless sentences, especially sentences describing places:

(5) *La rivière est déjà un estuaire, ça sent l'algue et la vase. <u>Plaisant paysage d'automne</u>.* (Diary, 21 September 1939)

The river is already an estuary; it smells of mud and seaweed. Nice autumn landscape.

events (existential VS):

(6) a. *Je me calme, je descends; les rues sont graves. Sur la place Montparnasse, <u>une bagarre</u>.* (Diary, 3 September 1939)

I calm down, I go down; the streets are somber. In Montparnasse Square, a fight.

b. *Il y a du clair de lune plein la chambre. <u>Soudain un grand cri</u> – je vais à la fenêtre, une femme a crié.* (Diary, 1 September 1939)

The room is full of moonlight. Suddenly a loud cry – I go to the window, a woman had cried out.

or the writer's emotion:

(7)   *Un petit spasme en m'asseyant ici: la fenêtre de Bost, Bost traversant la place* (*avant notre amour), je le revois si bien.* (Diary, 5 September 1941)

A little convulsion when I sit down here: Bost's window, Bost crossing the square (before our love), I can see him again so clearly.

Fictitious diaries often use VS and, more precisely, nominalizations:

(8)   a.   *Réunion mardi chez le curé D'Hébuterne, pour la conférence mensuelle.* (Bernanos, *Journal d'un Curé de Campagne*, cf. Lefeuvre 1999)

Meeting on Tuesday at the vicar D'Hébuterne's place, for the monthly conference.

b.   *Rapporté à Olivier ses affaires. Sitôt de retour de chez Passavant, travail. Exaltation calme et lucide.* (Gide, *Les Faux Monnayeurs*)

Took Olivier's belongings back to him. As soon as back at Passavant's, work. Calm and lucid exaltation.

It is the correlation with the corpus of letters that underlines the specific uses of VS in Simone de Beauvoir's diary.

Social reports do not belong to a discourse genre that allows for extensive VS use. Thus, the occurrence of a VS other than a title or subtitle suggests that the text is still in its draft version. Indeed, addressed to a judge and with an extremely high performative value, social reports have to formulate and develop all the key concepts relating to the situation of the child or its family. From this point of view, in any longitudinal analysis of a corpus of social reports, VS can be assumed to serve as cognitive-semantic writing tools.

3.   Corpora

The present study was supported by a collection of data in the form of:

–   the draft and final versions of 27 social reports on children at risk, each file containing 2–41 versions of the same text; all of them are considered as texts (see Mahrer & Nicollier Saraillon 2014);
–   Simone de Beauvoir's private diary (*Journal de Guerre*, September 1939 to January 1941);
–   Simone de Beauvoir's letters to Jean-Paul Sartre (*Lettres à Sartre*, 1930–1939, 1940–1963).

3.1   Social reports as an example of constrained professional writing

Social reports on children at risk are examples of professional writing that are both institutionally and socially constrained. The authors of our corpus were social

workers and youth workers belonging to an association who worked with more than 20 fostered children aged 5–21 years. The social workers had to write regular and final reports describing the situation and progress of each child, and recalling the family history and the reasons for fostering. Each report ended with a conclusion. All the reports had comparable structures, divided into the same sections and with comparable section headings. Indeed, there was even an institutional template, although this was not overtly followed.

47	ou le recours aux adultes n'est pas acquis d'emblée.	ou le recours aux adultes n'est pas acquis d'emblée.		n'est pas acquis d'emblée.	n'est pas acquis d'emblée.	
			Evolution			
48		II. Evolution		II. Evolution	II. Evolution	1 insertion
49		sur le groupe		sur le groupe :	sur le groupe :	2 insertions
50				Jade est une jeune fille de tout juste 15 ans.	Jade est une jeune fille de tout juste 15 ans.	1 insertion
51		Jade a		Elle a	Elle a	1 insertion 1 substituti
52		intégré le groupe de l'Origami le 18 janvier 2010.		intégré le groupe de l'Origami le 18 janvier 2010.	intégré le groupe de l'Origami le 18 janvier 2010.	1 insertion
53		Nous constatons à la dernière échéance que Jade s'était installée sur le groupe comme une jeune		Nous constatons à la dernière échéance que Jade s'était installée sur le groupe comme une jeune	Nous constatons à la dernière échéance que Jade s'était installée sur le groupe comme une jeune	1 insertion
54		" calme mais pas effacée ".		calme, mais cependant pas effacée.	calme, mais cependant pas effacée.	2 insertions 1 substituti 1 suppressi
55		La prolongation		La prolongation	La prolongation	1 insertion
56		du reccueil		du recueil	du recueil	1 insertion 1 substituti

Figure 1. Report template and sections

Our corpus of social reports contained 23 files of drafts recorded by the authors themselves and four files of drafts recorded by Inputlog, a keystroke-logging program that also records realtime data (cf. Leijten & Van Waes 2006; Leijten et al. this volume). The data recorded by the authors were computer processed and periodically saved (ideally, every 10–15 minutes).

Via all these documents, we were able to track the gestation of the social reports from the first short, schematic version to the full-length, final report. Although fragments of text might be either added or subtracted from one version to another, generally speaking, the text increased in volume between Version 1 and Version N. The existence of these successive drafts enabled us to adopt a longitudinal approach (see, among others, Fuchs et al. 1987; Prior 2004) by means of text alignment, using MkAlign[2] and Allongos[3] (Lardilleux et al. 2013), in order to study the way that VS are handled across the writing process. The results of the alignment, showing the various rewriting operations (marked with different colors), can be seen in Figure 1.

---

2.   Developed by Serge Fleury at Université Sorbonne nouvelle Paris 3 (http://www.tal.univ-paris3.fr/mkAlign/).

3.   This is automatic alignment software developed by Adrien Lardilleux within the Ecritures ANR program (http://www.univ-paris3.fr/allongos-221592.kjsp?RH=1295620557102).

The social reports had complex communicative features:

– collective writing: several authors wrote and revised the text. The body of the report was written by social workers, but the conclusion was written by the section head of the child protection unit. Authorship therefore had to be understood as a collective responsibility, implying collective norms of writing and assessment;

– an official addressee in the shape of the judge, who requested the report in the first place. In our corpus of social reports, there were two kinds of texts: regular progress reports, and final reports that summarized the overall situation. Although the latter were written with the judge in mind, they might also be consulted and used by the families during the court hearings;

– an unofficial addressee in the shape of the family, who were given the right to access social reports in the 2002 act reforming French social services. This means that while social workers must describe the child's situation as accurately as possible, with some judges even providing guidelines (cf. *Guide pratique pour la protection de l'enfance*, Huyette & Desloges 2009), they must also protect the child and its family, and avoid hurting them in the report.

Another layer of complexity concerns the data that has to be set out and used for the purposes of assessment and argumentation. Social workers collect field data while visiting the family/child, and during meetings with the parents. They then have to summarize this information, which evidently results in some VS, at least in the incipient versions of the reports.

3.2   The private diary and the letters: Two sides of the same life experience and informational content

Our other corpus provided two different versions of the same life experience: Simone de Beauvoir's private diary (*Journal de Guerre*, September 1939-January 1941) and her letters to Jean-Paul Sartre (*Lettres à Sartre*, 1930–1939, 1940–1963). Between September 1939 and January 1941, de Beauvoir wrote in her private diary almost daily, while at the same time sending letters to Sartre. The two sets of texts are opposed from the point of view of the communicative situation: in the private diary, the writer wrote for herself, whereas the letters were written for someone else. These features seem to have determined the writing process. For instance, in the diary, the grammatical subject is sometimes absent, inasmuch as the identity of the agent of the process is obvious:

(9)   *Arrive une petite script-girl du "Flore" qu'ils emmènent avec eux, elle est hagarde de tristesse et de peur, ils ne pensent qu'à foutre le camp. <u>Raconte</u> qu'il y a eu avant-hier soir un formidable accident de chemin de fer.* (*Diary*, 5 September 1939)

A little script girl comes from the "Flore" and they take her with them, she is crazed with sadness and fear, they think only of buggering off. Tell how the day before yesterday there was a terrific train accident.

The subject is regularly absent when it corresponds to the writer, as in the sentence below, where there is neither subject nor verb:

(10)   *Levée dès 8h. du matin — à 8h. ½ au fond du "Dôme", dans la pénombre avec un café et des journaux.* (*Diary*, 8 November 1939)

Up at 8 a.m. – at 8.30 a.m. in the "Dome", in the half-light with a coffee and newspapers.

By contrast, in the letters, the subject and verb are always formulated:

(11)   *Je me suis levée à 8 h. bien juste et à 8h. ½ j'étais au fond du "Dôme"* (*Letters to Sartre*, 8 November 1939)

I woke up at 8 a.m. precisely, and at 8.30 a.m. I was in the "Dome".

3.3   Methodology and quantitative data

For both corpora, we identified the VS on the basis of the description provided in the Introduction, and classified them according to the criteria set out below ("Corpus data and analysis: typology and textual development of verbless sentences"). We thus obtained three lists of VS, which were indexed according to their position and role in each corpus.

On the basis of data yielded by the social reports, we hypothesized that VS generally serve an instrumental purpose in this particular discourse genre.

There seemed to be no absolute correlation between the number of versions of a social report and the instrumental use of VS, inasmuch as some reports that had undergone only a limited amount of rewriting (Files 3, 4, 8, 9, 20…) still contained instances of the instrumental use of these constructions, as shown in Table 1.

Table 1. Correlation between the number of versions and the instrumental use of verbless sentences (VS) in social reports

Files	No. versions	Reports containing no VS
File 1	2	0
File 2	4	0
File 3	2	
File 4	4	
File 5	3	0
File 6	4	0

(*Continued*)

Table 1. (Continued)

Files	No. versions	Reports containing no VS
File 7	5	0
File 8	4	
File 9	4	
File 10	10	0
File 11	7	
File 12	6	
File 13	6	0
File 14	5	
File 15	24	
File 16	16	
File 17	30	
File 18	12	
File 19	41	
File 20	3	
File 21	5	
File 22	6	
File 23	5	
File I Inputlog	12	
File II Inputlog	14	
File III Inputlog	9	
File IV Inputlog	13	
Total: 27 files	**Total no. versions: 256; Mean: 9.5 versions per file**	**7 files containing 0 VS**

Table 2. Total number of verbless sentences correlated with the number of reports

Number of reports	27
Number of reports featuring the instrumental use of VS	20
Total number of VS affected by the rewriting process	126
Mean number per report (out of 20)	6
Cases of reduction	6

We also took verbless subclauses (Lefeuvre 1999, 29) into account. These were quite few and far between in our social report corpus, and were mostly concentrated in the realtime recorded Inputlog files. They mainly took the form of nominalizations used as compressing tools.

4. Corpus data and analysis: Typology and textual development of verbless sentences

This section offers a detailed insight into the different types of verbless sentences we were able to identify in our corpora on the basis of several heterogeneous criteria: referential content, grammatical structure (and grammatical and semantic features of the reformulated unit), and status in the text.

4.1 Referential content

One of the first questions that emerged when we studied the list of VS concerned their denotation: which types of referential content were most frequently denoted by the VS? We were able to identify three categories in the social reports, which also apply to other discourse genres, including diaries:

a. VS denoting topics to be developed, such as "Ophthalmology appointment"
   [*Bilan ophtalmo et orthoptiste*; *port de lunettes de confort* (File II), *les retours des week-ends* (Files15, 17, II)]

b. Adjectival VS characterizing the child or a member of its family, such as "Very opposed to the idea of meeting his father"
   [*Très opposée à l'idée de rencontrer son père* (File 10)]

Here, we can see that there are no subjective verbless sentences, such as:

> *Très bien élevé, ce prince monténégrin.* (Daudet, *Tartarin de Tarascon*, cf. Lefeuvre 1999)
> Very well brought up, this Montenegrin prince.
> *Une belle ville, Toulouse.* (Bernanos, *Un Crime*, cf. Lefeuvre 1999)
> A beautiful city, Toulouse.

> *Fameuse, hein,* […]*, cette soupe à l'oignon.* (Queneau, *Zazie dans le Métro*, cf. Lefeuvre 1999)
> Delicious, ain't it, […], this onion soup.

> *Le plus précis des hommes, ce Bernard.* (Mauriac, *Thérèse Desqueyroux*, cf. Lefeuvre 1999)
> The most precise of men, this Bernard.

c. Verbless sentences denoting events, such as "17.06.09: interruption of psychomotor rehabilitation sessions"
   [*Depuis le 17. 06. 09: interruption des séances de psychomotricité* (File 20)]

Although some of the VS in the "topics to be developed" category could also represent types of events, they had a specific status in the development of the text that allowed us to distinguish them from the "events" category.

4.2   Grammatical structure of the VS

Quite a few of the VS identified in the social reports and diary entries had a nominal form, and some were straightforward action nominalizations (see examples above). From a textual point of view, in the social reports, we could distinguish between the subtitles, which remained unchanged in most cases, but signaled the expansion of the text in the rewriting process, and the events (roughly speaking),[4] which were generally affected by the rewriting process and turned into verbal sentences in the majority of cases.

As shown in Table 3, only the subtitles had a saturated syntactic structure and could be represented by NPs. Bare nouns and action nominalizations characterized both categories (i.e. events and subtitles). Nonsaturated patterns corresponded to events that would be reformulated as verbal sentences in subsequent drafts of the text (see "*Grammatical structure and semantics of the reformulated unit*" below); they therefore constituted "to-be-filled-in" patterns.

Table 3. Syntactic structure and discursive function of verbless sentences in social reports

	Subtitles	Events
Noun phrase	*L'environnement familial, La scolarité, La santé* The family environment, Schooling, Health	–
Bare noun	*Problématique à l'admission, Fratrie* Problem at admission, Siblings	*RDV pédiatre, Dentiste, Poids* [File 15] Appointment with pediatrician, Dentist, Weight
Action nominalization (bare)	*Evolution sur le groupe, Prise en charge du quotidien* Progress within the group, Day-to-day care	*Evocation de vécus difficiles post-placement* [File 16], *Appel téléphonique, Départ et retour du dimanche* [Files 15, 17] Evoking difficult experiences post-fostering, Phonecall, Departure and return on Sundays

4.3   Grammatical structure and semantics of the reformulated unit

Most of the VS in the social reports were reformulated and expanded into verbal clauses. Similarly, some of the VS in de Beauvoir's private diary correlated with verbal clauses in her letters to Sartre. These reformulations and expansions took a

---

4.   Bare nouns and action nominalizations do not denote events in the same way. Bare nouns, featured in Table 3, only evoke events.

variety of forms, and seemed to depend on the nature of the VS as well as on the position it occupied in the text's formal and semantic structure.

Plain verbal paraphrase
In the excerpt below, a copula completes the adjectival predicate in the first sentence, and action nominalizations are reformulated as verbs in the following sentences:

(12)

Draft 2	**Final version**
Bonne évolution de Damien. Verbalisation de plus en plus importante de ses affects, peut être dans l'interrogation, s'intéresse de plus en plus à son environnement.	*L'évolution de Damien est bonne. Il verbalise de plus en plus importante ses affects. Il peut être dans l'interrogation et s'intéresse de plus en plus à son environnement.* [File 20; 3 versions]
Good progress by Damien. Increasing verbalization of his feelings, may be in a questioning phase, taking more and more of an interest in his environment.	Damien <u>is</u> making good progress. He is increasingly verbalizing his feelings. He may well have entered a questioning phase, and is taking more and more of an interest in his environment.

The same progression can be seen in the transition from diary to letters, with nominalizations being reformulated as verbal sentences:

(13)

Diary	**Letter**
On va au "Cintra": discussion sur le pacifism. (S. de Beauvoir, *Diary*, Friday 19 November 1939)	*Nous avons été au "Cintra" de la rue du fb Montmartre et nous avons eu une grande discussion sur le pacifisme.* (S. de Beauvoir, *Letters to Sartre*, Friday 19 November 1939)
We go to the "Cintra": discussion on pacifism.	We went to the "Cintra" in Faubourg Montmartre Street and had a great discussion on pacifism.

Text unfolding
The subtitles in the social reports were formulated as VS, as mentioned above. They were sometimes affected by the rewriting operations, but their predicative structure never changed. Subtitles were often already present in the very first draft, as elements of the basic outline, leaving empty fields to be completed as and when the care process and writing process advanced. In some cases, this expansion resulted in the production of several paragraphs filling an entire subsection of the report (see Figure 2).

Evolution sur le groupe	Evolution sur le groupe
	Annie est arrivée sur le groupe de l'Hortensia depuis le mois de juin 2008. Elle semble aujourd'hui moins souffrir de l'éloignement du domicile et commence à réfléchir sur les raisons du placement. E seg. 40 état 4 plus autant vindicative ou, tout au moins, elle ne semble plus dans une perspective immédiate de retour. Lors de la dernière audience, elle a bien entendu les positions communes de Madame LELEU, Monsieur PAUTY et notre service de continuer le placement pour la protéger. Cela a permis de dépasser certains enjeux de loyauté pour s'autoriser à mieux s'installer sur le groupe.
	Jusqu'aux vacances de la Toussaint 2009, nous avons vu Annie évoluer favorablement. En tant que jeune fille la plus âgée sur le groupe, elle s'est placée en leader positif, accessible et disponible. Cependant, Annie est une jeune fille qui cherche à être en bonne relation avec tout le monde et a toujours autant de difficultés à se défendre des autres ou à préserver ses espaces. Par exemple, lorsque les autres jeunes ont remis en cause cette place de grande, Annie n'a pas accepté et a préféré quitter le groupe pour se réfugier dans le parc. Annie reste fragile dans ses relations. Elle peine à s'affirmer face à des personnalités fortes et peut encore se laisser influencer. Par exemple, elle a été revue pour le vol de chaussures avec des autres jeunes du groupe.
	Le comportement d'Annie s'est modifié à partir du déménagement de Madame LELEU à MONTROUGE. Nous avons observé une régression dans ses positions auprès des adultes et des autres jeunes. À travers un discours conforme sur l'intérêt du déménagement, Annie a ressenti le besoin d'être plus souvent en lien téléphonique avec sa mère, allant jusqu'à utiliser excessivement le téléphone du groupe sans autorisation.
	De plus, pendant cette période, Madame LELEU ne disposait pas de véhicule et était dépendante des prêts possibles par sa famille. En conséquence, elle a estimé qu'Annie était suffisamment grande pour rentrer seule le week-end en bus. Etant donné la fréquence des bus, Annie a dû rentrer seule certains vendredis soir. Ces retours tardifs et longs l'ont particulièrement insécurisée.
	Pour ces raisons, nous nous sommes mobilisés pour organiser les transports et Madame LELEU a pu, courant janvier, acquérir un véhicule. Pendant cette même période, Annie a recommencé à affabuler sur ce qu'elle vivait le week-end. Nous retrouvions la jeune fille du début du placement, relatant des événements du week-end qui ne correspondaient pas au discours de la mère. De plus, elle relatait des propos très sexualisés sur son intimité qui ont mis à mal les autres jeunes du groupe.

Figure 2. Example of text expansion after a subtitle (*Evolution sur le groupe*)

Narrative verbal paraphrase
Adverbial phrases of time frequently accompanied the nominalizations. This facilitated the reconstruction of the narrative thread, in both the social reports and the diary entries versus letters:

(14)

Draft 2

Depuis le 17. 06. 09: interruption des séances de psychomotricité.

17.06.09: interruption of psychomotor rehabilitation sessions.

Draft 3

Depuis le 17 juin 2009, les séances de psychomotricité ont été interrompues. [File 20; 3 versions]

Since 17.06.09, there have been no further psychomotor rehabilitation sessions.

(15)

Diary

Travail au "Dôme" après déjeuner. (S. de Beauvoir, *Diary*, 19 November 1939)

Work at the "Dome" after lunch.

Letter

Je me suis donc mise au travail de nouveau à 1h. de l'après-midi. (S. de Beauvoir, *Letters to Sartre*, 19 November 1939)

I therefore resumed my work at 1 p.m.

Verbal paraphrase with inserted modalization
Whereas the social reports elided modalization and other subjectification markers throughout the rewriting process, the objectivation constraint did not apply to

the letters. For instance, a simple nonverbal statement in the diary might have the same modal and indexical adjunctions as its verbal paraphrase. A point of view, in the form of a personal assessment, might thus be added to a factual statement.

(16)

Diary	Letter
Matinée libre. (S. de Beauvoir, *Diary*, Friday 20 October 1939)	<u>*Heureusement*</u> *j'ai un petit temps* <u>*ce matin*</u>*.* (S. de Beauvoir, *Letters to Sartre*, Friday 20 October 1939)
Free morning.	Fortunately I have a little time this morning.

Guidelines for drafting social reports tend to suggest avoiding modalization. However, while some rewriting operations are aimed at deleting modalizations, others may also insert them (see Brunner and Pordeus-Ribeiro, this volume). In rare cases, alongside verbal paraphrases and other additions, modalization may be added to VS, as in the following:

(17)

Draft 14	Draft 15
Appel téléphonique, départ et retour du dimanche.	*Les départs en famille et les retours sur le groupe sont* <u>*généralement chaleureux*</u>*.* […] [File 15]
Phone call, departure and return on Sunday.	The departures to his family's place and his returns to the group are usually warm.

Verbal paraphrase with inserted adverbial

The following example resembles the previous one in several respects, including the addition of modalization and indexicality.[5] Moreover, an adverbial specifies the date. This adverbial supports the argumentation that follows the verbal paraphrase of the verbless sentences: when we compare the dates, we can understand why de Beauvoir found that the letters from Sartre took too long to arrive.

---

5. "As a genre, letters have specific forms of deixis, that is ways of referring to the writer and the intended reader and to space and time. […] Two words are invoked: the here and now of the writer and the here and now of the reader. Shared knowledge is referred to, often explicitely." (Barton & Hall 2000, 6)

(18)

Diary	Letter
Lettre de <u>Sartre</u> – 2 lettres de Bost – je leur réponds. (S. de Beauvoir, *Diary*, Monday 2 October 1939)	*J'ai reçu une <u>petite</u> lettre <u>de vous</u>, <u>du 24</u>; comme elles sont longues à arriver, celles de Bost mettent trois jours de moins.* (S. de Beauvoir, *Letters to Sartre*, Monday 2 October)
Letter from Sartre – 2 letters from Bost – I answer them.	I have received a little missive from you, dated the 24; they take so long to arrive, those from Bost take three days less.

The same strategy can be found in the social reports, although the outcome of the rewriting is more complex, with many formal modifications:

(19)

Draft 3	Draft 4
Très opposée à l'idée de rencontrer son père.	<u>*Depuis les vacances de Toussaint 2007 où Marine avait rencontré son père au ABCD, il n'y a plus eu de contact direct.*</u> [File 10; 10 versions]
Very opposed to the idea of meeting her father.	Since the autumn half-term holiday in 2007, when Marine met her father at the ABCD, there has been no direct contact.

Adverbials are also commonly added when the VS calls for the text to be expanded, and the resulting rewriting produces new paragraphs.

Development through verbal clauses
Verbal clauses can provide more detailed information about a given situation. For instance, in the letters, the writer can explain what has been happening or describe her feelings- in this case her disappointment -in greater detail:

(20)

Diary	Letter
Au concert, plus de place; <u>déception</u>; on va au "Cintra". (S. de Beauvoir, *Diary*, 19 November 1939)	*On a filé en taxi salle du Conservatoire, mais il ne restait plus une place; <u>on a été déconfits, c'était un peu raide d'avoir Delarue tout cru sur les bras</u>; nous avons été au "Cintra" de la rue du fb Montparnasse.* (S. de Beauvoir, *Letters to Sartre*, 19 November 1939)

At the concert, no seats left; disappointment; we go to the "Cintra".	We drove by taxi to the Conservatoire, but there weren't any seats left; we were frustrated, it was a bit much being stuck with Delarue in full throttle; we went to the "Cintra", in the Rue du Faubourg Montparnasse.

In the social reports, the detailed versions could contain far more information than was suggested by the VS. For instance, phone calls and weekend departures and returns were routine events in the association's fostering practice. They were also subject to assessment, inasmuch as the relationship with the family was regularly scrutinized in order to interpret the young person's situation more fully or discuss the pursuance of the social measure.

(21)

Draft 14	**Drafts 15–16**
Appel téléphonique, départ et retour du dimanche.	*Par ailleurs, les appels téléphoniques sont réguliers les mercredis.* [*Nous avons le sentiment que le placement est bien vécu.*] *Toutefois, Anthony exprime peu de choses de son quotidien au domicile tant sur ce qu'il y fait, ce qu'il y vit. Il parle surtout de ces moments privilégiés avec son père lors de dépannage.* [File 17]
Phone call, departure and return on Sunday.	In addition, there are regular phone calls on Wednesdays. [It looks to us as though the fostering is going down well] However, Anthony says little about his everyday life at home, be it about his activities or about his experiences. He mainly talks about the special moments he spends with his father while he is doing his breakdown repair work.

(*Dis*)ambiguating paraphrase

In both corpora, some – rare – cases of reformulation seem to contradict the semantic content of the verbless sentences. In the excerpt below, *post fostering* and *before fostering* are clearly opposed. Does this mean that the author formulated contradictory statements about the situation? This is not an easy question to answer, and we propose an alternative hypothesis: the condensed formula in Draft 6 is ambiguous, in that either the child could only evoke his/her difficult background since coming

into contact with the carers (i.e. post-fostering), or else he/she found being fostered a difficult experience. This allows us to view the reformulation as a means of clarifying the situation, with the adverb *recently* helping to situate the events chronologically.

(22)

Draft 6	**Draft 7**
Evocation de vécus difficiles __post__ placement.	*Depuis peu de temps Thierry peut évoquer de vécus difficiles avec ses frères et soeurs __avant le__ placement.* [file 16]
Evocation of difficult personal experiences <u>post</u> fostering.	Recently, Thierry has become able to talk about the difficult personal experiences he had <u>before</u> being fostering with his brothers and sisters.

In the same way, the letters to Sartre seem more precise than de Beauvoir's diary:

(23)

Diary	**Letter**
Lever <u>vers 8h. ½.</u> (S. de Beauvoir, *Diary*, 22 December 1939)	*On s'est levées <u>à 8h.</u>* (S. de Beauvoir, *Letters to Sartre*, 22 December 1939)
Up at around 8.30 a.m.	We rose at 8 a.m.

In the adverbial clause, we can see that the vague and semantically underdetermined preposition *vers* (~about) has been replaced by the preposition *à* (~at), which has a precise spatial or temporal value.

The types mentioned below were only observed in the social reports, owing to the specific nature of this written genre and the lengthy writing process.

Verbal paraphrase with evaluation and focus modification
The textual change signaled below in example (24) is probably due to the report's very long gestation period. All the cases involved long-term social measures, and it could take several weeks, if not months, to complete each report. Going by the dates, Draft 10 was written prior to or simultaneously with the room change. The verbless statement is factual, possibly anticipatory, whereas the statement in Draft 31 is clearly posterior and assesses the situation arising from the room change. This is therefore very different from the case of de Beauvoir's diary and her letters to Sartre, which were written on the same day.

(24)

Draft 10 (14 January)	**Draft 31 (17 March)**
Changement de chambre à compter du début janvier.	*Annie occupe depuis le début de l'année une chambre individuelle qu'elle a plutôt bien investi.* [File 19]

	Since the New Year, Annie has been in a single room that she has very much made her own.
Room change early January.	

Objectivating paraphrase

In the example cited below, the rewriting operations deeply affect the text. VS are not only reformulated as regular verbal clauses, but some semantic information is profoundly modified. Thus, the references to the author's/team's impressions ("doubts", "picture") are deleted, and elements forming part of the social assessment are replaced by descriptive statements. This strategy is in line with guidelines for drafting social reports (cf. Huyette & Desloges 2009).

(25)

Version 3	**Version 15**
~~*Interrogation de l'équipe*~~ *sur ses capacités d'élaboration et de compréhension sur les relations avec ses pairs.* ~~*Constat*~~ *d'une maturité affective faible et besoins éducatifs important. Positions éducatives fortes, vigilance sur le quotidien et constat d'une jeune pouvant se satisfaire de cet encadrement.*	*Annie est une jeune qui cherche à être en bonne relation avec tout le monde et à toujours des difficultés à se défendre des autres ou à préserver ces espaces. Elle reste assez fragile face à des personnalités fortes et peut encore se laisser influencer.* [File 19]
~~Team's doubts~~ about her ability to develop and understand her relationships with her peers. ~~Picture~~ of weak affective maturity and major educational needs. Strong educational stance, careful monitoring of her everyday life and picture of a youngster who can get by with this type of supervision.	Annie is a youngster who always wants to be on good terms with everybody and always has difficulty defending herself and preserving her private space. She remains vulnerable when she comes into contact with strong personalities and may still be easily influenced.

4.4   Status in the text

This section focuses on the longitudinal process of writing social reports, inasmuch as it offers a more direct access to text configuration strategies. It is not always easy to follow the progress of verbless constructions through the various draft versions, and text alignment is of little use if no direct track is available. We can nonetheless identify two kinds of writing situations: (i) the use of VS as key concepts, which can be tracked across the drafts using semantic analysis and keyword searching; and (ii) the use of VS and nonfinite sentences or clauses as "to fill in" and "to do" clues, which have more to do with social work practice than with the actual writing process.

4.4.1   *Verbless sentences as key concepts*

In File 15 (24 versions), for instance, two nonverbal constructions share comparable semantic and topical information, and are deleted at different stages in the writing process, while the content is fully developed in an intermediate draft, in a different section of the text:

Table 4. Example (26)

Drafts 1–2	*Draft 7*	*Draft 8*	*Draft 9*	*Draft 20*
	(*Gestion de l'agressivité, gestion d'un conflit*) [Line 32] (Aggressiveness management, conflict management)	*Par exemple, à la suite d'une discussion autour de l'inscription en club pendant les vacances, le jeune est rentré dans une <u>colère</u> avec des propos assez <u>agressifs</u> considérant que nous voulions l'éloigner du groupe.* [Line 35] *Nous observons également qu'Anthony est un jeune qui <u>se contient</u> énormément autant dans ses relations avec ses pairs qu'avec les adultes.*		Line 32 (see draft 7) deleted
		Il évite autant que possible les <u>conflits</u> ou il y a risque de <u>débordements</u>.		
La colère contenue, agressivité sur plus jeune [Line 142] Anger held back, aggressiveness toward the younger ones		*Néanmoins, il apparait par moment <u>des explosions de colère</u>par essentiellement <u>des propos</u>orduriers et quelques fois<u>des gestes agressifs</u> à l'égard des plus jeunes.* [Lines 44–53] For example, after a discussion about enrolling in a holiday club, the youngster became very angry and came out with some rather aggressive words, accusing us of wanting to distance him from the group. [Line 35] We have also noticed that Anthony is a youngster who controls his emotions a lot, in his relations with his friends as well as with adults. He avoids conflicts where there is a risk of things getting out of hand as much as possible. Nevertheless, he sometimes has fits of rage, using foul language and sometimes making aggressive gestures toward the younger children [Lines 44–53]	Line 142 (see drafts 1–2) deleted	

The final version of the social report preserves the text produced in Draft 8 and featured in the Table 4.

Comparable functioning is found in some of the other files. For instance, the fact that a child is disturbed by frequent moves and has a weak personality is conveyed in verbless sentences in File 19 (41 versions).

4.4.2   *Verbless sentences serving as "to fill in" and "to do" clues*

In the corpus of social reports, a sizeable proportion of VS (27.5%) take the form of section subtitles or lexical elements related to section subtitles that indicate the discourse thread and the sections to be fleshed out. VS function as topic markers. The example of text expansion commented on above illustrates this process.

Along with VS, nonfinite clauses are also worth mentioning. A few of the nonfinite clauses were unrelated to the writing process. For instance, a series of nonfinite segments appeared in Draft 1 of File 20 (3 versions) only to disappear immediately afterwards, leaving no obvious semantic or topical trace in that section:

(27)   *Maintenir les RDZ réguliers avec l'AF de Damien. Maintenir l'organisation actuelle des DVH. Soutenir la scolarité de Damien, et l'aider à adapter ses réactions parfois violentes avec les autres enfants. Accompagner Damien dans ses passages entre sa famille et sa famille d'accueil. Proposer un espace repéré par Damien à Mme VAILLANT.* [File 20, Draft 1, Line 172]

Keep [having] regular meetings with Damien's foster carer. Keep to the current organization of the visiting and staying contact. Support Damien's schooling, help him to modify his sometimes violent reactions to other children. Accompany Damien in the transitions from his family to his foster family. Propose a place spotted by Damien to Mme VAILLANT.

Even so, there is an obvious link with the contents of the conclusion, which is already clear to see in Draft 1:

(28)   [Conclusion]

Le maintien de l'accompagnement de Damien et des soins entamés dans un cadre thérapeutique et éducatif le concernant nous parait toujours adapté. La question de l'accompagnement scolaire de Damien reste centrale, car cette année va être décisive en terme d'orientation scolaire. Concernant les droits de visite, l'organisation actuelle paraît adaptée à Damien et en adéquation avec les capacités de Madame VAILLANT. [File 20, Draft 1, Lines 219–221]

We believe that maintaining Damien's educational and medical care, initiated within a therapeutic and educational framework conceived specifically for him is a suitable solution for the situation. The question of schooling support for Damien remains central, inasmuch as this is a decisive year for his future educational choices. As for visiting rights, the current system seems to suit both Damien's current state and Madame VAILLANT's capabilities.

Our data suggest that nonfinite clauses are future- and action-oriented, whereas nominalizations tend to refer to past or present facts or judgments.

5.   Interpretation

5.1   Role of the verbless sentences in text configuration

VS are, to all intents and purposes, textual elements. Talking about VS in finished texts, Behr (2011, 243) specifies that some of them are discursively nonautonomous, inasmuch as they are semantically – and probably also syntactically – entrenched in the preceding and following phrasal constructions, serving as a kind of intratextual link.

In several types of texts, we find that VS are used to scaffold the discourse (see Lefeuvre 2004, 2007). In media discourse, for instance, VS help to organize the text, as with *une certitude* in the following excerpt:

(29)   *Une certitude, le chômage va s'aggraver: il pourrait atteindre 10,5 % fin 2012 selon l'OCDE.* (*Le Nouvel Observateur*, 19–25 January 2012)

An absolute certainty, unemployment will get worse: it could reach 10.5% by the end of 2012, according to the OECD.

This can also happen in discourse genres where the text is mainly composed of verbal sentences, as in de Beauvoir's letters:

(30)   a.   *Un truc amusant: j'ai lu au vendredi 10 Novembre un passage sur vous qui m'a surprise par un certain ton légèrement pompeux, de pathétique contenu, de simplicité tendue* […]. (S. de Beauvoir, *Letters to Sartre*, 12 December 1939)

A funny thing: I read a passage about you on Friday 10 November which surprised me by its slightly pompous tone of contained pathos and strained simplicity.

b.   *Juste un petit mot avant de partir pour Quimper. J'ai eu une journée bien occupée* […]. (S. de Beauvoir, *Letters to Sartre*, 19 September 1939)

Just a word before departing for Quimper. I have had a rather busy day […].

Conversely, in the diary entries and draft social reports, VS are regularly used to relate events. Verbal sentences then replace these VS in the letters and final reports.

The fate of VS in the social reports is mainly tied up with chronological anticipation in the narrative plot. As the "life-giving" (Jespersen 1924, 139) elements of verbs, such as time, mood and person, are absent, the reconstruction of the chronology is free from all but contextual constraints (see Plane et al. 2010;

Cislaru 2014). Sometimes, as in the example evoking the room change (24), VS may anticipate or accompany social acts, thus enhancing the text's performativity. Linguistic changes may reflect changes in the nature of the social intervention, or in the criteria used to assess the situation. It is worth recalling that social reports help judges to reach decisions about children's situations and potential social worker involvement. Therefore, only relevant information should be kept in the final version, which is why some VS are deleted:

(31)   *Un week-end par mois à la Clairefontaine.* (… *L'aider à comprendre l'histoire de sa mère…*) [File 19]

One weekend per month at Clairefontaine (… Help her understand her mother's story…)

In the end, the mother's story was deemed to be irrelevant to the assessment. Other verbless and verbal constructions, concerning fantasizing verging on mythomania, were also deleted during the writing process. The deleted content might have seemed too subjective to the judge, and none of the information contained in the VS was included in the final report.

5.2   Writing for oneself versus writing for others

Adopting a more general approach, we can say that syntactic choices are contingent upon the communication situation, and that there are two very different communicative attitudes: writing for oneself (e.g. diary entries, draft reports) and writing for others (e.g. letters and final reports). For instance, when de Beauvoir was writing for herself, she could just jot down pieces of information without any further explanation, because she knew what she meant and was the regular reader of the diary. When she was writing to someone else, such as Sartre, she had to deploy those pieces of information and introduce details – possibly embellished with a few rhetorical flourishes – if she wanted her letters to be understood.

Similarly, draft social reports are for internal use only, and their authors are the texts' sole addressees. By contrast, final reports are addressed to judges, families, and the whole social service. They therefore need to conform to discourse norms and constraints, and be clear, relevant, and performative (see Bach 1992). Furthermore, when we compared the drafts recorded by the authors and the drafts recorded by Inputlog, we noticed slight differences in the way that the VS were employed. VS were less frequent in the Inputlog corpus, and in three out of four Inputlog files, they consisted exclusively of sequences (subclauses) within longer clauses containing verbs. The paradox of the observer, and the implicit presence of the linguist may account for the specific status of the Inputlog subcorpus, which was halfway between writing for oneself and writing for others. These data are

not entirely reliable because of the discrepancy in the volume of the two draft subcorpora. Nevertheless, this difference would be worthwhile exploring in future research.

6.   Conclusions

In conclusion, VS are useful tools for studying texts and the process of writing them. VS are ambivalent, in that they may either compress or elaborate a text. They may serve as a basis for either objectivation or subjectification. In our two corpora, the effect of the rewriting operations mainly took the form of linguistic and semantic deployment. The way that VS are handled is conditioned by

–   Communicative needs (writing for oneself vs. writing for others);
–   The conditions and chronology of the writing process (availability and relevance of facts and other information);
–   Genre constraints and rhetoric (for diaries).

Last but not least, the study of VS and their paraphrasing in two types of textual configurations offers new insights into the status of these constructions, and highlights the numerous hypotheses concerning the syntactic and semantic operations involved in VS, such as ellipsis, addition and predicativity.

References

Bach, Ulrich. 1992. "From private writing to public oration: The case of Puritan wills. Cognitive discourse analysis applied to the study of genre change." In *Cooperating with Written Texts. The Pragmatics and Comprehension of Written Texts*, ed. by Dieter Stein, 417–436. Berlin & New York: Mouton de Gruyter. DOI: 10.1515/9783110881196.417

Barton, David, and Nigel Hall. 2000. "Introduction." In *Letter Writing as a Social Practice*, ed. by David Barton, and Nigel Hall, 1–14. Asterdam – Philadelphia: John Benjamins. DOI: 10.1075/swll.9.01bar

Behr, Irmtraud. 2011. "Etude d'énoncé averbal à un terme: entre grammaire et discours." In *Les énoncés averbaux autonomes entre grammaire et discours*, ed. by Florence Lefeuvre, and Irmtraud Behr, 239–252. Paris: Ophrys.

Benveniste, Emile. 1966. "La phrase nominale." In *Problèmes de linguistique générale I.* Paris: Gallimard. DOI: 10.7202/500389ar

Cislaru, Georgeta. 2014. "Contraintes linguistiques et contextuelles dans la production écrite." *Les Carnets du Cediscor* 12: 55–74.

Delorme, Benjamin. 2004. *Les énoncés nominaux dans la fiction contemporaine de langue anglaise, implications sémantiques et pragmatiques de la prédication averbale*. Paris 4 Sorbonne: Ph.D. thesis.

Eid, Mushira. 1991. "Verbless Sentences in Arabic and Hebrew." In *Perspectives on Arabic Linguistics III. Papers From the Third Symposium on Arabic Linguistics*, ed. by Bernard Comrie, and Mushira Eid, 31–61. Amsterdam - Philadelphia: John Benjamins. DOI: 10.1163/157005893x00237

Fuchs, Catherine, Almuth Gresillon, Jean-Louis Lebrave, Jean Peytard, and Josette Rey-Debove. (Eds.). 1987 [1982]. *Genèse du texte: les modèles linguistiques*. Paris: CNRS.

Grinshpun, Yana. 2011. "Phrase averbale et presse écrite: le cas des constructions en [*UN*+N+expansion]." In *Les énoncés averbaux autonomes entre grammaire et discours*, ed. by Florence Lefeuvre, and Irmtraud Behr, 187–203. Paris: Ophrys.

Hengeveld, Kees. 1992. *Non-Verbal Predication. Theory, Typology, Diachrony*. Berlin & New York: Mouton de Gruyter. DOI: 10.1515/9783110883282

Huyette, Michel, and Philippe Desloges. 2009 [1997]. *Guide de la protection judiciaire de l'enfant – Cadre juridique, pratiques éducatives, enjeux pour les familles*. Paris: Dunod.

Jespersen, Otto. 1924. *The Philosophy of Grammar*. London: George Allen & Unwin. DOI: 10.1017/s0009840x00034715

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London - New York: Routledge. DOI: 10.1017/s0332586500003127

Lardilleux, Adrien, Serge Fleury, and Georgeta Cislaru. 2013. "Allongos: Longitudinal Alignment for the Genetic Study of Writers' Drafts." In *Computational Linguistics and Intelligent Text Processing* LNCS 7817, 537–548. Berlin: Springer. DOI: 10.1007/978-3-642-37256-8_44

Lefeuvre, Florence. 1999. *La phrase averbale en français*. Paris: L'Harmattan. DOI: 10.3406/igram.2001.2732

Lefeuvre, Florence. 2004. "Le prédicat nominal dans les articles de presse." *Syntaxe et Sémantique* 6: 181–198.

Lefeuvre, Florence. 2007. "Le segment averbal comme unité syntaxique textuelle." In *Parcours de la phrase, Mélanges en l'honneur de Pierre Le Goffic*, ed. by Michel Charolles, Nathalie Fournier, Catherine Fuchs, and Florence Lefeuvre, 143–158. Paris: Ophrys.

Lefeuvre, Florence. 2015. "Les nominalisations prédicatives en discours." In *Le français moderne*, special issue ed. by Véronique Magri-Mourgues: 126–145.

Leijten, Mariëlle, and Luuk Van Waes. 2006. "Inputlog: New Perspectives on the Logging of On-Line Writing." In *Studies in Writing: Vol. 18. Computer Key-Stroke Logging and Writing: Methods and Applications*, ed. by Kirk P.H. Sullivan, and Eva Lindgren, 73–94. Oxford: Elsevier.

Mahrer, Rudolf, and Valentine Nicollier Saraillon. 2014. "Les brouillons font-ils texte? Le cas des plans pré-rédactionnels de C. F. Ramuz." In *Faire Texte. Frontières textuelles et opérations de textualisation*, ed. by Jean-Michel Adam, 223–305. Besançon: Annales littéraires de l'Université de Franche-Comté.

Merchant, Jason. 2004. "Fragments and Ellipsis." *Linguistics and Philosophy* 27(6): 661–738. DOI: 10.1007/s10988-005-7378-3

Nichols, Joanna. 1988. "Nominalization and Assertion in Scientific Russian Prose." In *Clause Combining in Grammar and Discourse*, ed. by John Haiman, and Sandra A. Thompson, 399–428. Amsterdam - Philadelphia: John Benjamins. DOI: 10.1075/tsl.18.15nic

Nordlinger Rachel, and Louisa Sadler. 2006. "Verbless Clauses: Revealing the Structure Within." In *Architectures, Rules and Preferences: A Festschrift for Joan Bresnan*, ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen, 139–160. Stanford, CA: CSLI Publications. ⟨http://privatewww.essex.ac.uk/~louisa/newpapers/nordlinger-sadler-final.pdf⟩

Plane, Sylvie, Thierry Olive, and Denis Alamargot (Eds.). 2010. "Traitement des contraintes de la production d'écrits: aspects linguistiques et psycholinguistiques." *Langages* 177.
DOI: 10.3917/lang.177.0003

Prior, Paul. 2004. "Tracing Process: How Texts Come Into Being." In *What Writing Does and How It Does It. An Introduction to Analyzing Texts and Textual Practices*, ed. by Charles Bazerman, and Paul Prior, 167–200. Mahwah, NJ - London: Lawrence Erlbaum.
DOI: 10.1207/s15427625tcq1602_5

Schleppegrell, Mary J. 2007. "Grammar, the Sentence, and Traditions of Linguistic Analysis." In *Handbook of Research on Writing: History, Society, School, Individual, Text,* ed. by Charles Bazerman, 673–691. Mahwah, NJ: Erlbaum.

Re-writing operations and their effects of meaning

Pascale Brunner & Michele Pordeus Ribeiro

FoReLL A, University of Poitiers, France /
CLESTHIA-Cediscor, Université Sorbonne nouvelle Paris 3, France

This chapter studies the effects of meaning generated by re-writing operations (insertions, deletions, replacements) in a corpus of social worker reports. Intended to describe the situation of a child and his or her family, these reports are produced within the framework of educational guidance and assessment orders. The *proper* functioning of the judicial process is therefore contingent upon the *proper* composition of these reports. In this context, we focus our attention on the modifications undertaken by the reports' authors with a view to making the writing more precise or adjusting the intensity of certain expressions – a composition process that navigates between producing an effect of objectivity and producing an effect of subjectivity.

Keywords: re-writing operations; effects; precision; intensity; objectivity; subjectivity

1. Introduction

The professional discourse of social workers is characterized in part by re-writing procedures. In this chapter, we analyze the features that are modified during the composition of social worker reports, looking for regularities. Re-writing operations can take a variety of forms, including insertions, deletions, replacements, and text movements, as defined by Grésillon (1989). We will undertake a linguistic and discursive analysis of the modifications made by the authors of the reports by using a bottom-up model, which aims to elicit the effects of the modifications on meaning. Lastly, the results will be linked to generic and institutional constraints, which are characteristic of the production of these writings.

Social worker reports produced as part of an educational guidance and assessment process are intended to describe the situation of the child and his or her family. They are addressed to the judge who is to give a ruling on the child's situation,

as well as to the families, who, since 2002,[1] may consult the reports produced by the social services responsible for children and families. These writings are thus "subject to editing constraints which, in turn, are caught between the functions of help versus control – a paradox characteristic of social assistance brought about by the double recipient of the discourse, namely the judge and the family"[2] (Pugnière-Saavedra 2008, 28).

Within this context, our attention was drawn to the recurrence of modifications intended either to *intensify* the text or to make it *more precise*:

Precision:

(1)   *Elise est en bonne santé Ø → Elise est en bonne santé <u>physique</u>*
      'Elise is in good Ø health' → 'Elise is in good <u>physical</u> health'
      (DD, File 2, lines 70–71)

Intensity:

(2)   *Un rendez-vous de médiation a eu lieu entre Carine, ses parents et ses éducateurs au sujet d'une soirée d'anniversaire au domicile des parents qui a « <u>débordé</u> » → Un rendez-vous de médiation a eu lieu entre Carine, ses parents et ses éducateurs au sujet d'une soirée d'anniversaire au domicile des parents qui a « <u>dégénéré</u> »*

      'A mediation meeting took place between Carine, her parents and her youth workers regarding a birthday party at her parents' home which "<u>got out of hand</u>"' → 'A mediation meeting took place between Carine, her parents and her youth workers regarding a birthday party at her parents' home which "<u>ended badly</u>"'

      (HC)

2.   Modifying to be more precise

Strictly speaking, the term *precision* refers to "that rigorous precision in calculations and in the sciences"[3] (Bat-Zeev et al. 2010, 3), but in everyday language it has come to refer to "that which is or has been clearly identified and made explicit,

---

1.   More specifically, since the publication of Decree No. 2002–361 of 18 March 2002 (see Huyette & Desloges 2009, 86). For a more thorough discussion of the linguistic and contextual constraints on social workers' written production, see Cislaru (2014).

2.   "[…] soumis à des contraintes rédactionnelles subordonnées à un paradoxe caractéristique de l'aide sociale, prise entre la fonction d'aide et celle de contrôle, paradoxe matérialisé par un destinataire double du discours: le juge et la famille". Translator's note: all the French citations in the text have been translated.

3.   "[…] cette exactitude rigoureuse dans les calculs et les sciences".

leaving no room for any uncertainty or hesitation"[4] (*Trésor de la langue française*). Writers appear to make these modifications in order to avoid characterizations that are *approximate*, *unclear*, *vague*, *imprecise* or *uncertain* in the text, and make sure that nothing important is left out. As a lack of information is often perceived of and judged to be characteristic of a "vague" and "approximate" discourse (Brunner 2014), the quest for clarity contributes to the goal of providing the information needed to analyze the situation.

Thus, clarity constitutes a means of "attuning the discourse to reality",[5] and allows for the "*creation* of an *effect* of objectivity"[6] (Paveau 2013, 214). In her book on the relationship between morality and language, Paveau observes that "[n]umerous utterances are considered dishonest by others because the utterances do not seem to describe the world as it appears to them".[7] In order to be seen as virtuous, a discourse must express the truth, and this leads the writer to adopt a "virtue of exactitude"[8] (Paveau 2013, 212). The more details that are provided, the more precise the referents, and the more prominent this *virtue*. The quest to accurately *report the facts* is chiefly manifested in the addition and replacement of linguistic elements in order to make the information being transmitted in the text *more precise*.

As Nadeau (1999) indicates, it is difficult to determine the degree of precision of an utterance. It is, however, possible to establish that the degree of precision of Utterance 2 is greater than that of Utterance 1 (Nadeau 1999, 500).[9] This is the way

---

4.   "ce qui est, qui a été bien identifié, explicité et ne laisse place à aucune incertitude, à aucune hésitation".

5.   "ajustement du discours à la réalité".

6.   "*fabriquer* de l'*effet* d'objectivité".

7.   "[n]ombre d'énoncés sont considérés comme malhonnêtes par les agents car ils ne leur semblent pas décrire le monde tel qu'il leur semble être".

8.   "vertu d'exactitude".

9.   For the author, while it "may seem difficult to determine the degree of precision of an utterance, it is sometimes easy to establish that the degree of precision of Utterance E_2 is greater than the degree of precision of Utterance E_1. To do this, one merely needs to show that E_1 can be deduced from E_2, or that E_2 contains elements that make it potentially more easily refutable than E_1 (it being understood that E_1 and E_2 have the same degree of universality). For example, Utterance E_2 ('The train came through between 1 pm and 2 pm') is more precise than Utterance E_1 ('The train came through between noon and 3 pm'). If the train came through between 1 pm and 2 pm, it is possible to deduce that it came through between noon and 3 pm. By contrast, E_2 (singular statement) can be refuted more readily than E_1 (singular statement), because if the train came through between noon and 12.59 pm, or between 2.01 pm and 3 pm, E_2 would be refuted but not E_1" (Nadeau 1999, 500). "[s'il] peut paraître difficile de déterminer le degré de précision d'un énoncé, il est parfois aisé d'établir que le degré de précision de l'énoncé E_2 est plus élevé que le degré de précision de l'énoncé E_1. Pour ce faire,

that modifications have been tackled, - that is, an utterance can only be deemed to be precise in relation to an earlier version of that same utterance. In this context, we can define *precision* as the movement from the general to the particular, leading to a *restriction* of an entity's referential field. This movement can be applied to a range of different entities, including persons, objects, times, and places.

2.1 Precision concerning the identity of persons

This type of precision mainly occurs when a segment that was initially anaphoric is replaced by an expression with an absolute referential value (in particular a proper noun). This modification can be explained by a concern for textual consistency and a desire to reduce the cognitive efforts required of the reader. This process is therefore intended to avoid confusion and make the identification of the protagonists as clear as possible, as the following examples show:

(3) *Vive et curieuse <u>elle</u> participe pleinement aux activités proposées par sa famille d'accueil → Vive et curieuse, <u>Elise</u> participe pleinement aux activités proposées par sa famille d'accueil*

'Lively and curious, <u>she</u> fully participates in the activities proposed by her foster family'→ 'Lively and curious, <u>Elise</u> fully participates in the activities proposed by her foster family'

(DD, File 2, lines 43–45)

(4) *<u>Il</u> est originaire de Rennes où il vit → <u>Monsieur Herisson</u> est originaire de Rennes où il vit*

'<u>He</u> is a native of Rennes where he lives' → '<u>Mr. Herisson</u> is a native of Rennes where he lives'

(DD, File 11, lines 158–159)

(5) *<u>Son</u> souhait est de s'orienter en « Mécanique parc et jardin » → <u>Le souhait de Didier</u> est de s'orienter en « Mécanique parc et jardin »*

'<u>He</u> wants to train as a garden machinery mechanic'→ '<u>Didier</u> wants to train as a garden machinery mechanic'

(DD, File 11, line 29)

---

il suffit de montrer que E_1 est déductible de E_2, ou encore que E_2 contient plus de réfutateurs potentiels que E_1 (étant entendu que E_1 et E_2 ont le même degré d'universalité). Par exemple, l'énoncé E_2 'Le train est passé entre 13h et 14h' est plus précis que l'énoncé E_1 'Le train est passé entre 12 h et 15 h': en effet, d'une part, si le train est passé entre 13h et 14h, on peut en déduire qu'il est passé entre 12 h et 15 h, et, d'autre part, E_2 (énoncé singulier) est plus réfutable que E_1 (énoncé singulier), puisque, si le train est passé entre 12 h et 12h 59, ou encore entre 14h 01 et 15 h, E_2 se trouve réfuté, mais non E_1".

In these examples, replacing the pronouns with proper nouns eliminates a potential ambiguity for the reader regarding the identity of the referent. This is because the proper noun is cognitively more stable (Gary-Prieur 2009, 157; Charolles 2002, 215–216) in that the designation of the referent is independent of contextual variations and, as a result, univocal. Thus, the shift from a pronominal form, in which the identification of the referent relies on contextual features, to a denominal form, by means of a proper noun, is part of this process of making the writing more precise. The final version of the utterance therefore has a greater degree of precision than the initial one.

2.2   Precision regarding objects

Certain referents can be made more precise by attributing a characteristic to them, as in (6) and (7):

(6)   *Elise est en bonne santé Ø → Elise est en bonne santé <u>physique</u>*
      'Elise is in good Ø health' → 'Elise is in good <u>physical</u> health'
      (HC, File 2, lines 70–71)

The adjective *physique* (physical) in (6) modalizes the head noun *santé* (health) and thus restricts its extension. By making this addition, the author emphasizes that the girl's *physical* health is not a problem – an emphasis that could be seen as implying that her *psychological* health is not quite as good. This is a hypothesis that the final recipients of the written product (i.e. the judge and the family) will be able to set against the rest of the information contained in the social worker report.

Example (7) is of particular interest regarding the pragmatic aim of the utterance. It contains the same precision mechanism as in (6), but as the addition of the prepositional phrase here clarifies not simply the nature of the excesses, but also the consequences for the girl, this information could well influence the judge's decision:

(7)   *Elle sait profiter de ses relations d'école et amicales et s'autorise des débordements Ø → Elle sait profiter de ses relations amicales et s'autorise des débordements, <u>sans gravité ni conséquence néfastes pour elle</u>*
      'She knows how to take advantage of her relationships with friends and classmates and exhibits certain excessive behaviors Ø' → 'She knows how to take advantage of her relationships with friends and classmates and exhibits certain excessive behaviors, <u>without any serious or harmful consequences for herself</u>'
      (DD, File 10, lines 55–56)

In the following extract, the modification concerns the definite determiner, which is replaced by a demonstrative determiner: <u>*le* placement</u> (<u>the</u> placement) thus

becomes *ce placement* (this placement). This seems to be part of the same process of trying to make the writing more precise:

(8)   *De grandes difficultés relationnelles sont rapidement apparues entre les mem-*
bres de la nouvelle famille d'accueil et Robert qui n'acceptait pas le placement
→ De grandes difficultés relationnelles sont rapidement apparues entre les
membres de la nouvelle famille d'accueil et Robert qui n'acceptait pas ce place-
ment

'Significant relational difficulties quickly arose between the members of the new foster family and Robert, who did not accept the placement' → 'Signifi-cant relational difficulties quickly arose between the members of the new foster family and Robert, who did not accept this placement'

(DD, File 8, 7–8)

This modification has the effect of removing a potential ambiguity in French between the specific and generic readings generated by the use of the definite article (Riegel et al. 2005 [1994], 154). The replacement of the definite article by a demonstrative makes for greater precision, because the latter removes the potential generic value from the definite phrase and maintains the specific refer-ence: Robert is not against placement in general, but against the placement with this particular family. Precision is therefore achieved by specifying the referent, and it contributes to an accurate evaluation of the situation. Given that the new family has already been mentioned, and this in itself encourages the reader to opt for a specific reading, this example shows just how concerned the author is to eliminate ambiguity from the text as far as possible, in order to avoid the slightest uncertainty.

2.3   Precision concerning time

Temporal and spatial expressions undergo modifications of the same type. Either spatio-temporal information is added that was absent from the previous version, as in (9), or else absolute landmarks are added, allowing the date in question to be clearly identified. No such clarity is provided by the deictic "last year" in (10):

(9)   *Il s'est laissé prendre en flagrant délit Ø → il s'est laissé prendre en flagrant*
délit le 11

'He was caught in the act Ø' → 'he was caught in the act on the 11th'

(HC)

(10)   *L'année dernière Ø, Anne était scolarisée au Collège → L'année dernière*
(99–2000), Anne était scolarisée au Collège

'Last year Ø Anne attended middle school' → 'Last year (99–2000) Anne attended middle school'

(HC)

The addition of the term *désormais* (as of now) provides important information in (11) that makes the events temporally more precise:

(11) *Pour autant, Didier montre Ø certaines capacités d'autonomie qu'il n'avait pas il y a peu → Pour autant, Didier démontre <u>désormais</u> certaines capacités d'autonomie qu'il a récemment acquises.*

'For all that, Ø Didier shows certain autonomous abilities he did not have until recently' → 'For all that, <u>as of now</u> Didier demonstrates certain autonomous abilities he recently acquired'

(DD, File 11, line 72)

Here, attention can be focused on the value of the verbs *montrer* (to show) and *démontrer* (to demonstrate), which can have a *deictic* value in the present tense, insofar as the verb is "closely linked to the moment of speech"[10] and "indicates an event or a state of things contemporaneous with the utterance act",[11] and may even have an *extended* value that is still centered on the moment of speech, but occupies "a more or less extended time span"[12] (Riegel et al. 2005 [1994], 299). The addition of the complement of time *désormais* (as of now) allows the duration to be restricted: Didier did not show autonomous abilities at the beginning of the placement, but has done since Time X, as indicated by the expression *as of now*. This idea of temporal restriction is, however, already present in the phrase *qu'il a récemment acquises* (that he recently acquired), so the addition of *désormais* (as of now) once again attests to the writer's desire to emphasize the change in the child's behavior by accentuating the temporal restriction through the use of several linguistic means, including the adverbs *désormais* (as of now) and *récemment* (recently) and the verb *acquérir* (to acquire) indicating progress.[13]

The replacement of the negative particle "pas" (no) by the temporal adverb "plus" (no longer) in (12) makes the statement more precise by restricting the temporal field for the experience of worrying. In the first version of the utterance, there is no cause for concern at any point in the placement, whereas in the second version, the presupposition generated by *no longer* also implies an improvement in the situation.

---

10. "étroitement repéré par rapport au moment de la parole".

11. "il indique un événement ou un état des choses contemporains de l'acte d'énonciation".

12. "un espace de temps plus ou moins large".

13. As Née, Sitri & Veniard (2013) observed, the adverb *désormais* (as of now) is used fairly frequently in the corpus, where it appears in a particular *pattern* ('as of now' + predicate expressing the ability), as in Example 11. The authors emphasize that this pattern expresses a "common place of progress" which is central to the assessment process undertaken by the social workers, insofar as it highlights the youth workers' interventions with the child.

(12)   *il n'y a pas lieu de s'inquiéter* → *Il n'y a plus lieu de s'inquiéter pour sa santé*
           physique

        'there is <u>no</u> reason to worry' → 'There is <u>no longer</u> any reason to worry
        about his physical health'

        (DD, File 10, lines 7–19)

The presence of temporal elements,[14] as well as elements intended to specify the
identity of the actors and the nature of the objects, allows the reader to clearly
perceive the reality of the events. In all the examples cited above, the operations
of replacement and addition lend an effect of precision and accuracy to the text,
together with an effect of congruency between what is said and what actually took
place, and an *impression* of an objective description of reality.

3.   Objectivity and subjectivity, coexisting effects

This precision produces an *effect of objectivity*. By objectivity, the following quality
is intended: "that which exists independently of all knowledge and ideas"[15] and
which is therefore valid for everyone, not just one or two individuals (Nadeau
1999, 451). In the present context, producing an effect of objectivity involves
implementing a set of modifications intended to adapt the text to the situation
experienced by the child and his or her family. The writer adopts the posture of an
observer and regards the situation that he or she is supposed to describe and give
an opinion on as an *object* that exists outside of the mind and must be understood
as such.

    We do not mean by this that the objectivity sought by the writers can actually
be attained. This idea does, however, influence the composition process, which
is why we prefer to talk about the *effect* of objectivity or objectification. In gen-
eral, objectivity is viewed positively, as the quality of that which is true, rational
and credible. Objective speech implies the use of a language "that is clearly and
precisely defined, that is straightforward and direct", without ambiguity and cor-
responding to reality (Lakoff & Johnson 2003 [1980], 187).

    This positive conception of objectivity is to be found in the *Guide de la pro-
tection judiciaire de l'enfant* (Huyette & Deslogés 2009). The authors of this guide
acknowledge the importance of social worker reports for the sound functioning
of the judicial process: "it is indeed possible to say that, in too many cases, there is

---

14.   Operations making the spatial field more precise were likewise observed.

15.   "ce qui existe indépendamment de toute connaissance ou idée".

only a semblance of an open discussion in the chambers of the family court judge. All too frequently, the debate consists of generalities and sterile rejoinders, as the written evidence on which it is based is *too vague, too imprecise*, such that it prevents the full complexities of each of the concrete difficulties from being properly addressed."[16] Thus, in order for there to be a full debate on both sides, which is a prerequisite for the validity of the judicial process, "it is not sufficient that people meet together. It is not even sufficient that prior to the meeting all the people present had at their disposal the written documents that would serve as the starting point for the debate. For these documents must be sufficiently *detailed* and *specific* to allow for an exchange of precise arguments."[17] Furthermore, experience has shown that families "accept severe criticism when it *corresponds to reality*. By contrast, they do not accept more superficial remarks when they are couched in *imprecise* or *caricatured* terms"[18] (Huyette & Desloges 2009, 357–359, our italics).

Thus, the conditions under which the reports are composed demand that the writing be objective. Social workers are required to provide precise, detailed, well argued, and substantiated written reports, to allow the judge to undertake a comprehensive analysis of the family situation about which he or she has to render a decision. Analysis of the corpus does indeed show that the reports' authors endeavor to describe the situation in as *precise* and *depersonalized* way as possible, in order to leave no room for doubt or for subjective interpretation by the reader.

While the author's position can be that of an observer who describes the situation by regarding it as an object, it is above all that of a subject who evaluates, taking charge of the discourse and leaving evidence of his or her subjectivity in the textual material. The reinforcement of what we can call the *effect of subjectivity* involves a set of modifications intended to adapt the text to the perception the author has of the child's situation. As the situation may be perceived of as more or less serious, this subjective perception will be reflected in the degree of intensity of

---

16. "il est bien possible de dire que, dans un trop grand nombre de cas, il n'y a pas dans le bureau du juge des enfants qu'un simulacre de débat contradictoire. On en reste trop souvent à des généralités, à des réparties stériles, à cause d'une base de départ constituée par des écrits *au contenu trop vague*, *trop imprécis*, qui interdit d'aborder chacune des difficultés concrètes dans toute son ampleur".

17. "il ne suffit pas que des personnes se réunissent. Il ne suffit même pas que toutes les personnes présentes aient eu à disposition avant la rencontre les documents écrits qui serviront de point de départ au débat. Encore faut-il que ces documents soient suffisamment *détaillés* et *précis* pour permettre un échange d'arguments pointus".

18. "acceptent des critiques sévères quand elles *correspondent à la réalité*. À l'inverse, des remarques plus superficielles ne sont pas acceptées lorsqu'elles sont énoncées de façon *imprécise* ou *caricaturale*".

the words the writer chooses to describe the situation. The effect of objectivity is thus accompanied by an effect of subjectivity that is related to the evaluation of the situation. The writer is called upon to choose the *correct* words (of greater or lesser intensity) to describe the situation.

4.   Modifying to adjust the intensity

In recent years, many studies have focused on the theme of intensity and its modulation, which includes both intensification and attenuation.[19] A volume of *Langue française* was recently devoted to this theme (Anscombre & Tamba 2013), highlighting the heterogeneity of its manifestations, as well as the problems of defining intensity, given that it is frequently in competition with the concepts of quantification and scalarity. In order to grasp the diversity of the phenomena encountered when identifying rewriting operations, we propose a broad definition of intensity.

Analyses of our corpus of drafts lead us to conceive of intensity as *accentuation*. It concerns the transition from an indistinct and neutral utterance to a distinct and accentuated one, via a process of rewriting and modification (Kiesler 2000, 232). This accentuation is thus above all a *transformation* that "consists in changing the intensity of the effect that the utterance gives to the interlocutor"[20] (Kiesler 2000, 225). Intensity also establishes a relationship – a *tension* – between two states or properties: "the intensity of a phenomenon X consists of the divergence (or the difference) between two states x_1 and x_2 relative to this phenomenon"[21] (Romero 2007, 59). This difference can cause either a *reinforcement* (an intensification) or an *attenuation* of the content of the initial utterance.[22]

---

19.   See, in particular, Hadermann, Pierrard, and van Raemdonck (2007); Kleiber & Schnedecker (2007); Romero (2001, 2007).

20.   "consiste à changer, chez l'interlocuteur, l'intensité de l'effet de l'énoncé".

21.   "l'intensité d'un phénomène x consiste dans l'écart (ou la différence) entre deux états x_1 et x_2 relatifs à ce phénomène".

22.   A terminological explanation is needed here: Kiesler uses the term *intensification* to refer to accentuation. According to him, this intensification "can take two opposite directions, such that the effect is either reinforcement (increase in intensity), or attenuation (decrease in intensity)" (Kiesler 2000, 225). In this chapter, however, the term *intensification* refers solely to the processes of reinforcement, and thus contrasts with the term *attenuation*.

Utterance$_1$ (state$_1$)		Utterance$_2$ (state$_2$)
content not (or not very) intense	intensification →	*Content intense (or more intense)*
Utterance$_1$ (state$_1$)		Utterance$_2$ (state$_2$)
content intense	attenuation →	*content less intense*
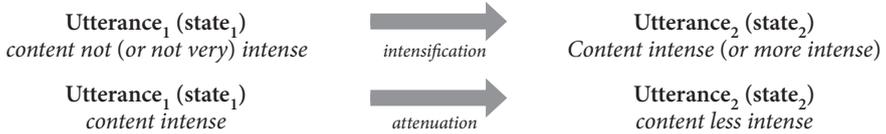
As noted by Romero, intensity is conceived of as a difference – either quantitative or qualitative – between two states and two properties.

In quantitative terms, intensity thus implies an idea of scale: (A) *peu de courage* (not much courage) → (B) *beaucoup de courage* (much courage); (A) *assez gentil* (quite kind) → (B) *très gentil* (very kind),[23] with State B being *quantitatively* greater and more intense than State A. In the corpus, this aspect is expressed either by the addition of a linguistic marker denoting intensity or by the replacement of one element by another of greater or lesser intensity.

In qualitative terms, and from a perspective that is better suited to the data, the contrast is between two lexemes: (A) la situation *déborde* (the situation *gets out of hand*) → (B) la situation *dégénère* (the situation *ends badly*), where the intensity effect results from the opposition between two lexemes with semantic contents of different intensities. It is thus the choice of vocabulary that is decisive here. In the corpus, this aspect of intensity is reflected in the replacements made over the course of composition. While the effect of intensity is brought out explicitly in the analysis through the opposition between the two versions (versions displayed side by side), it also remains perceptible in the final version of the text (the one that is supplied to the judge and the concerned parties), as the terms themselves have a distinct intensity that contrasts implicitly with a norm of neutrality.[24]

In the following section, we list the different types of intensity phenomena, according to whether they are quantitative or qualitative.

4.1   Using quantifiers to increase or decrease intensity

Our analysis of the corpus brought to light a set of modifications dealing with states or properties in terms of *quantity*. In these cases, intensity can thus be defined as

---

23.   As we have said, intensity as we define it here refers to properties and states (not much courage → much courage). Here, intensity contrasts with quantification, which deals with substances, objects, and events (few cars → many cars).

24.   *Norm* refers here both to the idea of *normal* (e.g. the mean on a scale) and to the idea of *normativeness* (prescribed usage in a specific genre).

the "quantitative determination of a property or a state"[25] (Kleiber 2013, 71), and is produced by the insertion of a quantifier, generally an adverb of intensity, which introduces a difference between two states of the same property:

(13)   *Sur le plan du langage nous constatons des progrès importants chez* [*Marjorie*] *notamment depuis son entrée à l'école maternelle. Elle sait Ø se faire comprendre → Sur le plan du langage nous constatons des progrès importants chez Marjorie notamment depuis son entrée à l'école maternelle. Elle sait* <u>très bien</u> *se faire comprendre*

'Regarding language, we have observed that [Marjorie] has made significant progress, particularly since she started nursery school. She knows Ø how to make herself understood' → 'Regarding language, we have observed that Marjorie has made significant progress, particularly since she started nursery school. She is <u>very good</u> at making herself understood'

(DD, File 3, line 39)

The intensification here comes from the addition of the adverb *très* (very), which makes the content of the second utterance more intense in comparison with the initially neutral content of the first utterance. Here is a process of rewriting that leads to the quantification of a property (in this case, *knowing how to make herself understood*, being clear). Other linguistic markers may generate the same effect, as in the following example:

(14)   *En ce sens, la demande de PSP* [...] *qui permettrait ensuite une prise en charge par le service d'accompagnement à l'extérieur reste Ø d'actualité → En ce sens, la demande de PSP* [...] *qui permettrait ensuite une prise en charge par le service d'accompagnement à l'extérieur reste* <u>totalement</u> *d'actualité*

'In this sense, PSP's request [...] that would then allow the child escort service to take charge remains Ø relevant' → 'In this sense, PSP's request [...] that would then allow the child escort service to take over remains <u>entirely</u> relevant'

(DD, File 13, line 123)

The addition of the adverb *entirely* signals the difference between the initial, neutral state and the final state with the more intense content. In this example, the determination of quantity concerns the property of *remaining relevant*: rather than modify the nature of the property, the inserted adverb determines its quantity.[26] In other words, the adverb intensifies the property.

---

25.   "détermination quantitative d'une propriété ou d'un état".

26.   Markers of quantity do not qualify the entity itself, because they are not a property of it. Rather, they determine, within a category X, what the quantity of X is. "This

The deletion of a quantifier leads to the opposite effect: attenuation. In the following example, the adverb of intensity is deleted over the course of successive modifications, thus making the final version of the utterance less intense:

(15)   *Elle a eu un discours <u>très</u> similaire à son père à propos de son malaise et sur le fait que peut être son état de santé était <u>très</u> préoccupant → Nous avons observé des similitudes dans le discours d'Evelyne, sur sa santé, et le discours habituel de Monsieur Roux, qui se veut inquiétant pour ses enfants*

'She had a <u>very</u> similar discourse to her father's regarding her collapse and the fact that perhaps her state of health was <u>very</u> worrying' → 'We observed similarities between Evelyne's discourse regarding her health and Mr. Roux's habitual discourse, which gives his children cause for concern'

(DD, File 16a, line 389)

A substitution can be seen in the following example:

(16)   *Adèle est plus ouverte aux autres qu'auparavant, moins angoissée dans son rapport à autrui et <u>beaucoup</u> plus capable de démarches extérieurs qu'autrefois (sorties avec des copines, prendre le bus…) → Adèle est plus ouverte aux autres qu'auparavant, moins angoissée dans son rapport à autrui et <u>désormais</u> plus capable de démarches extérieures qu'autrefois (sorties avec des copines, prendre le bus…)*

'Adèle is more open to others than before, less anxious in her relationships with others and <u>much</u> more capable of going out in public than before (outings with friends, travelling by bus, etc.)' → 'Adèle is more open to others than before, less anxious in her relationships with others and <u>as of now</u> more capable of going out in public than previously (outings with friends, travelling by bus, etc.)'

(DD, File 13, line 27)

This example highlights two distinct phenomena: attenuation resulting from a deletion, and temporality resulting from an addition. In the first version of the text, the presence of the adverb *beaucoup* (much) indicates an intensification because

---

may seem a trivial point, but, besides the fact that it is regularly highlighted to identify expressions that mark intensity, it crucially signifies that the quantitative variation or modulation is performed not on something that is heterogeneous but on something that is homogenous". "Ce fait peut paraître trivial, mais, outre qu'il se trouve régulièrement mis en avant pour identifier les expressions qui marquent l'intensité, il est essentiel dans la mesure où il signifie que la variation ou modulation quantitative s'effectue, non pas sur de l'hétérogène, mais sur de l'homogène" (Kleiber 2013, 65).

it denotes the degree of a property, namely that of *being capable*. Therefore, its deletion results in an effect of attenuation, an operation that reduces intensity. A second operation is subsequently undertaken and concerns a completely different feature: the addition of the adverb *désormais* (as of now) serves to reinforce the temporal aspect of the utterance. After the modifications, the content of an utterance that had expressed intensity now expresses only temporality.[27]

The following example features the substitution of a modal adverb[28] *forcément* (necessarily) with an adverb of intensity *excessivement* (excessively):

(17) *Ses résultats ne sont pas <u>forcément</u> mauvais → Les résultats scolaires de Didier ne sont pas <u>excessivement</u> mauvais*

'His grades are not <u>necessarily</u> poor' → 'Didier's grades at school are not <u>excessively</u> poor'

(DD, File 11, line 21)

This modification has the effect of intensifying the content of the utterance. While in the first version, the focus was on the possible *mauvaiseté* (poorness) of the grades (*they are poor… but not necessarily*), in the second version the emphasis is on the degree of *poorness* of the child's grades (on the *poorness* scale, his grades are not *excessively* poor, *they do not exceed the norm*).

4.2   Using qualifiers to increase or decrease intensity

Effects of intensification or attenuation can be seen in other types of modification, too. For instance, they can be produced by lexemes whose semantic content includes features that express intensity. Here, it is a question of a *quality* of the word itself - a quality that is perceived of as more or less intense in comparison with a norm that is either implied or made explicit. The issue here is one of *designation* (Kleiber 2001), of the choice of the word that seems most suitable to describe the situation. Thus, we are no longer concerned with the quantitative difference between two states on a scale, but with the difference in semantic intensity between two words.

In the following example, the replacement of the term *rencontre* (encounter) by the term *confrontation* has an effect of intensification, owing to the fact that confrontation includes not only the idea of meeting, but also the idea of a potentially conflictual encounter. Thus, compared with the word *rencontre* (encounter),

---

27.   Given that intensity and temporality are so disparate, we wondered whether we could really talk about replacement, or whether we should conclude that there are two different rewriting operations going on here, owing to the absence of a semantic link between the two.

28.   *Modal* in the sense that the adverb concerns the degree of certainty of the assertion.

the semantic content of the word *confrontation* includes a feature of greater intensity, implying that the encounter is *very* difficult…

(18)   *Anne a été perturbée par la <u>rencontre</u> entre leurs deux problématiques différentes → Anne a été perturbée par la <u>confrontation</u> de leurs problématiques différentes*

'Anne was perturbed by the <u>encounter</u> of their two different sets of issues → Anne was disturbed by the <u>confrontation</u> of their different sets of issues'

(HC)

In the following example, the verb *affirmer* (to assert or maintain) combines the action of *dire* (to say) with the values of conclusiveness and assurance. We should also emphasize that in reporting the father's words, the writer interprets the illocutionary force of the utterance.[29]

(19)   *Il <u>dit</u> craindre de nouvelles révélations de la part de sa fille → Il **affirme** craindre de nouvelles révélations de la part de sa fille*

'He <u>says</u> he fears fresh disclosures from his daughter' → 'He **maintains** that he fears fresh disclosures from his daughter'

(DD, File 2, line 106–107)

The adjective *inquiétant* (disturbing) in the example below qualifies the behavior as being *particulier* (peculiar, unusual), and reinforces this idea of peculiarity by introducing an emotional and alarming dimension that will have a greater impact on the judge's decision:

(20)   *Bertrand manifestait des comportements <u>particuliers</u> → Bertrand manifestait des comportements <u>inquiétants</u>*

'Bertrand displayed <u>peculiar</u> behaviors' → 'Bertrand displayed <u>disturbing</u> behaviors'

(DD, File 16, line 149)

In the following example, the perceived intensification is due to the fact that the verb *dégénerer* (to end badly) implies – in comparison to the verb *déborder* (to get out of hand) – a profound change in nature that begins with boundaries being exceeded. Both verbs have an intense value (as attested by the presence of the

---

29.   The same type of modification can be observed in (11), where the verb *montrer* (to show) is replaced with the verb *démontrer* (to demonstrate). The latter can be perceived of as more intense, insofar as it implies the idea of establishing a truth in a clear and rigorous manner, accompanied by proof.

prefix *dé* in French), but the contrast between the two generates a greater effect of intensity for the latter.[30]

(21)   *Un rendez-vous de médiation a eu lieu entre Carine, ses parents et ses éduca-*
       teurs au sujet d'une soirée d'anniversaire au domicile des parents qui a
       « <u>débordé</u> » → Un rendez-vous de médiation a eu lieu entre Carine, ses
       parents et ses éducateurs au sujet d'une soirée d'anniversaire au domicile des
       parents qui a « <u>dégénéré</u> »

       'A mediation meeting took place between Carine, her parents and her youth
       workers regarding a birthday party at her parents' home which "<u>got out of</u>
       <u>hand</u>"' → 'A mediation meeting took place between Carine, her parents and
       her youth workers regarding a birthday party at her parents' home which
       "<u>ended badly</u>"'

       (HC)

Moreover, the presence of inverted commas here emphasizes the challenge of select-
ing the *right* word. The social worker uses autonymy to indicate that the choice of
these terms is not self-evident and may not adequately reflect the reality of the
situation. All these examples are, in fact, evidence of this concern with correctly
evaluating the child's situation in order to allow the report's recipients to familiarize
themselves with it, too.

5.   Conclusion

In this chapter, we have shown that the composition process navigates between
producing an effect of objectivity and producing an effect of subjectivity. This
tension can be attributed to the generic constraints imposed by the *social worker
report* genre, which requires the text to report the situation in a precise way that
adequately conveys the nature of the situation through the choice of more or less
intense terms. In other words, the social worker has to tailor the text not just to the
reality itself, but also to the perception he or she has of that reality. By proceeding in
this manner, the social worker fulfills the two writing goals here, namely enabling
the judge to grasp the reality of the child's situation in order to give a ruling on it,
and allowing the family to become acquainted with the facts and understand the
decision made by the judge. Emphasizing that there is *no longer* any reason to be

---

30.   *Déborder* literally means *to exceed the boundaries or limits*. As for the verb *dégénerer*, it
comes from the Latin verb *degenerare*, where the prefix *dé* (indicating distancing) is added to
the morpheme *genus* (race, species). The verb therefore contains the idea of a change in nature
and the loss of natural qualities, which is not present in the English expression "end badly".

concerned about the child's physical health implies that there has been a change for the better in his situation (cf. Example 12). Similarly, describing a situation as *disturbing* prompts the judge to take the necessary measures and opens the family's eyes to the difficulty of the situation (cf. Example 20). The *correct* composition of the social worker report is thus a *sine qua non* for the *correct* functioning of the judicial process.

To conclude, we return to the idea of *effect*. The effects of objectivity and subjectivity generated by the modifications undertaken during the composition process not only take account of the perceptual dimension of the interlocutor (including the analyst) but also contain a reference, be it conscious or unconscious, to the social values that constrain professional writing. The analysis thus shows that several factors, not least institutional ones, come into play in the composition process, leading the writer to make modifications. As a result, the text is imbued with objectivity and, at other times, with subjectivity, and these *effects* are perceptible to the interlocutor.

References

Anscombre, Jean-Claude, and Irène Tamba (Eds.). 2013. "L'intensification." *Langue française* 177. DOI: 10.3917/lf.177.0003

Bat-Zeev Shyldkrot, Hava, Silvia Adler, and Maria Asnes. 2010. "Approximation et precision: un champ à explorer." *L'Information grammaticale* 125: 3–5. DOI: 10.2143/ig.125.0.2049244

Brunner, Pascale. 2014. *Le vague, Vagheit. Du mot au concept – pragmatique et folk linguistics.* Limoges: Lambert-Lucas.

Charolles, Michel. 2002. *La référence et les expressions référentielles en français.* Paris: Ophrys.

Cislaru, Georgeta. 2014. "Contraintes linguistiques et contextuelles dans la production écrite." *Les Carnets du Cediscor* 12: 55–74.

Gary-Prieur, Marie-Noëlle. 2009. "Le nom propre, entre langue et discours." *Les Carnets du Cediscor* 11: 153–168. DOI: 10.1017/s0959269512000129

Grésillon, Almuth. 1989. "Fonctions du langage et genèse du texte." In *La Naissance du texte*, ed. by Louis Hay, 177–192. Paris: Librairie José Corti.

Hadermann, Pascale, Michel Pierrard, and Dan van Raemdonck (Eds.). 2007. "La scalarité: autant de moyens d'expressions, autant d'effets de sens." *Travaux de linguistique* 54. DOI: 10.3917/tl.054.0007

Huyette, Michel, and Philippe Desloges. 2009. *Guide de la protection judiciaire de l'enfant: cadre juridique, pratiques éducatives, enjeux pour les familles.* 4th ed. Paris: Dunod.

Kiesler, Reinhard. 2000. "Où en sont les études sur la mise en relief?" *Le français moderne. Revue de linguistique française* 68 (2): 224–238.

Kleiber, Georges. 2001. "Remarques sur la dénomination." *Cahiers de praxématique* 36: 1–22.

Kleiber, Georges. 2013. "À la recherche de l'intensité." *Langue française* 177: 63–76. DOI: 10.3917/lf.177.0063

Kleiber, Georges, and Catherine Schnedecker (Eds.). 2007. "L'intensité." *Travaux de linguistique* 55. DOI: 10.3917/tl.055.0007

Lakoff, George, and Mark Johnson. 2003 [1980]. *Metaphors We Live By.* Chicago: University of Chicago Press. DOI: 10.1007/s11616-003-0093-6

Nadeau, Robert. 1999. *Vocabulaire technique et analytique de l'épistémologie*. Paris: Presses Universitaires de France. DOI: 10.7202/004947ar

Née, Émilie, Frédérique Sitri, and Marie Veniard. 2013, November. "Pour une approche des routines discursives dans les écrits professionnels." Paper presented at the conference entitled "La phraséologie: les collocations aux séquences figées," Grenoble, France.
DOI: 10.1051/shsconf/20140801195

Paveau, Marie-Anne. 2013. *Langage et morale. Une éthique des vertus discursives*. Limoges: Lambert-Lucas.

Pugnière-Saavedra, Frédéric. 2008. "Quelques régularités des écrits du signalement." *Les Carnets du Cediscor* 10: 21–36. DOI: 10.1177/14614456100120040603

Riegel, Martin, Jean-Christophe Pellat, and René Rioul. 2005 [1994]. *Grammaire méthodique du français*. Paris: Quadrige/PUF. DOI: 10.1017/s0959269500003549

Romero, Clara. 2001. "L'intensité en français contemporain: analyse sémantique et pragmatique." Ph.D. Dissertation, University of Paris 8. DOI: 10.3406/igram.2002.2687

Romero, Clara. 2007. "Pour une définition générale de l'intensité dans le langage." *Travaux de linguistique* 54: 57–68. DOI: 10.3917/tl.054.0057

Linguistic forms at the process-product interface

Analysing the linguistic content of bursts of production

Thierry Olive & Georgeta Cislaru

Centre National de la Recherche Scientifique & Université de Poitiers, France / CLESTHIA, Université Sorbonne nouvelle Paris 3

In the present study, we adopt a cognitive-discursive approach to analyse the linguistic structures of bursts of production in a corpus of reports by social workers about children at-risk. Bursts were identified as periods of fluent writing between pauses of at least two seconds, and were coupled with textometric analyses of the final texts. We focused on repeated segments (RS) of texts, i.e. sequences of at least two linguistic units that are repeated at least twice in the corpus. Preliminary analyses showed that the number of bursts with identical or nearly identical content to repeated segments in the texts was limited. Morphosyntactic and semantic descriptions of RSs and bursts indicated that short and medium-sized bursts mainly corresponded to complete syntactical constituents, whereas short and medium RSs often correspond to incomplete syntactical constituents. All together, this study offers information on the structure of the language that is produced during bursts, and thereby raises further questions about the status of routines both at the discursive and psycholinguistic levels.

Keywords: bursts of writing; repeated segments; routines

1. Introduction: Linguistic forms at the process-product interface

This chapter aims to examine the nature and behaviour of linguistic forms on both sides of the process-product interface: the writing process in a real-life situation vs. text and its specific features. The comparison is structured in relation to a more general question concerning writing practices and the relationship between textual data and data related to the writing process. To what extent are text data relevant to the understanding of writing practices? And conversely, are writing

practices predictive of the configuration of the text as a finished product of the writing process? The aim is to add a bit more linguistics and linguistic analysis to the description of the process, and to connect the text more directly to the conditions of its production.

Our study falls within the principles of a pluridisciplinary corpus analysis. The questions addressed in this chapter are anchored to four domains: linguistics, textual genetics, natural language processing (NLP), and psycholinguistics. Here we combine data from all four of these fields to assess how concepts from each of them may be put together to better understand writing practices, but also to produce a new research heuristic.

We collected a corpus of reports on children at risk written by social workers. The corpus was annotated with real-time data recorded by a keylogging program (Inputlog; see Leijten, Van Waes & Van Horenbeeck this volume) while the social workers were typing the reports. In order to detect specific or recurrent linguistic structures, we performed a twofold analysis. First, we analysed the content of the bursts of writing. The term "bursts" of writing refers to strings of text that are produced without major interruption. In other words, bursts are segments of text that are produced between two consecutive pauses. Second, we analysed the content of repeated segments, which are linguistic strings that are reiterated within a text or a corpus.

Our goal was to determine whether bursts and repeated segments are similar by comparing their content. If they are, then bursts and repeated segments, which are behavioural and textual observations respectively, should be considered to reflect similar phenomena in both writing and text. However, if the two do not coincide, they should be considered to reflect distinct writing phenomena. Thus, we compared the linguistic composition of bursts with re-occurring segments of discourse and, more specifically, with the linguistic forms present in repeated segments.

The first Section (§ 2) of the paper presents the notions of burst and repeated segment as well as a brief state of the art of the theoretical questions to which they are related: respectively, writing skills and discourse routines. These are followed by a Section (§ 3) which clarifies the methodological framework and the nature of the corpus. Finally, we present (§ 4) and discuss (§ 5) the results of the analysis, which are both quantitative and qualitative.

2. Bursts of writing and repeated segments of text

2.1 Bursts of writing

At the behavioural level, the activity of a writer can be described as a sequence of periods of handwriting (or typing) – i.e. bursts of production – separated by

pauses. Pauses usually take circa 50% of composing time, and they generally occur for cognitive reasons, although they can also result from socio-psychological or physical causes (see Schilperoord 2002). Cognitively speaking, pauses signal the occurrence of writing processes that cannot be carried out simultaneously with handwriting/typing; they may also be the consequence of memory decay, the writer having forgotten what s/he wanted to write. In the latter case, pauses are used to re-instate the intended message. In sum, pauses "are fundamental moments of conceptualization, formulation or control of the message" (Chanquoy, Foulin & Fayol 1996, 37).

By contrast, bursts of production are moments during which writers produce text as such, making up the remaining 50% of writing activity. Bursts of production are thus periods of handwriting (or typing) during which a segment of text is written. It is important to note that, during bursts of language, writers are not only transcribing what has been prepared earlier in the writing process. Instead, the writing cognitive processes of planning, translating and revising[1] can be implemented while handwriting or typing, at least when these latter skills are sufficiently automatized (Olive 2014). For example, Olive and Kellogg (2002) showed that adult writers, but not 9-year-old children or adult writers using an unfamiliar calligraphy, can simultaneously apply planning, translating and revising during handwriting (for similar evidence with writers' eye movements, see Alamargot, Dansac, Chesnet & Fayol 2007). In fact, in adults, translating occurs mostly during handwriting, whereas planning and revision mainly occur during pauses (Alves, Castro & Olive 2008; Olive, Alves & Castro 2009).

Kaufer, Hayes, and Flower (1986) conducted the first study that investigated bursts of production. They showed that adult writers typically compose by producing segments of text with an average length of 9 words. They also observed that more skilled writers composed using larger bursts (four words more on average) than less skilled writers. Since the texts written by the experts were generally rated of better quality than those composed by novices, the authors interpreted this increase in burst size and length as evidence of more efficient translating processes. This interpretation was later confirmed by Chenoweth and Hayes (2001), who found that undergraduate students are more fluent and produce longer bursts when composing in their first language than in their second language (L2); the same finding was observed with students who were more skilled in L2 in comparison to less skilled

---

1. "Planning" refers to psychological processes that operate at a conceptual level for retrieving and organizing ideas. "Translating" refers to the psycholinguistic processes that formulate written language (see also Galbraith and Baaijen this volume). Revision processes are engaged when reviewing the text and assessing its match with the writers' communicative goals (see Brunner and Pordeus Ribeiro this volume; Fenoglio this volume). A more detailed account of these writing processes can be found in Alamargot and Chanquoy (2001).

students. Chenoweth and Hayes (2003), as well as Hayes and Chenoweth (2006), completed these findings by showing that impairment in verbal working memory, a system that is required in translating, consistently decreased burst length and writing fluency. Notably, Hayes and Chenoweth (2006) did not find bursts in the production of expert typists, as if these writers were able to prepare their text completely while typing, and thus to compose without pausing. Using a passive-to-active sentence conversion task, Hayes and Chenoweth (2007) also concluded that translating is strongly involved in bursts of execution.

Bursts are also determined by writers' handwriting/typing skills. Several studies have shown that having a low level of handwriting or typing skills directly constrains writing fluency and text quality. The detrimental effect on text quality is due to the resulting need for a large amount of cognitive resources or attention when handwriting, which cannot be devoted to planning, translating or revising (Olive 2014). Accordingly, writers with limited handwriting or typing skills do not have enough processing capacity to activate high-level writing processes in parallel. They therefore produce their text in short bursts, during which they mainly produce the text prepared during a previous pause: i.e. with a *thinking-and-then-writing* strategy. By contrast, because writers with high level of transcription skills need little if any cognitive resources to produce the text itself, they can activate high-level writing processes as they do so, with a *thinking-while-writing* strategy (Olive 2014). Thus, they are better able to produce longer bursts.

Consequently, automatizing transcription also leads to longer language bursts. For example, writers with a high level of typing skills compose in larger bursts, on average three words more (Alves, Castro, Sousa & Strömqvist 2007). Similarly, fourth-graders with a high level of handwriting skill show larger written language bursts, compose text more fluently, and produce better stories (Alves, Branco, Castro & Olive 2011). More recently, in Alves, Olive, and Castro's (2008) study, half of the participants composed by handwriting and the other half by typing. In the handwriting group, handwriting skill was manipulated by asking writers to use either an uppercase cursive script or their usual calligraphy. In the typing group, typing skill was manipulated by using either a normal or a scrambled keyboard layout. In both modalities, the low-skill groups showed similar reliable decreases in burst length – about six words less – and received lower ratings for text quality.

As this short review shows, at least translating and handwriting determine the duration and length of bursts. Writers with a high level of handwriting skills can devote their available cognitive resources to translating, which can then be maintained longer while handwriting. Moreover, a high level of translating skills

allows writers to prepare longer segments of texts concurrently to handwriting. In sum, the greater a writer's level of translating and handwriting skills, the longer the bursts they produce.

Less is known, however, about the content of bursts, and more specifically about the linguistic structure of the portions of text that are produced during bursts. In fact, only one study has analysed the linguistic structures of bursts (Kaufer et al. 1986). The authors showed that these parts or segments tend to correspond to clauses, since they showed a strong tendency to end at clause boundaries, and less so at phrase boundaries. Thus, according to these authors, writers compose sentences by first selecting a topic, and then by producing and evaluating sentence parts that fit grammatically with the part of the sentence that has already been prepared. If the evaluation is negative, the writer has to either revise the current part or produce an alternative part to follow it. If the evaluation is positive, then the sentence part is added to the current sentence that has already been produced, or that is still in the writer's mind (i.e. in verbal short-term memory).

In this context, we explored the process-product interface by investigating for a possible association between bursts of production and the linguistic forms or structures that are produced during these execution periods. In particular, we aimed to determine whether texts are produced in segments of text that share common structural characteristics. The main questions about bursts that we addressed in this study were the following:

–   What are the linguistic forms produced in bursts?
–   Are there regularities in the content of bursts?
–   Is bursts' content predetermined by specific, defined text structure/patterns?
–   Are there ready-made linguistic structures that can be retrieved from long-term memory and directly written out?

2.2   Repeated segments

Following Lafon and Salem (1983) and Salem (1986), we define "co-occurrences" as "couples of forms that function almost exclusively within idioms"[2] (Lafon & Salem 1983, 162). This definition selects one specific subtype of co-occurrence, namely, repeated segments (RSs): i.e. strings of at least two graphical units that occur together at least twice in a text or a corpus. RSs represent ready-to-speak

---

2.   "[…] des couples de formes fonctionnant presqu'exclusivement à l'intérieur d'expressions figées" (Lafon & Salem 1983, 162).

units (which are somewhat different from collocations): in the framework of textometry and discourse analysis, they are considered as discourse routines that
characterize either a studied language or a type of discourse. In the framework
of corpus linguistics, Sinclair (1991, 2004) showed that all linguistic productions,
oral or written, are half constituted of prefabricated sequences, following an "idiomaticity principle" (cf. Erman & Warren 2000; Kuiper 2009). Biber's corpusdriven studies on multi-word regular sequences (Biber 2009; Biber et al. 2004)
also show a high prevalence of various types of formulaic language in both oral
and written corpora.

The repetition principle suggests the hypothesis of a routinization of discourse, as defined by Wray (2002, 9): sequences of words or other units that seem
to be prefabricated, which are memorized and reproduced "as is" in the text, and
not generated ad hoc. Along the same lines, Mayaffre (2007, 10) wrote "Repeated
segments of significant length are linguistic tunnels where the creativity of the
speaker/writer is reduced in favour of a kind of recitation."[3] These are strong
hypotheses, which we will test here by comparing the bursts and repeated segments in our corpus. Accordingly, repeated segments are viewed as key elements
of text organization in the framework of discourse analysis and corpus linguistics,[4]
inasmuch as they signal discourse routines related to genre, social sphere of activity, professional domain and occupation, etc.

RSs constitute a "formal" approach to linguistic routines, inasmuch as they are
detected by their graphical form. There are several ways to broaden the insights
that can be drawn from the study of such units. On the one hand, generalized
"grammatical patterns" may be identified as regularity in some subset of the
repeated segments (Hunston & Francis 2000). In this we follow Biber's (2009) and
Biber et al.'s (2004) work on lexical bundles. On the other hand, the semantic types

---

3.   "Les segments répétés de longueur importante sont des tunnels linguistiques dans lesquels
la créativité du locuteur recule au profit d'une forme de récitation." (Mayaffre 2007, 10)

4.   The study of co-occurrences is "[…] the first thing to do in order to underline semantic
nets that are shaped in a text or, more precisely, which *shape the text*; the first thing to do
in order to reach the essential features of the text (i.e.: what makes the text a meaningful
linguistic sequence (a 'completeness of meaning' [Détrie, Siblot, Verine 2001, 349]) and a coherent and cohesive assembly of words" (Mayaffre 2007, 8).

"[…] premier mouvement pour pointer les réseaux sémantiques qui se forgent dans
un texte, ou plus précisément *qui forgent un texte*; le premier mouvement pour toucher à
l'essentiel de ce qu'est la textualité (*i.e.*: ce qui fait d'un texte une suite linguistique signifiante
('une complétude de sens' [Détrie, Siblot, Verine 2001, 349]) et un assemblage de mots à la fois
cohérent et cohésif)." [our translation]

of repeated segments may be considered, depending on their lexical-grammatical contents and discursive profile (see also Cislaru et al. 2013), such as:

–   The "waffle" (doublespeak) RSs, determined by the genre or the topic of discourse (*être en/be in, can+speech verb*).
–   RS-genre clichés, related to a type of cognitive activity: analysis, evaluation (*nous avons/we have, nous pensons/we think*).
–   RSs representing structural clichés in French (*de la, lieu de, part de, une fois, quant à, en effet*).
–   RSs representing individual discourse habits (*ce dernier/cette dernière – the latter,* etc.).

We are concerned here with the particularities of a type of professional discourse, the reports of social workers on at-risk children. The early presence of the RS in the drafts might signal a stereotyped form of discourse, reflecting a strongly constrained professional discourse. At first glance, the drafts of social workers' reports in our corpus do not really seem to correspond to such a discourse type. For instance, the longest RS recorded by the machine contains 11 forms. Whereas a few of them emerge beginning in the first two versions of a report, most of these RSs appear no earlier than version 4–6. This deserves to be underlined, inasmuch as it suggests that ready-mades are not automatically activated in the first stages of the writing process (and they are probably not genuine ready-mades).

2.3   Bursts versus repeated segments

To summarize, by contrasting the defining criteria and interpretive hypotheses regarding bursts and repeated segments, we highlight the differences between the two categories, and show that some of the criteria of definition and identification are undecidable, inasmuch as they need to be tested on corpora, and, concerning bursts more particularly, because they have not yet been submitted to detailed linguistic analysis.

As can be seen in Table 1, bursts and repeated segments are not obviously aspects of the same psycho-linguistic interface. A handful of existing studies run in the same direction. For instance, Schmitt et al. (2004) selected recurrent target clusters from corpora and then tested them during psycholinguistic tasks involving native speakers and second language speakers (see also Schmitt 2004). Their results suggest "it is unwise to take recurrence of the clusters in a corpus as evidence that those clusters are also stored as formulaic sequences in the mind" (Schmitt et al. 2004, 147). Schmid also underlines the weak interpretive impact of the notions of recurrence or frequency, and writes: "we seem

to be quite far from having a good grip on the relation between frequency and entrenchment.[5] This is mainly due to the unclear interaction between absolute and relative frequency, or cotext-free and cotextual entrenchment, respectively." (Schmid 2010, 123).

As suggested by Schmid (2010, 102), "patterns of frequency distributions of lexico-grammatical variants of linguistic units correspond to variable degrees of entrenchment of cognitive processes or representations associated with them". Accordingly, because RSs may constitute prefabricated written forms or discourse routines, from a psycholinguistic point of view, they may refer to language forms that are retrieved in a single block from the writer's long-term memory, and that therefore can be written in a single burst. Consequently, if RSs are discourse routines, then the linguistic form of bursts and of RS may be expected to be relatively similar.

Table 1. Contrasting the features of bursts and repeated segments

Criteria	Bursts	Repeated segments
Recognition	Pauses, real-time data	Repetition and identity
Frequency	Does not apply	Defining
Newly created	Yes (No)	No
Memorization	Does not apply	Presupposed
Familiarity	Not expected	Yes, presupposed
Conventional meaning	?	Yes
Context dependence	?	Yes
Competences required	Writing	Discourse
Standard methodology	Real-time analysis	Corpus-driven

3. Corpus and methodology

3.1 Global description of the corpus and of the method of analysis

Our study is based on a corpus of six reports written by social workers (see Table 2) with a key-stroke logging program (Inputlog; see Leijten & van Waes 2006; Leijten,

---

5. (Langacker 1987; see Schmid in press for a discussion). "Entrenchment refers to the ongoing reorganization and adaptation of individual communicative knowledge, which is subject to exposure to language and language use and to the exigencies of domain-general cognitive processes and of the social environment." (Schmid in press).

Van Waes and Van Horenbeeck this volume) during their regular activity of monitoring and evaluation of the situation of foster children (see also Brunner and Pordeus Ribeiro this volume, and Cislaru and Lefeuvre this volume, for a detailed presentation of the writing situation). Inputlog is a computer tool that records all actions that a writer performs with the computer when composing a text using a word processor and keyboard. In the present case, each key press, along with its timing, was recorded, as was each move of the mouse in the text or in the menu of the word processor that the writers were using. All textual operations to modify the text were also recorded, as were all interactions with other programs (web browser, email client, etc.) available on the computer.

Table 2. Corpus description

Reports	6 (from 2 to 6 pages)	Sentences	979
Pages	30	Words	13,701
Paragraphs	305	Words per sentence	14

The linguistic analysis, supported by the application of NLP tools to a corpus annotated with real-time data, is quite new in the field, and requires several methodological adjustments. To perform these, we used a natural language processing program developed by Adrien Lardilleux (Lardilleux et al. 2013) to extract bursts from Inputlog's log files, automatic detection of repeated segments (initially extracted by Le Trameur, a textometry tool),[6] and the alignment of repeated segments and bursts, as shown in Figure 1. The right frame gives counts of repeated segments (columns to the left of text) and bursts (columns to the right of text) recorded in a complete file of drafts for one report. The outermost numbers indicate the total number of units that are represented as by a given set of related repeated segments (far left of the frame) and bursts (far right of the frame). In the left frame, the upper part shows context for the selected repeated segments, while the bottom part shows the neighbourhood of the selected bursts, with temporal data (the first column indicates the timing within the time log for the writing session, and the second column the length of the burst in seconds).

---

6. Developed by Serge Fleury, Université Sorbonne nouvelle Paris 3, http://www.tal.univ-paris3.fr/trameur/ (see Née et al. 2012 for an application on this corpus; see also Doquet and Poudat this volume).

Figure 1. Alignment of bursts and repeated segments, with direct access to the text

Our study is based on hybrid methods of corpus analysis (close to pattern grammar studies: cf. Biber 2009) based on real-situation text production. The list of repeated segments is corpus-driven, and we operate with raw data, without sequence pre-selection. The bursts were produced during a real-time and real-situation writing activity. Both bursts and repeated segments were also described grammatically (in the terms of constituent analysis by using pre-defined grammatical categories) and semantically. Particular attention was paid to cases of homonymy, such as *qu'elle ne* (~ that she not), which introduces a noun determiner as a burst (*le sentiment qu'elle ne* – the feeling that she Verb not…) versus reported speech as a repeated segment (*elle dit qu'elle ne* – she says that she Verb not…). The fact that the same writers produced both bursts and repeated segments is crucial, ensuring that the data for both correspond to the same discourse genre and social activity. Indeed, the homogenous conditions of text production for bursts and repeated segments reinforce the conclusions that can be drawn on the (non-)correlation between the two.

The threshold of frequency for repeated segments was fixed at 2, given the size of the corpus. For the real-time analysis, we opted for pauses between bursts of more than 2 seconds. Bursts were therefore defined as periods of typing separated by pauses longer than 2 seconds. This threshold allowed us to exclude all pauses that resulted from typing movements from the analyses (for example, moving the hands and fingers on the keyboard to reach the next key, or preparing a combination of keys to type a diacritic character), and whose origin thus did not lie in the operations of one of the writing processes.[7]

Table 3 sums up the quantitative characteristics of the studied corpus. The reports were written in 34 sessions of an average of 23 minutes in length, corresponding to a total composing time of 12h57. Within this time, roughly 40% was spent pausing (5h10) and during the remaining 60% (7h47), the social workers typed their text. They produced their text at a speed of 17,6 words per minute, which falls within the normal range for common compositional fluency in adults. Bursts were long (22 sec.) and contained few words (2.6) suggesting that at least some of the writers were not highly skilled typists.

Table 3. Quantitative and temporal parameters associated to the reports

Writing sessions	34		
Total composition time	12h57	Number of analysed RSs	1506
Total pause time	5h10	Number of pauses	5157
Total writing time	7h47	Mean pause length	13 s
Mean session length	23 min.	Mean pre-writing pause	8 s
Writing fluency	17,6 wpm	Mean within-words pause	10 s
Number of analysed bursts	1014	Mean between words pause	11 s
Mean burst duration	22 s	Mean between sentences pause	13 s
Mean burst length	2.6 words	Mean between paragraphs pause	20 s

3.2 Text progression

Text progression is rather linear in contrast to the evolution of drafts, and many authors who work with the text as a finished product consider that "texts are linearly… and also non-linearly developed" (Hoey 2004, 395). However, a detailed study on the rewriting operations involved in each draft for a larger series of social reports (twenty-nine reports in total: see Brunner and Pordeus Ribeiro

---

7. For a recent discussion on the several ways that this threshold has been defined, see Chenu, Pellegrino, Jisa, and Fayol (2014).

this volume) shows that text progression "in process" is not necessarily linear (see also Fenoglio this volume), and that chunks of text are frequently displaced more than once within the text.

The text progression of the reports we analysed is shown in Figure 2. As can be seen, it took the writers between 1 and 11 sessions to compose their reports, suggesting the adoption of very different writing strategies. It may also be noticed that revision sessions can be detected, i.e. when the curve becomes flat, particularly at the end of writing sessions for the reports that were written over a larger number of sessions. By contrast, some reports increased greatly in length between sessions. In sum, social workers differed in their way of completing their reports.
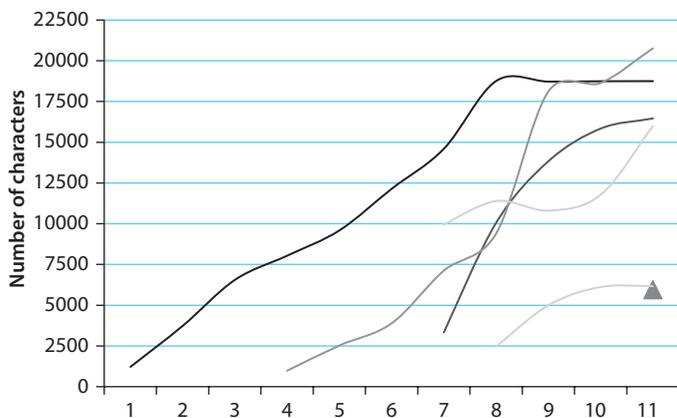
Figure 2. Text progression. The x-axis represents the writing sessions, and the y-axis the number of characters produced. For example, one report was written in 11 sessions, while another was written in a single session

3.3   Pause analysis

Before analysing the content of the bursts and the repeated segments, we first looked at pause data to assess whether the writers who composed the reports could have been differently drawing on the cognitive processes involved in writing. As a first observation, it is interesting to notice that globally, the prewriting pause is the shortest one, even shorter than the mean within-words pause (see Table 3). This may indicate that the writers had already a plan in mind when they began composing their reports. This is not so surprising since the reports include a set of predefined sections on specific topics (life history, daily life in the group, health, school, relations with others, conclusion), which helped the writers structure their reports. This resulted in reduced planning efforts.
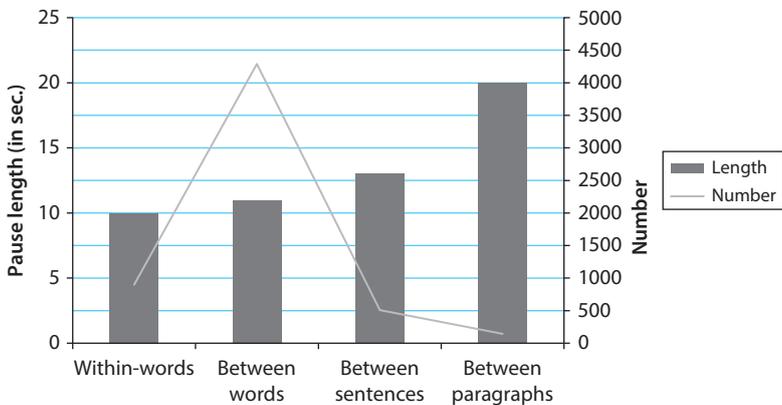
Figure 3. Number and length of pauses longer than 2 seconds

Replicating previous findings on pauses during writing, we observed that the location of pauses in the text strongly influenced their length (Foulin 1995, 1998; Schilperoord 2002; van Hell, Verhoeven & van Beijsterveldt 2008). The shortest pauses were within words, followed by pauses between words, and then pauses between sentences, while the longest pauses were those that preceded a paragraph (see Figure 3). As suggested above, since pause length may be taken as an index of the mental effort that the writer is exerting in constructing the text, our data indicate that pauses before paragraphs involved more mental effort by the writers. This is presumably due to the fact that before writing a paragraph, writers engage planning processes that are cognitively more costly than the formulating pro-cesses activated before producing sentences and words (see Olive 2004, 2012 for a review on the cognitive demands of writing processes). Despite the low planning demands of this writing situation, it is nevertheless possible to conclude that the writers studied here engaged cognitive writing processes in a rather standard way.

4. Linguistic analysis

Less than 3% of bursts and repeated segments converged, i.e. were 75 % similar from a formal/graphic point of view. Given that this ratio is very low, we searched for various comparison criteria, in order to establish a more complex linguis-tic view on the phenomena that the two represent. Our first choice was to con-trast syntactic structures that were specific to bursts and repeated segments. This approach is based on corpus linguistic methods, and more specifically on the col-ligation principle (Hoey 2005; Yamasaki 2008), which allows the identification of pattern types via constituent analysis.

The following is a sample of the most common constructions:

– Noun phrase (NP)
– Prepositional phrase (PP)
– Noun
– Noun phrase + Verb (NP+V, close to sentence-type)
– Verb (auxiliary, participle or other incomplete verb form)
– Verb phrase (VP)
– Adjective
– Adverb
– Adverbial phrase, Connector, Connector & NP or Connector & NP+VP
– Conjunction
– Clause
– Determiner
– NP + preposition
– Preposition + determiner
– Etc.

The classification of lexical strings – either RSs or bursts – demands complex criteria and a number of adjustments to the types of syntactic structures to which they are assigned. The main criterion applied was syntactic saturation, due to its analytical accessibility. This means that we mainly distinguished two categories: saturated strings, which correspond to phrase-type constructions (noun phrases, prepositional phrases, sentences, etc.), and unsaturated strings, which correspond to syntactically irrelevant constructions and units that associate two grammatical groups (phrases), such as *NP+preposition*, or that stop ahead of the boundary of a grammatical group, such as *Preposition+determiner*. However, we are aware that saturation is a notion that is subject to further negotiation. First, syntactic saturation does not always coincide with semantic (i.e. informational) saturation. Some noun phrases, for instance, may be saturated out of context, but unsaturated in discourse use/context; thus, semantically, "*her/his difficulties*" may be either saturated or unsaturated (e.g. "*her difficulties in…*"). We tried to take such cases into account, but a much more thorough semantic analysis is needed. Second, lexical strings are never discursively and interdiscursively saturated: they always maintain and evoke connexions with other words or lexical strings, and this might be an important cognitive/memory factor. Another difficulty is related to the ambiguous status of connectors, which represent discursive functions rather than grammatical categories. Unfortunately, a detailed discourse analysis was not really manageable within the framework of this study. These two main difficulties imposed a limitation on the syntactic criterion. Nevertheless, we took into account a range of semantic data, and thus subdivided the

saturated and unsaturated categories into several subcategories. Full stops and capital letters were treated as graphical criteria, marking sentence boundaries. These criteria allowed us to identify different types of breaks, such as …*other children. She*….

Non-saturated constructions are markers of discontinuity, and can also include:

–   Break before and after full stop: … *other children. She*…
–   Break after coordination: *Alex shows some signs of sadness and/but* [he]…
–   Break after concatenation between a saturated unit and a connector: *She decided to leave. Therefore*…
–   Etc.

Some strings, like clauses, may be regarded as saturated although they are not autonomous. Things get more complicated with longer strings that are peculiar to bursts, which can contain saturated sentence-type strings followed by unsaturated strings, for instance.

It is interesting to study the distribution of these constructions in repeated segments and bursts, in order to verify the linguistic particularities of the two categories, by distinguishing: (i) the distribution of each type of construction; (ii) the distribution of saturated vs. non-saturated constructions.

–   ***RSs and bursts for each linguistic structure***. Table 4 presents the percentage of saturated and unsaturated linguistic structures in RSs and bursts. Overall, the number of saturated and unsaturated structures in the corpus significantly differed ($\chi^2 = 203.6$, $p < .001$). Accordingly, unsaturated bursts and RSs are more numerous than saturated ones. The distribution of bursts and RSs between these two types of structures also significantly differed ($\chi^2 = 42.9$, $p < .001$). More precisely, around 42.6% of bursts are saturated, whereas 32% of RSs are saturated. This difference between bursts and RSs also indicates that the grammatical structures found in repeated segments of text and in bursts differed.

Table 4. Number and percentage (in parentheses) of saturated and unsaturated linguistic structures among all bursts and repeated segments

	RS	Burst
Saturated	483 (32%)	1083 (43%)
Unsaturated	1016 (68%)	1458 (57%)
Total	1499	2541

– **RSs and bursts with saturated structures.** Sentences and clauses are specific to burst production in our corpus; no saturated pattern of this category is attested as a repeated segment. Although numbers may correspond to a single graphical unit (e.g. *2009*) and could have been produced in a single burst, they could not be classified as repeated segments, which required a string of at least two units. It may be noted that saturated noun phrases and prepositional phrases are the structures most frequently found in both repeated segments and bursts, and they are the most frequent ones. Connectors and adverbials are also common to burst and RSs but are less frequent. Additionally, these three types of structures were more present in bursts than in RSs. Verb phrases, NP + conj + NP, clauses, sentences and linked sentences constructions are not frequently used in RSs. Of these, only verb phrases, clauses and sentences appeared in bursts, albeit to a lesser extent than the types shared with RSs.
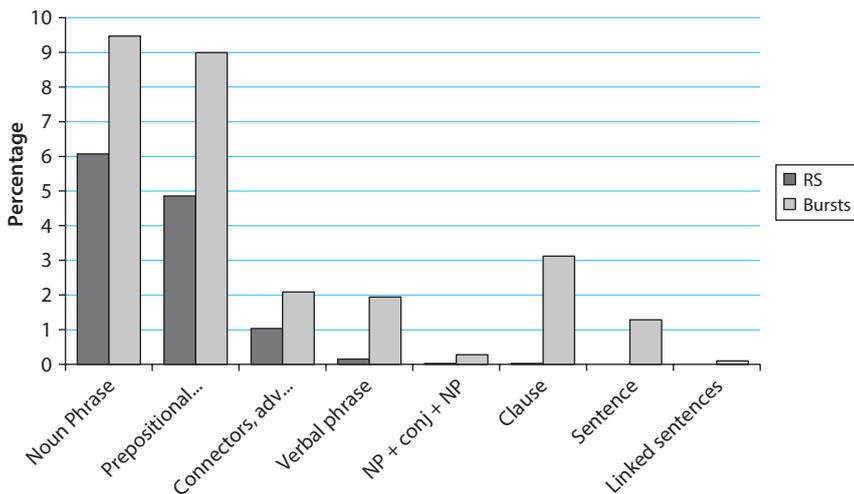
Figure 4. Percentages of bursts and repeated segments represented by different types of saturated linguistic structures

– **RSs and bursts with discontinuities (unsaturated).** Unsaturated constructions appeared in both bursts and RSs. Long items and breaks or double breaks in the context of a boundary marker (full stop, connector, conjunction, comma, etc.; see above) were clearly exclusive to bursts. Repeated segments had a greater association than bursts with only two types of discontinuities: incomplete noun phrases and prepositions followed by determiners or various other items.
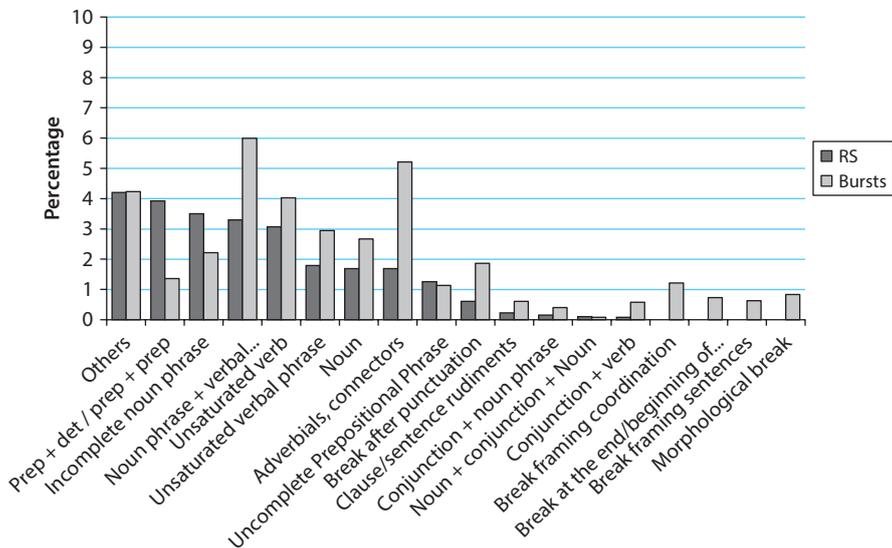
Figure 5. Percentages of bursts and repeated segments in the various unsaturated linguistic structures

5.   Discussion

5.1   Saturated and unsaturated patterns

The concept of unsaturated constructions is similar to that of lexical bundles, as defined by Biber (2009). First, lexical bundles are by definition extremely common (in contrast to most idioms and many 'grammar patterns', which tend to be rare). Second, most lexical bundles are not idiomatic in meaning and not perceptually salient. For example, the meanings of bundles like *do you want to* or *I don't know what* are transparent from the individual words. And finally, lexical bundles usually do not represent a complete structural unit (Biber 2009, 283).

Talking about the idiom-collocation principle, Partington (1998, 19 et sq.) suggests that the use of prefabs facilitates communication processing on the part of the speaker as well as the hearer. But what seems to be obvious – at least at first glance – for oral communication does not function with written communication, where process and product are clearly separated both materially and chronologically. According to Biber (2009), lexical bundles in writing, such as the construction *in the light of,* usually serve to bridge pairs of phrases, and are open-choice oriented on their right border. Indeed, some of the unsaturated constructions in our corpus are open-choice repeated segments. For instance, adverbials followed

by prepositions or subordination markers, as well as prepositions followed by determiners or various other items, are "filled with" relevant units in context; in Biber's terms, they provide a kind of "pragmatic head" for larger units, thus assuming the role of "interpretive frames".[8]

From a cognitive point of view, bursts may be considered to function in a similar way, with the writer having to pause in order to choose the contextually relevant development. However, this interpretation is nuanced by the very limited number of units that bursts and repeated segments share. The affirmation of this similarity thus appears to be cognitively valid and semantically weak, as suggested in Section 5.2. Discontinuity constructions highlight cases where the connection between facts is pre-constructed, and only the discourse elements that are to be connected are selected from a list of possibilities.

Cognitive linguistics (see Schmid 2010 and in press for discussion) is interested in the degree of routinization and automation in the formation and use of a unit. The hypothesis that repetition favours entrenchment is close to the assumptions of corpus linguistics (see above). Yet Schmid (2010, 125) sums up his paper as follows: "What I have tried to show here, however, is that so far we have understood neither the nature of frequency itself nor its relation to entrenchment, let alone come up with a convincing way of capturing either one of them or the relation between them in quantitative terms."

The results of the present study can be discussed first in terms of the communicative competence/performance contrast, along the lines of Hymes' (1971) proposals. Hymes distinguished four kinds of competence skills: knowledge of what is formally possible given the constraints of the language system, on feasibility, on appropriateness and, last but not least, on actually produced sequences. Our study focuses on actual performance, and offers the possibility to confront performance with competence hypotheses. It may be assumed, in the light of the results presented here, that the produced sequences can be separated into two distinct classes, those of process-performance and product-performance. The fact that we found a great number of repeated segments which did not have an equivalent burst may reflect strategies of communicative adaptability (Mey 1998, Verschueren & Brisard 2009), which fits particularly well with the fact that most of the repeated segments did not emerge in the first drafts of writing. This point may suggest an overlap between the use of linguistic prefabs and the shaping of text to conform to social norms. Finally, as noted by Schmid,

---

8.   See Biber et al. (2004) for the three primary discourse functions for lexical bundles in English: (1) stance expressions, (2) discourse organizers, and (3) referential expressions.

> […] what frequency counts in a corpora reflect more or less directly are degrees of *conventionalization* of linguistic units or structures. Conventionalization, however, is a process taking place first and foremost in social, rather than cognitive, systems, and it requires an additional logical step to assume that degrees of conventionalization more or less directly translate into degrees of entrenchment.   (Schmid 2010, 116–117)

The writer may thus search for the appropriate formulae, first describing the individual situation of the youth they are monitoring, and then adapting the particular situation to social norms, in terms of both assessment and language choices. This hypothesis can be verified by tracking the contents of revision bursts (Baaijen, Galbraith & de Glopper 2012; see also, under a different perspective, Galbraith and Baaijen this volume).

5.2   Cognitive-semantic analysis and discussion

A complete semantic analysis of bursts and repeated segments would require an entire study unto itself. Space does not allow us to present such a full semantic description of our data here, but in what follows we will highlight a few relevant semantic phenomena.

First, common RS-burst constructions related to certain specific denotational domains. Noun phrases were frequent among these common constructions, although they also included N + V and adverbial constructions. Among the noun phrases, common RS-burst constructions were often Poss.det. + Noun. Most of these referred to the foster child's family:

ses parents, sa mère, sa famille, ses sœurs, ses inquiétudes pour, de ses sœurs, chez sa mère	*his/her parents, his/her mother, his/her family, his/her sisters, his/her worries for, of his/her sisters, at his/her mother's home*

Items focusing on the child were also frequent; they were often configured as Subject Noun + Auxiliary Verb constructions, although various patterns were attested. Some patterns seem to be gender-oriented, such as the preferential use of *to be* (potentially followed by a characterization) with a female subject, and the preferential use of *to have* (potentially followed by an event-verb) with a male subject:

<u>Eloïse/Fanny/il est</u>; elle peut; se montre plutôt; <u>Kevin/Alexis a</u>; il a demandé; né/e le; Enfin, Fanny; qu'elle doit; qu'elle ne; du jeune	*Eloïse/Fanny/he is; she can; tends to be rather; Kevin/Alexis has; he asked; born on; Finally, Fanny; that she must; that she not; of the youth*

Some constructions shared by bursts and RSs refer to stereotypical realia in social work, such as the collective (a) or an institution (b); characterize the child and support the social worker's evaluation (c); or refer the educator's subjective involvement (d):

(a) groupe de, sur le groupe (x2), du groupe, de l'internat, sur l'internat, aux autres (x2), les autres, des contacts	(a) *group of/from, on the group (x2), of the group, of the boarding residence; on the boarding residence; to others (x2), others, (of) contacts*
(b) les éducateurs, au SAFE, en IME, par le SESSAD, de l'Orangerie, le placement,	(b) *educators, at the SAFE, in IME, by the SESSAD, of the Orangerie, the fostering*
(c) se réfugie dans, la question, les difficultés, les raisons, reste difficile	(c) *to take refuge in, the question, the difficulties, the reasons, remains difficult*
(d) nous observons, nous constatons, nous avons, nous avions, nous lui avons	(d) *we observe, we notice, we have, we had, we had…to him/her*

Complex proper names and other designative phrases (*de Balleroy, Me Alleaume, Mme de Souza, Mme Chaudin, Mme X, Mr Y*, etc.) were also included among common RS-burst constructions: not only did names occur in all parts of the text, but they also seem to have been written within bursts.

All these denotative types are highly entrenched in the situation and professional practice of these writers, putting the child, his/her family and situation, and fostering at the core of the writing process. It is the institution, however, that determines and dominates the denotative domain.

Moreover, some idioms were found in the corpus both as bursts and as repeated sequences – either formulaic, like most of the adverbials below, or some type of verbal lexical bundles. These are the only systemic elements that were at the same time memorized and reproduced as bursts, and produced as repeated segments:

un peu, suite à, à chaque fois, du fait, à ce sujet, lors des, à plusieurs reprises, d'autre part, de ce fait, mais aussi, De plus, En effet (x2), Par ailleurs (x2), Pour autant (x2) a été (x2), a été prononcé, il faut, se trouve	*a bit, following* (*the*), *each time,* (*because*) *of the fact, on this subject, during, several times, on the other hand, as a result/therefore, but also, moreover, indeed, furthermore, nevertheless* *has been/was, has been/was pronounced, he must/it is necessary, ~as a matter of fact*

Secondly, the types of patterns that were specific to bursts present some intriguing profiles which merit semantic and cognitive analysis. The cases of discontinuity, which were more peculiar to bursts, are in this respect semantically relevant. For instance, the discontinuity after or on both sides of a full stop, a coordinative marker or a connector indicates that concatenations were anticipated, but

their content was not pre-formulated: *et son agressivité contenue; importantes.* **Le médecin…** –> **and** *her contained aggressiveness; great. The physician…*

Many non-saturated noun bursts corresponded to lists or subtitles, and signal the existence of genre-specific pre-defined structures in the reports of social workers.

We also attested two infra-grammatical types of bursts, the plural marker "*s*" and the feminine marker "*e*" as breaking points, which suggest the occurrence of cognitive processing before these two grammatical markers. This is quite surprising, since a pause before morphological markers of gender or number suggests that the marker was not retrieved along with its noun but in a separate step. This suggests that in some cases calculating the morphological marker is effortful for writers, a behaviour that is more typical of novice writers.

6. Conclusion

Less than 3% of the units in the analysed texts were shared by bursts and repeated segments. This means that corpus data are not psycholinguistically valid here, and that the memorization and automatisation principles supposedly associated to clusters and collocations cannot objectively define the recurrent occurrences attested in the body of discourse that we examined.

However, some relevant features can be formulated here:

– The most frequent types of repeated segments of text and of bursts shared the same linguistic structures;
– The most frequent linguistic structures in bursts and RSs were syntactically unsaturated strings;
– The longest bursts were made up of complete syntactic structures such as sentences, and clauses, but these bursts were less frequent.

The "nominal" dimension of the discourse in our corpus is quite intriguing. The prevalence of saturated noun phrases and prepositional phrases (which usually contain a noun phrase) as repeated segments, and the high proportion of noun-based constructions among those that appeared both as repeated segments and as bursts, point to a topic that will be worth examining in depth in following studies.

It must be mentioned that the comparison performed here between RSs and bursts strongly depends on the definition of bursts. As mentioned above, bursts were identified in real-time data by the presence of a fluent transcription period, i.e. without any pause. The question of to define a threshold for identifying pauses is still a matter of debate in psycholinguistics (see Chenu, Pellegrino, Jisa & Fayol

2014). Our analysis of bursts is hence dependent on the two-second threshold used here. 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. Further research should examine the possible slot choices in open-ended productions, more thoroughly confronting the qualitative contents of bursts and repeated segments.

Finally, our study offers new information about linguistic entrenchment (Schmid, in press). Entrenchment is the psychological consolidation of linguistic structures which, as such, are retrieved from writers' or speakers' long term-memory in a single chunk and consequently expected to be executed fluently. We explored entrenchment by comparing process and product performances, and showed that, in our corpus, RSs and bursts share very few linguistic structures. The relationship between social conventionalization, on the one hand, and cognitive automatisation and retrieval, on the other hand, remains to be clarified. Our study shows that both play a role in both the product and the process of writing.

The prospects for generalization from our data are limited, inasmuch as we were dealing with professional writing, with domain-specific norms, instructions, and constraints (see Cislaru 2014). Nevertheless, it may be that the constructions seen here to occur both as repeated sequences and as bursts play a particular role in communication. In the case of our corpus, they may represent text segments that are conventionally shared by all the addressees of the reports.

Acknowledgments

References

Alamargot, Denis, and Lucille Chanquoy. 2001. *Through the models of writing*. Dordrecht (NL): Kluwer.

Alamargot, Denis, Christophe Dansac, David Chesnet, and Michel Fayol. 2007. "Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production." In *Writing and Cognition: Research and Applications*, ed. by Mark Torrance, Luuk Van Waes, and David Galbraith, 13–29. Amsterdam: Elsevier. DOI: 10.1108/s1572-6304(2007)0000020003

Alamargot, Denis, Patrice Tellier, and Jean-Marie Cellier (Eds.). 2007. *Writing in the work place*. Amsterdam: Brill. DOI: 10.1163/9789004253254

Alves, Rui A., Marta Branco, Sao Luis Castro, and Thierry Olive. 2011. "Children of high transcription skill compose using bigger language bursts." In *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*, ed. by Virginia W. Berninger, 389–402. New York: Psychology Press.

Alves, Rui A., Thierry Olive, and Sao Luis Castro. 2008. The Transcriber contributes to burst length in written language production. *Paper at the Writing Research Across Borders Conference. Santa-Barbara*. CA, USA. February 22–24.

Alves, Rui A., Sao Luis Castro, and Thierry Olive. 2008. "Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill." *International Journal of Psychology* 43: 969–979. DOI: 10.1080/00207590701398951

Alves, Rui A., Sao Luis Castro, Luisa Sousa, and Sven Strömqvist. 2007. "Typing skill and pause-execution cycles in written composition." In *Writing and Cognition, Research and Applications*, ed. by Mark Torrance, Luuk Van Waes, and David Galbraith, 55–65. Dordrecht: Elsevier Sciences Publishers.

Baaijen, Veerle M., David Galbraith, and Kees de Glopper. 2012. "Keystroke Analysis: Reflections on Procedures and Measures." *Written Communication* 29: 246–277.
DOI: 10.1177/0741088312451108

Biber, Douglas. 2009. "A corpus-driven approach to formulaic language in English. Multi-word patterns in speech and writing." *International Journal of Corpus Linguistics* 14 (3): 275–311.
DOI: 10.1075/ijcl.14.3.08bib

Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. "If you look at…: Lexical bundles in university teaching and textbooks." *Applied Linguistics* 25 (3): 371–405.
DOI: 10.1093/applin/25.3.371

Chanquoy, Lucile, Jean-Noel Foulin, and Michel Fayol. 1996. "Writing in adults: A real-time approach." In *Writing Research: Theories, Models and Methodology*, ed. by Gert Rijlaarsdam, Huub van den Bergh, and Michel Couzijn, 36–43. Amsterdam: Amsterdam University Press. DOI: 10.5117/9789053561973

Chenoweth, Anne N., and John R. Hayes. 2001. "Fluency in writing." *Written Communication* 18: 80–98. DOI: 10.1177/0741088301018001004

Chenu, Florence, François Pellegrino, Harriet Jisa, and Michel Fayol. 2014. "Interword and intraword pause threshold in the writing of texts by children and adolescents: a methodological approach." *Frontiers in Psychology* 5: 182. DOI: 10.3389/fpsyg.2014.00182

Cislaru, Georgeta. 2014. "Contraintes linguistiques et contextuelles dans la production écrite." *Les Carnets du Cediscor* 12: 55–74.

Cislaru, Georgeta, Frédérique Sitri, and Frédéric Pugnière-Saavedra. 2013. "Figement et configuration textuelle: les segments de discours répétés dans les rapports éducatifs." In *Across the Line of Speech and Writing Variation. Corpora and Language in Use – Proceedings 2*, ed. by Catherine Bolly, and Liesbeth Degand, 165–183. Louvain-la-Neuve: Presses universitaires de Louvain.

Détrie, Catherine, Paul Siblot, and Bertrand Verine. 2001. *Termes et concepts pour l'analyse du discours: une approche praxématique*. Paris: Honoré Champion. DOI: 10.1075/li.34.1.05lee

Erman, Britt, and Beatrice Warren. 2000. "The idiom-principle and the open-choice principle." *Text* 20: 29–62. DOI: 10.1515/text.1.2000.20.1.29

Foulin, Jean-Noel. 1995. "Pauses et débits: les indicateurs temporels de la production écrite." *L'Année Psychologique* 95: 483–504. DOI: 10.3406/psy.1995.28844

Foulin, Jean-Noel. 1998. "To what extent does pause location predict pause duration in adults and children writing." *Current Psychology of Cognition* 17: 601–620.

Hayes, John R., and Anne N. Chenoweth 2006. "Is working memory involved in the transcribing and editing of texts?" *Written Communication* 23: 135–149.
DOI: 10.1177/0741088306286283

Hoey, Michael. 2004. "Lexical priming and the properties of text." In *Corpora and Discourse*, ed. by Alan Partington, John Morley, and Louann Haarman, 385–412. Bern: Peter Lang.

Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
DOI: 10.4324/9780203327630

Hunston, Susan, and Gill Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: John Benjamins.
DOI: 10.1017/s0022226701001001

Hymes, Dell. 1971. *On Communicative Competence.* Philadelphia: University of Pennsylvania Press.

Kaufer, David S., John R. Hayes, and Linda Flower. 1986. "Composing written sentences." *Research in the Teaching of English* 20: 121–140.

Kuiper, Koenraad. 2009. *Formulaic Genres*. New York: Palgrave Macmillan.
DOI: 10.1057/9780230241657

Lafon, Pierre, and André Salem. 1983. "L'inventaire des segments répétés d'un texte." *Mots* 6: 161–177. DOI: 10.3406/mots.1983.1101

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*. Stanford: Stanford University Press. DOI: 10.1017/s0022226700000177

Lardilleux, Adrien, Serge Fleury, and Georgeta Cislaru. 2013. "Allongos: Longitudinal Alignment for the Genetic Study of Writers' Drafts." *Computational Linguistics and Intelligent Text Processing*, Springer LNCS 7817: 537–548. DOI: 10.1007/978-3-642-37256-8_44

Leijten, Mariëlle, and Luuk Van Waes. 2006. "Inputlog: New Perspectives on the Logging of On-Line Writing." In *Studies in Writing: Vol. 18. Computer Keystroke Logging and Writing: Methods and Applications*, ed. by Kirk P.H. Sullivan, and Eva Lindgren, 73–94. Oxford: Elsevier.

Mayaffre, Damon. 2007. "L'analyse de données textuelles aujourd'hui: du corpus comme une urne au corpus comme un plan. Retour sur les travaux actuels de topographie/topologie textuelle (partie I)." *Lexicometrica* 9. http://lexicometrica.univ-paris3.fr/numspeciaux/special9/mayaffre.pdf. (March 2, 2012).

Mey, Jacob L. 1998. "Adaptability." In *Concise Encyclopedia of Pragmatics*, ed. by Jacob L. Mey, 5–7. Oxford: Elsevier. DOI: 10.1017/s0022226799277742

Née, Emilie, Erin McMurray, and Serge Fleury. 2012. "Textometric Explorations of Writing Processes: A Discursive and Genetic Approach to the Study of Drafts." *Lexicometrica* JADT 2012 767–778.

Olive, Thierry. 2004. "Working memory in writing: Empirical evidences from the dual-task technique." *European Psychologist* 9: 32–42. DOI: 10.1027/1016–9040.9.1.32

Olive, Thierry. 2012. "Writing and working memory: A summary of theories and of findings." In *Handbook of Writing: A Mosaic of New Perspectives*, ed. by Elena L. Grigorenko, Elisa Mambrino, and David D. Preiss, 125–140. New York: Psychology Press.

Olive, Thierry. 2014. "Toward an Incremental and Cascading Model of Writing: A review of research on writing processes coordination." *Journal of Writing Research* 6: 173–194.

Olive, Thierry, and Ronald T. Kellogg. 2002. "Concurrent activation of high- and low-level production processes in written composition." *Memory & Cognition* 30: 594–600. DOI: 10.3758/bf03194960

Olive, Thierry, Alves Rui A., and Castro Sao Luis. 2009. "Cognitive processes in writing during pauses and execution periods." *European Journal of Cognitive Psychology* 21: 758–785. DOI: 10.1080/09541440802079850

Partington, Alan. 1998. *Patterns and Meanings*. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1017/s0272263101221068

Salem, André. 1986. "Segments répétés et analyse statistique des données textuelles." *Histoire & Mesure* 1(2): 5–28. DOI: 10.3406/hism.1986.1518

Schilperoord, Joost. 2002. "On the cognitive status of pauses in discourse production." In *Contemporary Tools and Techniques for Studying Writing*, ed. by Thierry Olive, and C. Michael Levy, 59–85. Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-010-0468-8_4

Schmid, Hans-Jörg. 2010. "Does frequency in text instantiate entrenchment in the cognitive system?" In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, ed. by Dylan Glynn, and Kerstin Fischer, 101–133. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110226423.101

Schmid, Hans-Jörg. In press. "A framework for understanding linguistic entrenchment and its psychological foundations in memory and automatization." In *Entrenchment, Memory and Automaticity. The Psychology of Linguistic Knowledge and Language Learning*, ed. by Hans-Jörg Schmid. Walter de Gruyter.

Schmitt, Norbert (Ed.). 2004. *Formulaic Sequences*. Amsterdam, Philadelphia: John Benjamins. DOI: 10.1093/applin/ami018

Schmitt, Norbert, Sarah Grandage, and Svenja Adolphs. 2004. "Are corpus-derived recurrent clusters psycholinguistically valid?" In *Formulaic Sequences: Acquisition, Processing and Use*, ed. by Norbert Schmitt, 127–152. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/lllt.9.08sch

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press. DOI: 10.1177/003368829302400207

Sinclair, John. 2004. *Trust the Text. Language, Corpus and Discourse*. London/New York: Routledge. DOI: 10.4324/9780203594070

van Hell, Janet G., Ludo Verhoeven, and Liesbeth M. van Beijsterveldt. 2008. "Pause Time Patterns in Writing Narrative and Expository Texts by Children and Adults." *Discourse Processes* 45: 406–427. DOI:10.1080/01638530802070080

Verschueren, Jeff, and Frank Brisard. 2009. "Adaptability." In *Key Notions for Pragmatics*, ed. by Jeff Verschueren and Jan-Ola Östman, 28–47. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hoph.1.02ver

Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511519772

Yamasaki, Nozomi. 2008. "Collocations and colligations associated with discourse functions of unspecific anaphoric nouns." *International Journal of Corpus Linguistcs* 13(1): 75–98. DOI: 10.1075/ijcl.13.1.05yam

Tracks and traces of the writing process

From writing under production to the finished product

A processual threshold

Irène Fenoglio

CNRS (ITEM/ENS)

The study of manuscripts allows us to analyze the initial materialization of textuality, that is, the way in which text mass/material is built up, then the way in which an actual text is constructed that can then be turned into a book. This issue raises several questions: By what process are the enunciative forms constructed? Given the diachrony of the graphic delinearization brought about by successive additions and revisions, how does the author maintain enunciative linearity?

Through the analysis of the manuscripts of Pascal Quignard, Emile Benveniste and Ferdinand de Saussure, we show that while the product of the writing act is *always* linear, the writing process is *never* linear.

Keywords: manuscript; delinearization; enunciation; Benveniste; Quignard; de Saussure

1. Introduction

The preliminary notes and successive drafts of manuscripts contain all manner of traces of hesitations, resumptions and forgettings. We can assume that these traces emanate from the psychic (i.e. simultaneously cognitive, reflective and unconscious) gesture of writing. Scientific manuscripts, such as those produced by linguists, allow us to examine all the methodological constraints they face. Moreover, we can see where they hesitate over terminological considerations, requirements, and even injunctions arising from discourse production.

All these constructs become invisible in the published and consultable text, even if they continue to make their presence felt. In the case of scientific texts, these constructs reveal the inner workings of the theorizing process.

By studying manuscripts, we can analyze the initial materialization of textuality: how the text is gradually put together, from the first words to the actual book. In this investigation, I restrict myself to the pen movements across the surface of

the page, scrutinizing the writing gesture (cf. Fenoglio 2008, 2009b; Fenoglio & Boucheron 2002). The latter, while seemingly natural, is actually very complex. Both physical and psychic at the same time, this gesture produces a trace that can be used to track textuality. I only deal here with the writing gesture that is graphically materialized. We cannot say anything about what happens before or during this gesture in the mind, "behind the eyes", as Pascal Quignard eloquently puts it. By scrutinizing the manuscripts of Quignard, Ferdinand de Saussure and Emile Benveniste (see also Fenoglio 2006, 2007a, 2009a, 2009c, 2009d, 2011, 2012a) I demonstrate how a sentence, a text and a book are put together.

This type of examination can provide definitive answers to a number of questions, and allows us to analyze certain linguistic issues (Fenoglio 2011, 2012a, b, c). By what process, for instance, are the enunciative forms constructed? Intriguingly, despite the diachrony of the graphic delinearization brought about by successive additions and revisions – as in this draft manuscript by Quignard –, the enunciative linearization is never *lost*: we can always recognize and follow its tracks.

The complexity of what happens when we write and when we attempt to produce a legible text can sometimes seem inextricable at first glance. The legible text provided for the Other and destined for the reader never willingly gives up its data.

Regardless of the identity of the author and regardless of the writing object, the delinearization process is absolutely necessary, and constitutes an integral part of writing production. By picking up all the *incidents* that occur during writing, which, in an ideal world, would be absolutely cursive and linear, we can understand the processes behind all forms of writing. This understanding does not entail finding out about the author's psychology. For instance, the repetitive use of certain words that is subjectively specific to a given author can be regarded as part of his or her style. However, since no word is perfect, all writing is subject to alterations, and all the various reformulations, breaks, repetitions, word omissions and ellipses that inwardly leave their traces, inhabit and add new layers to the linearity of a text provided to the reader.

For example, in Benveniste manuscripts, how do we move from the tangle of alterations in the draft[1] stage to the printed[2] text which is directly legible?

How do we go from Quignard's preliminary handwritten notes to this *draft*, and thence to at least 11 other versions of his *initial content*:[3]

---

1. BnF, Pap Or, boite 45, env. 117, f⁰ 3.

2. Benveniste, Émile. 1980. *Problèmes de linguistique générale*, 2. Paris: Gallimard (coll. Tel), p. 43.

3. *Sur le désir de se jeter à l'eau*, *op. cit.*, p. 45 and p. 48, respectively.

CHAPITRE III

Sémiologie de la langue *

I

*La sémiologie aura beaucoup à faire
rien que pour voir où se limite son
domaine.*

Ferdinand de Saussure [1].

Depuis que ces deux génies antithétiques, Peirce et Saussure, ont, en complète ignorance l'un de l'autre et environ le même temps [2], conçu la possibilité d'une science des signes et travaillé à l'instaurer, un grand problème a surgi, qui n'a pas encore reçu sa forme précise, n'ayant même pas été posé clairement, dans la confusion qui règne sur ce domaine : quelle est la place de la langue parmi les systèmes de signes ?

Peirce, reprenant sous la forme *semeiotic* la dénomination Σημειωτική que John Locke appliquait à une science des signes et des significations à partir de la logique conçue elle-même comme science du langage, s'est adonné toute sa vie à l'élaboration de ce concept. Une masse énorme de notes témoigne de son effort obstiné pour analyser dans le cadre sémiotique les notions logiques, mathématiques, physiques,

* *Semiotica*, La Haye, Mouton & Co., I (1969), 1, pp. 1-12 et 2, pp. 127-135.

1. Note manuscrite publiée dans les *Cahiers Ferdinand de Saussure*, 15 (1957), p. 19.
2. Charles S. Peirce (1839-1914); Ferdinand de Saussure (1857-1913).

until we eventually reach this first edition that is materially featureless and constantly legible, no longer sending our gaze off in myriad directions?

Nevertheless, even when we are confronted with such apparent multilinear illegibility, we can still make out the presence of a text. The process we are interested in consists in switching from *a* text to *the* text. In other words, from textualization in progress – a potentially infinitely expanding process – to a constructed and discrete text, even though it can be revised and modified thereafter.

2.   Constructing a text

2.1   Putting pen to paper

We first need to talk about writing materials. Nowadays, authors rarely make do with pen and paper. These have largely been supplanted by the computer, and this has had a significant impact on writing. Benveniste and Saussure wrote entirely by hand, while Quignard combines handwriting and word processing. All three punctuate(d) their preliminary writings with small drawings that were an integral part of the construction process.

Notes, be they scattered or jotted on a single sheet, are often the starting point. These notes express the writing project, the desire that underpins it and the questions it raises. They do not, however, constitute the start of the drafting process *per se*.

Thus, in Benveniste's preliminary[4] notes for his article on the utterance as a *formal device*,[5] the linguist lays the foundations for his thinking, expressing his anxieties about how and what to say:
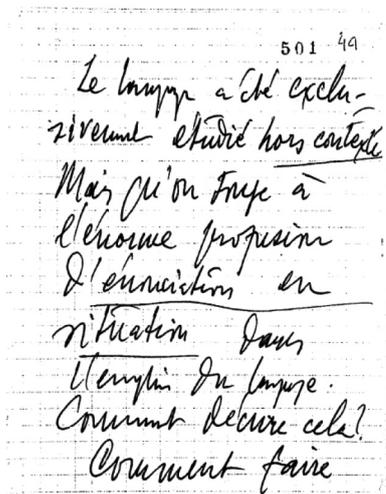
Language has been exclusively studied <u>out of context</u>
However, we need to think about the huge profusion of <u>utterances in situation</u> in language use
How can we describe that?
How can we proceed?

---

4.   BnF, Pap Or boite 51, env. 198, f⁰ 501.

5.   *Langages* 17: 1970, reprinted in *Problèmes de linguistique générale*, *op. cit.*, p. 79–88.

This additional note is more constructive:[6]

Actually it is a change, but it is not a change in the substance of language itself. A change ⟨more subtle, more radical⟩ owing to the fact that it is set in motion, that somebody seized it, and that someone takes hold of it, and moves it, puts it into action, that this device which is lying there, potential, but inert, consistent in signs on the one hand (lexical and different signs), and inflectional and syntactic models on the other hand ⟨suddenly comes to life and becomes suddenly actual⟩ suddenly takes on existence ⟨a speech is constructed restoring around itself a living movement⟩; of language. Something is born in the world then. A man expresses (lat. exprimere "to press out, to bring forth"), it brings language forth into the utterance

This important note shows just how moved the linguist-writer is by the discovery of the researcher, and how he explores his thoughts by setting them down on paper and secures their foundations ("lat. exprimere to press out, to bring forth"). This process is expressed in a rough, awkward, unconventional and astonished way. Although none of this emotion will be visible in the final text, it nonetheless contributes to the embedding of thought in textuality.

Quignard's preliminary notes for his work on Butes, a direct translation of Apollonius of Rhodes' *Argonautica*, constitute one of the first visible foundations of the text that will eventually be published:

---
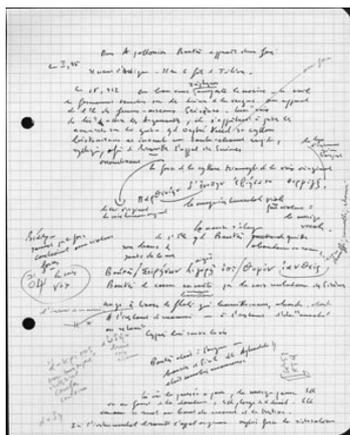
6.   BnF, Pap Or boite 51, env. 198, f⁰ 486.

Although the fictional content established in this inscription-translation does not appear in this form in the published book, this initial translation constitutes a genuine foundation on which the text will later be constructed. It has a real *mnemonic and textual depth.*

These preliminary jottings are divided up into sections, and their order is indicated on the sheets of paper, as it is in Saussure's manuscripts:[7]

Let us consider, for example, this preliminary note for a lecture that Benveniste delivered at the Warsaw Symposium:[8]

---

7.   Manuscript of "De l'essence double du langage", Geneva Library (Engler classification), f. 4 a and f. 10 a(1), respectively.

8.   BnF, Don 0429, f⁰ 39. The second [selon F. Dosse] symposium on semiotics held in Warsaw in the summer of 1968.

or these notes in Quignard:[9]
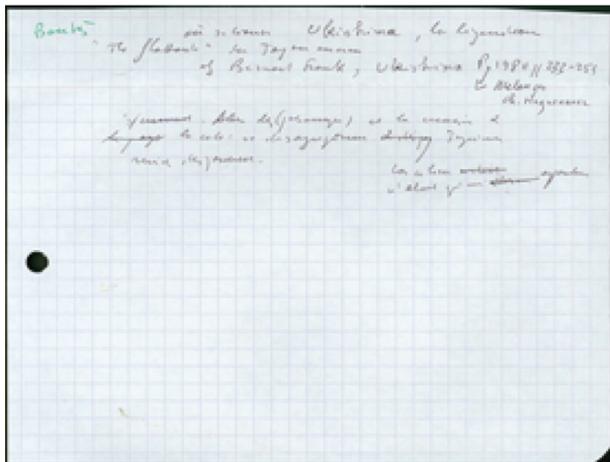
In the writing stage, the text is picked apart, revealing the dislocation/reconstruction that takes place at the different levels of text elaboration.

Here is an example from Benveniste:[10]

We **will** ~~set out~~ **have to explain** a relationship
Saussure ~~is characterized~~ appears
Saussure ~~appears~~ is presented from the start
Saussure presents ~~from the start~~
Saussure presents **from the start**

We can see that a choice had to be made between "set out" and "explain". Furthermore, this lexical choice is joined by a modal one. We therefore switch from a simple assertion to an obligation ("have to").

In addition to thinking about lexical semantics and modality, we can see that Beneveniste hesitated over whether or not to add the information "from the start": we can count three operations related to "from the start", as it was first written down, then crossed out, then written down again.

---

9.   V3, f⁰ 2 of the manuscript for *Boutès*, *Sur le désir de se jeter à l'eau*, *op. cit.*, p. 50.

10.   Draft of "Sémiologie de la langue", BnF, Pap Or, boite 45, env. 117, f⁰ 7.

These word choices in Saussure's preliminary manuscript are significant, as each change implies a total shift in point of view:
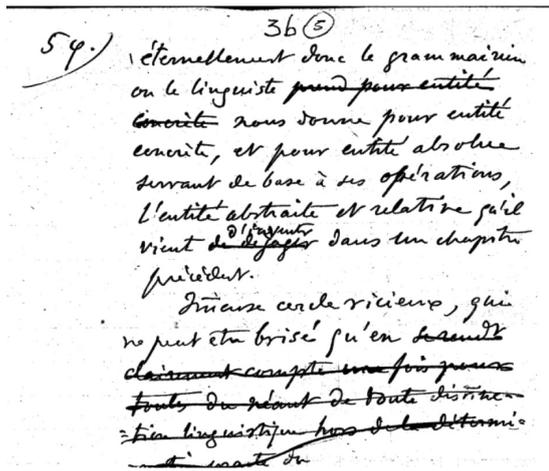
the grammarian or the linguist ~~takes as a concrete entity~~**gives us** as a concrete entity
[…]
The abstract and approximate entity he just ~~highlighted~~**invented** in a previous chapter.

Both substitutions correspond to the same change in direction. The linguist switches from a positivist position ("takes", "highlights" an entity assumed to reflect reality) to an epistemological point of view, where everything takes place in the linguist's scientific discourse: he "gives us" as an entity, the entity he "has just invented" in and through his reflexive speech.

Substitution (suppression + insertion) simultaneously expresses a hesitation and a choice: it redirects the utterance. The written trace of this movement reveals two textual possibilities (or two possible texts). The first reflects a positivist discourse, the second a distanced awareness of the linguist's function. In manuscripts, we can observe the utterance of two words at the same time, just as we can in speech, with slips of the tongue, for example. The saying that "you cannot say two words at the same time" is often transgressed in manuscripts.

In literary texts, another element is played out, as we can see in this example taken from Quignard's manuscript for the first chapter of *Sur le désir de se jeter à l'eau* (*On the Desire to Take the Plunge*);[11]

The author makes two related corrections. These parallel corrections follow the rhythm of the sentence. In literary manuscripts, substitutions generally involve stylistic modification and rhythm. The two successive substitutions clearly highlight the language constraint whereby the substituting phrase has to have the same

---

11.   Manuscript, private collection.

construction. However, at the literary level, "if we think about" clearly opens up a wider field of possibilities for the reader than "compared with".
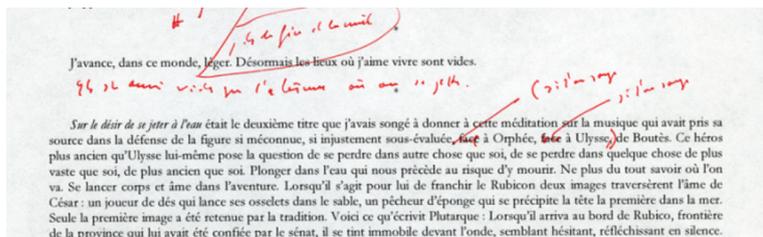
"*On the desire to take the plunge*" was the second title I had thought of giving to this meditation on music that had emanated from the defense of the figure of Butes, so little known, so unfairly underestimated, <u>compared with</u> ~~regarding~~ ⟨if we think about⟩ Orpheus, <u>compared with</u> ⟨**if we think about**⟩ Ulysses.

2.2   Writing

Writing is about the emergence of textuality through the binding together of words. There may be inescapable language constraints, but there are also endless possible combinations.

Benveniste, who was the first to deal linguistically with the sentence, thus creating the enunciative perspective, noted in preparation for his article on "The Semiology of Language"[12] that a sentence (utterance) has to be put together, or "created". On the basis of the Saussurean language system, he demonstrated the need for a "formal apparatus" for creating sentences. He called this the "formal apparatus of the enunciation", underscoring the impossibility of shifting directly from sign to sentence without having an understanding of the act of discourse. To consider the latter, he exploited the paradox between the finite number of signs that emerge from the language system and the infinite number of possible sentences:

Sign
Is it necessary to completely dissociate sign theory (= semiotics) and the theory of enunciation and sentence production, which will be a science of generation, a

---

12.   BnF, Pap Or, boite 45, env. 117, f⁰ 92.

Then in the preliminary notes for the Warsaw Symposium (1968):[13]

[…]
In our opinion, language ⟨(in the Saussurean meaning of the word)⟩ does not include the sentence – whether it is a proposition or an utterance, it conveys a message. This is because a sentence (in the broadest possible meaning) is always actualization. It presupposes ⟨refers to⟩ a situation that requires a realization of thought in language. We are in the field of speech here, and therefore out of language. The sentence exists only as realization, in and by the connected speech. This notion involves a correlative distinction in terms of meaning: instead of the universally referred notion ("*sens*" or "meaning"), we have to introduce a distinction between the "meaning" of the "sign" and the "meaning" of the speech act. The sign is in language, outside its use. It ⟨has and only has⟩ to be recognized as a sign (= like having a meaning) under the terms of a yes-no choice: krã (cran), yes; krɛ (crin) yes; krõ no. Thus krõ (cron) is only a vocal noise. This definition ~~excludes all~~ ⟨in terms of the sign's status in language⟩ exempts us from having to define ~~it~~ the "meaning" of the "sign". It just has to be identified.

A sign can be the object of several distinct identifications when it is reduced to its phonic meaning: thus krɛ "crin/crains" – craint crains crains/ crains imper/ craint craints/ point out to eight ~~signi~~ identifications. This does not matter. Polysemy does not exist on this level, which represents the assertion of existence or inexistence of the sign. I call this feature: semiotics.

This is the composition paradox highlighted by Benveniste: the sentence is not just a collection of linguistic signs but a careful piecing together of words to form a whole that is more than the sum of its parts. Quignard puts it rather differently: "What we find when writing, looking for words and crossing them out, is not what we experience when we are swamped by the sentence that we ignore." (Quignard & Lapeyre-Desmaison 2001, 111).

In addition to rewriting operations, we sometimes come across within-sentence additions indicated by a variety of insertion signs. Thus, in Saussure's manuscript:[14]
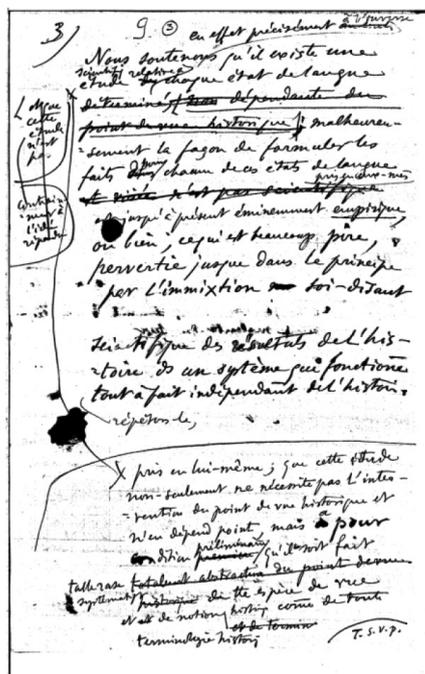
We also find devices that hold the construction process in abeyance. Saussure's manuscripts, for example, are known for their blanks and empty spaces. Whatever the reason for their presence, these blanks implicitly leave their mark: the space left for a missing word will later be filled without touching the rest of the text:[15]

---

14.   Manuscripts of "The Double Essence of Language," *op. cit.* f. 9(3).

15.   *Ibid.* f. 29 b (3).

We can observe the same process at work in Quignard. In the manuscript for the first chapter of *On the Desire*, for instance, by examining several successive rewritings of the same passage, we see how a particular expression is constructed and gradually evolves into its finished form:[16]

Version 1

My soul is contained in a small [**drawing**] that weighs two grams.

---

16.   *Op. cit.*

The more I scrutinized this manuscript, with its almost indecipherable hand-writing, the more this drawing came to resemble a beetle-an entirely subjective vision. Version Two did nothing to enlighten me, as the *beetle* simply turned into a question mark and thus into a missing word:

Version Three contains several alterations concerning the definitions of both the book and the text. Furthermore, the question mark is replaced by the name of the previously unidentified object: "My soul is contained in a small memory stick that weighs two grams":

Version Five

"I call My 'soul' that which is contained in a small memory stick that weighs two grams."

Version Six

A closer look at these operations reveals the existence of writing blocks. These blocks remain stable and are reminiscent of *cut and paste*.

Here are the successive transformations this sentence undergoes:

*My soul is contained in a small [**drawing**] that weighs two grams*
My soul is contained in a small ? that weighs two grams
My soul is contained in a small memory stick that weighs two grams
My I call 'soul'that which is contained in a small memory stick that weighs two grams
I call 'individual soul' that which is contained in a small memory stick that weighs two grams

Moreover, in a rerun of the Saussurean blanks, space is left for the appropriate term, which is not yet available. This term may change across the versions, but its place in the contextual construction does not.

## 2.3   Assembling the finished sentences

The nature of the *meta-inscription*, which highlights the target of the textual project, can vary from one manuscript to another. The author-critic may manifest interest in what the author-writer has just written, thereby drawing attention to him- or herself as a future reviewer, as in the Saussurean manuscripts,[17] where the *meta-margin* contains the word "Capital" written several times, along with drawings of a pointed finger:

Then again, the author-reviewer may ensure that the passages are read in the right order ("TSVP", Please turn the page) and marks the textual linearity with a guiding thread, as in these two consecutive preliminary pages by Saussure:[18]

17.   Op. cit., f. 6 e (7) and f. 29 h (1).

18.   *Op. cit.*, f. 9 (5) and f. 9 (6).

Similarly, Quignard uses this finished version of the *Butes* manuscript to sketch out the first version of the first chapter of *On the Desire.*[19] By following what is literally a guiding scarlet thread linking together the various blocks of text, we can glimpse its future structure:

19.   Op. cit., V32, f^0 2, p. 232.

Authors also set aside completed fragments, so that they can be assembled with other ones at a later time. We find an illustration of this process in Saussure, with "not to sacrifice":[20]

Quignard's manuscripts have a particular configuration. Many versions end in a section headed "FALLS" (CHUTES), a recurring title that separates these sections from the continuous text that precedes them.

In addition to everything that is removed, added or transferred either within a version, or from one version to another, fragments of texts may be temporarily or definitively set aside to be reused elsewhere one day – perhaps. This is precisely what happens to "FALLS".[21] It is in the fourth version of *Butes* that the title first appears, added in red ink:

---

20.   Ibid. f. 22 a (1).

21.   *Sur le désir de se jeter à l'eau*, op. cit., "Chutes" chapter, p. 263–276.

"FALLS" is initially written on the lefthand side, but from the sixth version onwards, it is centered:
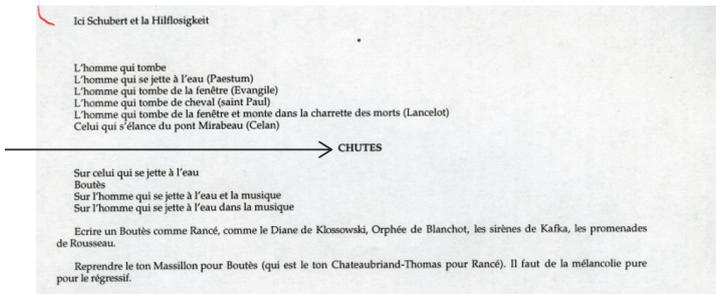
From the eighth version onwards, the author's name appears just above the title, thus marking a radical separation between the work in progress and the FALLS section, set aside for the time being. The author's name serves to confirm the text's construction:

3.   Constructing the text. Assembling the book

Once the texts are ready, they all have to be joined up. In other words, the book has to be put together.

3.1   Constructing the text

A set of meta-scriptorial or meta-discursive features attest to this transphrastic construction process. For example, Saussure writes "New paragraph":[22]

Putting the paragraphs together mostly involves cutting and pasting. Benveniste did this by hand, and by meta-textually noting all the necessary operations ("Insertion here", "Insert p. 26"):[23]

---

22.   *Op. cit.* f. 9 (3).

23.   Draft of "The Semiology of Language," BnF, cote Pap Or, boite 45, env. 117, f⁰ 28.

None of these movements, displacements and additions are, of course, visible in the final text provided to the reader. Only the manuscripts *betray* the alterations needed to achieve textual unity and cohesion.

Cutting and pasting has always existed, but the computer has obviated the need for actual glue and scissors. Now that it can be done (and undone) virtually automatically, the possibilities for putting text together, both within the sentence and within its container (the paragraph), have expanded enormously. Whole paragraphs can now be cut and pasted.

In Quignard's manuscripts, we find evidence of both tools (i.e. word processor and pen and paper), with digital cutting and pasting but also its manual version in the form of Proustian paperolles (see figure page 147).

Umberto Eco, slightly tongue-in-cheek, highlights the need to allow for future cutting and pasting by leaving out links and connectives until the last step of the second reading-writing phase:

> … with a computer […]: you write, then you decide to insert the paragraphs using cut and paste. Suddenly, you find "nevertheless" and "however", which should not be there. Then you start to delete them. But the next time you come to write, you leave these syntactic links out, so that you are free to move the paragraphs around. Thus, you eliminate "as a consequence", "but", "since", etc. In a certain sense, you therefore produce paragraphs that are ready to be shifted […] These electronic writing incidents will gradually lead us to cut adverbs and prepositions to the bare minimum, in order to maintain our freedom to move components around […] We will have to wait a century before we are able to judge whether this practice has had an impact on the way we think. (Eco 2007, 175)
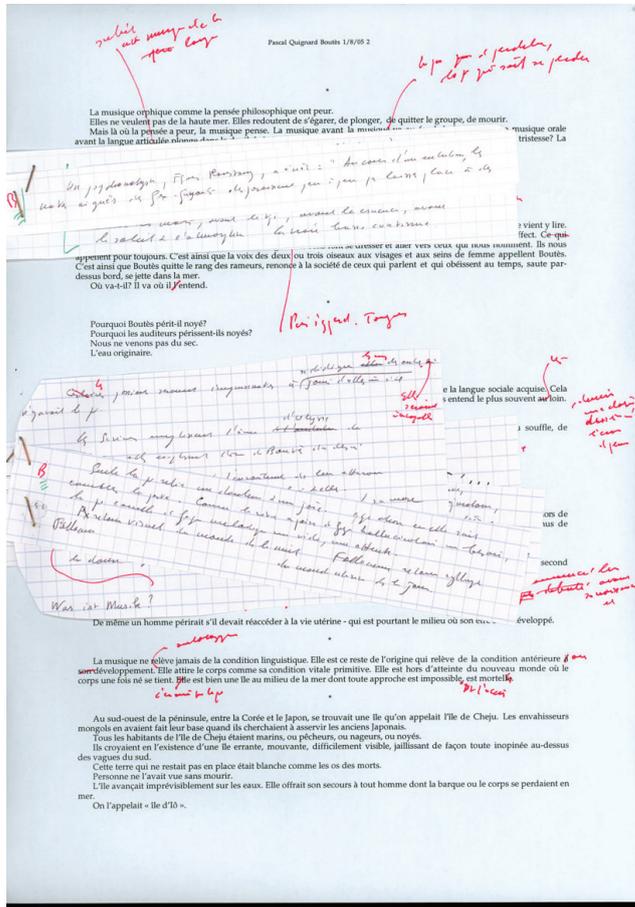
3.2   Assembling the book

In Jean-Claude Milner's *Clartés de tout*, we can read:

> In *Clartés de tout* two psychoanalysts, Fabian Fajnwaks and Juan Pablo
> Lucchelli, question the author about his research experience and the
> role that Jacques Lacan played in it […]. The book's project gradually
> emerges through these questions (and answers), which had to be carefully
> *honed* and *adjusted* in order to achieve this project. "Clearnesses of
> all" (the book's literal title) is the result of this work (my underlining).
> (Milner 2009, cover)

Quignard constructed *Butes* by dividing his text into chapters, then assembling
and numbering them. The insertion of a new chapter meant that he then had to
renumber the ensuing chapters. It therefore altered the book's structure:
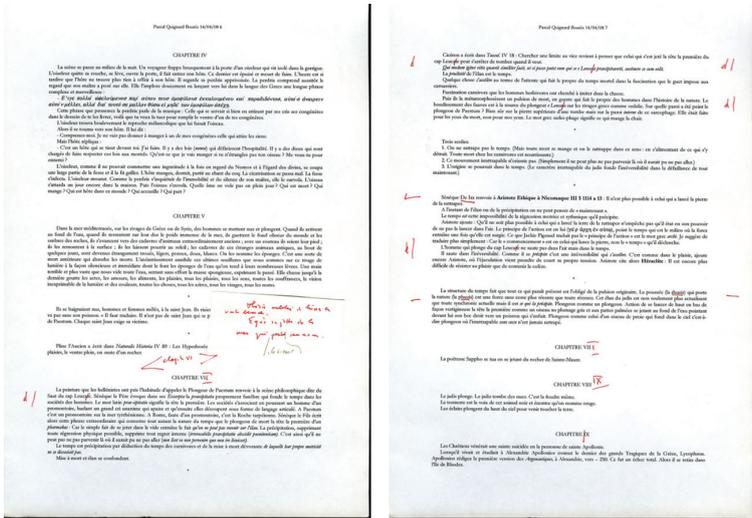
In a very eloquent article entitled "Montaigne, cannibals and grottoes",[24] Carlo Ginzburg explores the way in which Montaigne, despite seeming very close to us on account of his peculiarities, actually *escapes* us:

> Rejection of symmetry, inflation of details, violation of classical norms: Serlio[25] would have approved the loose structure as well as the uneven stylistic texture of Montaigne's essays. The abrupt juxtapositions may be compared to the alternate use of polished and rough stone in Giulio Romano's Palazzo del Te [in Mantua], representing respectively, as Serlio remarked, "works of art" and "works of nature" […]

> Readers of the first edition of the *Essais* (Bordeaux, 1580) were confronted with a text in which each essay was printed as a single, unbroken typographical unit. By splitting the sequence into two different paragraphs, modern publishers have attenuated the original harsh tone, but without making it disappear entirely.

> "Une marqueterie mal jointe", an inlay badly joined: this definition which Montaigne gave to his own writings […] reveals, in addition to his customary teasing tone, a remarkable literary self-awareness. Montaigne was referring to the uneven stylistic texture of the *Essais*, an unevenness exacerbated by his compulsive habit of inserting additions (*allongeails*) of various lengths in subsequent editions.
> (Ginzburg 2010)

---

24.   First printed in 1993, *History and Anthropology* 6(2–3): 125–155.

25.   Sebastiano Serlio, the Renaissance architect and architectural theorist (1475–1554) summoned by Francis I to oversee the construction of Fontainebleau castle.

4. Conclusion

What is textualization? It is the process whereby elements of language are mobilized, tested, and distributed after numerous hesitations, in order to achieve textual linearity. This textual linearity is sufficiently broken up into manageable chunks, and has sufficient coherence and cohesion, to be read, transmitted, quoted, and so forth.

Beginning with a finite number of letters and vowels, language and writing are both structured in the same way. Similarly, discourse is constructed from a finite number of normalized linguistic signs. At this basic level, without which no text can exist, the number of combinations and recombinations is virtually infinite. In other words, all texts are the result of a construction process that takes place on several scales and at several levels. Manuscripts bearing irrevocable and irreversible traces of a text's *gestation* represent a rich seam for understanding not so much how a text is made but how it is gradually developed, put together and consolidated. Until it is stabilized by the author's imprimatur, a text remains a work in progress. By the same token, stability is organized and indeed embodied in the finished text.

Beyond the imprimatur, the text's stability makes it reproducible, but it still needs to be inhabited by the reader, just as a house built on firm foundations is given a different meaning by each successive occupant.

A book is like a house, in that we cannot see either its vital foundations or its structure, with its complex textuality. It is this multilayered textuality that will be inhabited by the reader.

Above and beyond the text's stated ambition, it is in the semiotic functioning of the writing (even scientific writing) and its iterative, but progressive, construction processes that the author leaves his or her imprint. Writing offers both legibility and opacity, accessibility and resistance. Writing simultaneously reveals and produces a direct linearity, a kind of teleology that is immediately materialized in the content and, at the same time, an absolute, or rather inescapable, *wandering*. The four operations–addition, deletion, substitution, and displacement–manifest themselves in this process as breathing spaces, breaks, pauses and moments of apnea. They also flag up the ever-present tensions between the essential and the nonessential, freedom and constraint: there is a gap between the rigid, minimally required linearity, and the experiments, wanderings, additions and crossroads encountered in the course of the writing activity.

Although the *product* of that writing activity has a single voice and is necessarily linear, the *process* of writing is never linear.

References

Benveniste, Emile. 1966 [1974]. *Problèmes de linguistique générale*, vol. 1 and 2. Paris: Gallimard.
DOI: 10.7202/500389ar

Benveniste, Emile. 2012. *Dernières leçons, Collège de France, 1968 et 1969*, ed. by Jean-Claude Coquet, and Irène Fenoglio. Paris: EHESS-Gallimard-Seuil. DOI: 10.1080/09639489.2013.861397

Benveniste, Emile. *Archives and manuscripts*, National Library of France.

Eco, Umberto. 2007. "En quoi l'usage de l'ordinateur complexifie la genèse d'un texte?" In *L'écriture et le souci de la langue*, ed. by Irène Fenoglio, 167–191. Louvain la Neuve: Academia Bruylant.

Fenoglio, Irène. 2006. "Archives, brouillons, traces, interprétation. Pour une méthodologie de la textualisation." In *Seules les traces font rêver. Enseignement de la littérature et génétique textuelle*, ed. by Jean-Michel Pottier, 25–39. Reims: CRDP Champagne-Ardenne.

Fenoglio, Irène. 2007a. "*Fête des Chants du Marais*, un conte inédit de Pascal Quignard. Genèse in vivo et traitement de texte." *Genesis* 27: 73–95.

Fenoglio, Irène. 2007b. "Du texte avant le texte. Formes génétiques et marques énonciatives de pré-visions textualisantes." *Langue française* 155: 8–34. DOI: 10.3406/lfr.2007.6678

Fenoglio, Irène. 2008. "Observer un manuscrit. Transmettre un 'document de genèse.'" In *L'édition du manuscrit. Du manuscrit de création au scriptorium électronique*, ed. by Aurèle Crasson, 53–64. Louvain la Neuve, Academia-Bruylant.

Fenoglio, Irène. 2009a. "Les notes de travail d'Emile Benveniste. " In "Ecritures scientifiques. Carnets, notes, ébauches," ed. by Claire Doquet, special issue, *Langage & Société* 127: 23–49. DOI: 10.3917/ls.127.0023

Fenoglio, Irène. 2009b. "Textual Genetics and Manuscript in the Text Processing. A New Definition of a Text? Essay on the Avant-Texte of Pascal Quignard's Novel." In "New approaches in text linguistics," ed. by Sylvie Mellet and Dominique Longree, special issue, *Belgian Journal of Linguistics* 23: 45–61. DOI: 10.1075/bjl.23.05fen

Fenoglio, Irène. 2009c. "Conceptualisation et textualisation dans le manuscrit de l'article 'Le langage et l'expérience humaine' d'Emile Benveniste. Une contribution à la génétique de l'écriture en sciences humaines." *Modèles linguistiques* 59: 71–99. DOI: 10.4000/ml.335

Fenoglio, Irène. 2009d. "Les notes de travail d'Emile Benveniste." *Langage & Société* 127: 23–49. DOI: 10.3917/ls.127.0023

Fenoglio, Irène. 2011. "Déplier l'écriture pensante pour relire l'article publié. Les manuscrits de 'L'appareil formel de l'énonciation.'" In *Relire Benveniste. Réceptions actuelles des Problèmes de linguistique générale*, ed. by Emilie Brunet, and Rudolf Mahrer, 261–302. Louvain la Neuve: Academia (coll. Sciences du langages. Carrefours et points de vue).

Fenoglio, Irène (Ed.). 2012a. "Le geste linguistique," special issue, *Genesis* 35. DOI: 10.4000/genesis.1033

Fenoglio, Irène. 2012b. "Les manuscrits de travail des linguistes: un nouveau champ d'investigation génétique." In "Le geste linguistique," special issue, *Genesis* 35: 7–10.

Fenoglio, Irène. 2012c. "Genèse du geste linguistique: une complexité heuristique." In "Le geste linguistique," special issue, *Genesis* 35: 13–40. DOI: 10.4000/genesis.1033

Fenoglio, Irène. 2014. "Le fonds Émile Benveniste de la BnF est-il prototypique? Réflexions théoriques et méthodologiques sur les potentialités d'exploitation d'archives linguistiques." In *Archives et manuscrits de linguistes*, ed. by Valentina Chepiga, and Sofia Estanislao. Louvain la Neuve: Academia.

Fenoglio, Irène, and Sabine Boucheron. (Eds.). 2002. "Processus d'écriture et traces linguistiques," special issue, *Langages* 147.

Ginzburg, Carlo. 2010. *Le fil et les traces*. Lagrasse: Verdier.

Milner, Jean-Claude. 2009. *Clartés de tout. De Lacan à Marx, d'Aristote à Mao*. Lagrasse: Verdier. DOI: 10.3917/ess.028.0217

Quignard, Pascal, and Chantal Lapeyre-Desmaison. 2001. *Pascal Quignard le solitaire*. Paris: Flohic. DOI: 10.7202/008808ar

Editorial genesis

From comparing texts (product) to interpreting rewritings (process)

Rudolf Mahrer, Rossana De Angelis, Andrea Del Lungo,
Almuth Grésillon, Jean-Louis Lebrave,
Valentine Nicollier Saraillon, Thierry Poibeau,
Frédérique Mélanie-Becquet & Bénédicte Vauthier
"Manuscript – Linguistics – Cognition" team
ITEM (ENS/CNRS)
LATTICE (CNRS/ENS/Université Sorbonne nouvelle Paris 3)

> Nothing is ever definitively said while we still
> have time ahead of us, held out to the speaker
> as a promise.
>
> (*Rien n'est jamais définitivement dit tant que le
> temps est devant soi, donné comme espérance
> au locuteur.* Peytard 1993, §121)

In literary genetics, "editorial genetics" deals with the "public life" of texts, whereas the writing process is affected by edition and diffusion. Editorial genetics frequently has to deal with cases of "editorial rewriting": in the literary domain for example, authors frequently modify previously published works, so that several versions may co-exist. We are especially interested in Balzac's *La Bourse* (translated in English as *The Purse*) since we know three authorized versions of this specific work.

By comparing different texts associated with a single work, the literary geneticist is facing different products that are themselves the result of a writing process. However, different specificities should be outlined: (1) the writing process does not leave any trace: we just have access to different products/texts and (2) since the texts we compare seem to be achieved, differences must be referred, not to programmatic or temporary linguistic structures, but to the reconfiguration of a pre-existing textuality.

Do such products still reflect the processes that have given birth to them? Does the comparison between two texts considered as variations of a same text give access to this transformation's processes? After describing the objects of this particular textual comparison and the terminology that permits to give an account of such phenomenon, this contribution suggests to express these

questions differently, as a matter of *poetics of transitions between texts*, or, further digging, an *hermeneutics of the transition between texts*.

Keywords: editorial genetics; textuality; variation

1.   Genetic criticism and editorial genesis

Genetic criticism emerged in France at the end of the 1960s. Rather than describing written signs and their use, it focuses on their production. In other words, instead of looking at writing as a product, it treats it as a process. This theoretical approach consists in gathering together documents and observations, in order to identify the author's writing habits, linguistic reformulations, and cognitive mechanisms. This offers us an insight into the creative process and allows us to accurately characterize the preparatory documents that constitute the so-called *avant-texte* (also known as the pre- or fore-text).

As text composition is a process that takes place over time, it can be divided into several different *phases*. These are defined not only by the technological devices that are used, but also by the writing purpose and the properties of the successive preparatory activities.[1] According to Lebrave (2009)'s schematization, for example (see Figure 1), the initial stage of *traceless gestation* is followed by the first written drafts. Then comes the so-called *accommodations* phase paving the way for publication. The fourth phase consists of *revisions* to the published text. The fifth and final phase (*alterations*), concerns unauthorized modifications, in particular those made after the author's death. In this phase, which lies beyond the scope of *genetic criticism*, the text's story continues, with new writers and therefore new geneses.
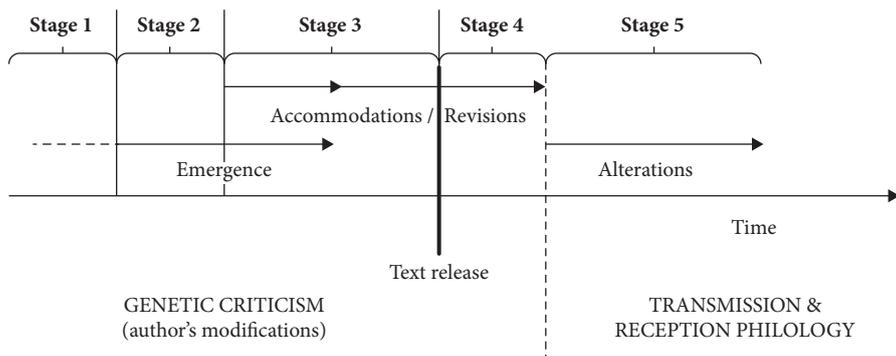
**Figure 1.** Lebrave's schematization (2009, 18)

---

1.   According to our chosen theoretical perspective, writers do not just prepare a written sign (product) open to interpretation. They prepare a trace, but also – *in fine*, a reading (process).

On first examination, *editorial genesis* deals with the fourth stage in Lebrave's schema, wherein "textual changes [are] made by the author after the text's publication" (Lebrave 2009, 18). More specifically, it concerns the point when a written text emerges from the writer's private sphere and embarks on the route to publication. In the draft phase, the writer (or writing authority) is – in principle – the sole agent and guarantor of all aspects of text production, but in this new phase, these processes are ordinarily carried out by a plethora of actors, including typists, correctors, proofreaders, printsetters, master printers and publishers.

2. Rethinking the frontier between product and process, text and *avant-texte*

By viewing writing as a process rather than a product, genetic criticism has encouraged research on preparatory documents (in particular manuscripts) bearing visible signs of text generation. As we have seen, geneticists use the term *avant-texte*, or even *dossier génétique* (Grésillon 1994, 109), to refer to the set of (more or less) chronologically ordered preparatory documents for a given text. In this opposition between *avant-texte* and text, the former is associated with the process and the latter with the product, but how does one move from the one to the other?

2.1   *Avant-texte* and text

Characterized by its original approach to literary work, textual genetics traditionally views the pass for press as a cut-off point. When the writer signs this pass, he or she agrees to the publication of the text as it stands. This contract marks the frontier between the third and fourth stages in Lebrave's schema (see Figure 1).

> Thus, the preparatory stage, where everything remains possible, gradually moves into a *new dimension* where the author's intervention is (with a few *exceptions*) far more limited. (De Biasi 2005 [2000], 45, our italics.)

> If we compare the physical, finished text with the virtual and unfinished manuscripts from a genetic point of view, we can say that *its emergence coincides with its publication* – an operation that turns a private, autographic object into a public, allographic text. (De Biasi 2000, 29, our italics.)

Such a representation of the genetic domain assumes that the reassessment of a published text is very much an exception, and further questioning of the editorial process beyond this cut-off point is discouraged. Nevertheless, authors do intervene after publication as attested by tree different and nonexlusive manifestations (1) variously justified requests by the author for *corrections*, (2) documents found in the author's personal papers containing superficial or in-depth revisions, (3) variations between different editions.

Research undertaken by the Manuscript – Linguistics – Cognition team, which has been studying post-publication rewriting since 2012, has consistently confirmed its prevalence. It concerns all periods and all genres. This observation has a corrosive effect on the genetic field, as both Grésillon (2007; Montaigne's *Essais* (*Essays*), Eluard's *Donner à voir*, Aragon's *Communistes* and Ronsard's complete works), and Vachon (2009; Balzac's work) noted:

> But the *terminus ad quem* [of the *avant-texte*] is equally complex. The criterion that was initially selected for the pass for press would rapidly prove illusory. Two aspects of textual genetics show just how fragile it was: successive editions (reviewed, corrected and augmented by the author), and the production process of theatrical works. In the former, the pass for press would be followed by publication, but the author would frequently annotate his/her personal copy of the work, introducing changes and additions to be incorporated in the next edition. (Grésillon 2007, 32)

> In other words, the Balzacian corpus leads us to formalize a specific theory for post-publication genetics that challenges the opposition between text (the focus of literary criticism) and *avant-texte* (focus of genetic criticism) and even makes it irrelevant, by showing that these are just two sides-one private, the other public-of the same reality. (Vachon 2009, 41)

We could also mention authors such as Erasmus, La Bruyère, Molière, Flaubert, Claudel, Ramuz, Cendrars, Césaire, Duras and Genet, who were all rewriters to some degree. From the *editorial genetics* point of view, we need to reconsider at least two aspects of the connection between text and *avant-texte*, that is, between written documents deemed to contain evidence of processing and documents regarded as products. First of all, we have to agree that, in theory, nothing can bring the process to an end once and for all. Nothing can sever the relationship between a creator and his or her work–apart from the disappearance of one or other. This is the meaning behind Peytard's opening quote which sums up the condition of human creation in the field of linguistic works.

Second, we need to distinguish between two categories of documents within the genesis of a work. Some writings are produced in preparation for the upcoming production of an oral or written discourse. They contain discursive forms of the preparation (preparatory notes, plans, scenarios, drafts, etc.), and use writing tools (manuscripts, pencils, quill pen, typewriters, screen and keyboard, etc.) and techniques (crossing out, erasing, copying, pasting, etc.) that connect the writing space and writing gesture with the content (invention and organization) and verbal expression.[2] Other genetic documents contain the completed form of a given

---

2.   Mahrer and Nicollier (2015) refer to this document family as *écrits de la préparation du discours*, or simply *écrits de préparation*, and suggest treating it as a type of discourse that encompasses several genres.

discourse genre, and can only be regarded *a posteriori* as part of the *avant-texte*, when we observe that they are not, after all, the final link in a chronological chain of steps ending with the final textual product.

Editorial genetics challenges critics who study writing processes via documents principally illustrating the second case. There is often no way of telling whether or not a text that has been published (once or even several times) will subsequently be revised by its author. The latter may not even have planned to do so him- or herself. Balzac, for example, writing in 1834 about the fourth edition of *La Peau de chagrin* (*The Magic Skin*) (published by Werdet in 1835), stated that "its purpose was to give once and for all a final form to the texts belonging the large general edition of [his] works". However, regarding the fifth edition (by Delloye and Lecou in 1838), he commented that "the text […] has been revised with such care that it must be regarded as the only one that exists, such is the difference between it and previous editions",[3] although it is worth mentioning that Balzac later went on to produce two more versions of that same novel. From the writer's perspective at least, these editorial genetic *steps* are viewed as definitive *stages*:

> A final state can very easily revert to a draft, and what seemed to imply the end of a creative process can evolve into a new writing episode. The text returns to being an *avant-texte*. The pen replaces the lead type.   (Grésillon 2007, 32)

2.2   Publication seen in a genetic light

Every writing can be rewritten. This observation leads us to reject the idea that, in principle, publication marks the end of the writing process. This does not mean that the geneticist should regard the published text as just another draft. In accordance with the objective of genetics as a discipline, the study of editorial genesis can be defined as the description of the effects of publishing on the various writing processes and, at the same time, the study of these processes under specific circumstances. As this chapter is too short for us to discuss it in detail, we simply outline its key aspects:

1. At a semiotic level, in the publishing phase, the piece of writing is modified in that its medium and even its semiotics change. The bundles of sheets or handwritten notebooks may become a poster, a leaflet, or a book. This aspect extends from spatial constraints and techniques to the rewriting gesture.
2. At a pragmatic level, the piece of writing is modified after having been put to one side for a time owing to publishing constraints (commercial and/or

---

3. *Lettres à l'étrangère*, Paris, Calman Lévy, p. 195–196 & 454–455 (quoted by Falconer 1969, 72).

technological).[4] The persistence of the graphic trace, combined with the work's prestige, allow for its continuing genesis. The longer this period lasts, the more "the I who was previously writing changes", as Grésillon (2007, 32) wrote, thinking not only about Montaigne writing his "*allongeails*" (expansions), but also about himself. The historical context and writing fashions (genre, style, etc.) change as well, giving the writer further reasons for revising his or her text.[5]

3.   Writing is modified when it becomes *public*. This adjective has to be understood in two different ways.

   a.   On the addresser's side, the enunciatory entity becomes openly collective, as the production of the artefact requires a range of skills (the writer is joined by the copyist, publisher, director of the collection, editor, printer, bookseller, etc.). The picture of the writer preparing his manuscript alone at his desk no longer matches reality. Henceforth, the publication of a text does not simply involve its author, be it on a practical, economic, legal, or even moral level. Genetics should be not only about the history of texts, but also about the history of the work itself. It cannot ignore the collective dimension of the publication, but must describe the product of *editorial enunciation* (Souchier 2007).

   b.   The presence of an audience means that there is also an addressee's perspective: once published, a text can prompt questions and elicit a response, be it positive or negative. Rewriting can then be interpreted as a sort of ratification or resistance to the criticisms formulated by the entities to which the work was submitted.

4.   Finally, in *literacy*, publication is an important sociocultural act. The author and publisher commit themselves to the release and, more particularly, to the published book itself. Thus, the symbolical value that the author places on publishing (his or her *imagining* of the publication act) will condition his or her post-publication rewriting (from Erasmus, who judged it as necessary and boundless, to Hugo who regarded it as vain and refused to do it).

---

4.   The contract for the second edition of *La Peau de chagrin* (*The Magin skin*) was signed in August 1831, two weeks after the first edition had sold out. The newer version came out the same year. Balzac made his revisions to the Furne edition sometime between the publication of the 15th volume of the *La Comadie humaine* (*The Human Comedy*) in 1846 and his death in 1850.

5.   Regarding the rewritings and reprints of *La Peau de chagrin* spanning the period from 1831 to 1846, Falconer makes the hypothesis that "a whole side of a 'flamboyant romanticism' from the primary text" that was probably pleasing to the 1831 readership, was gradually abandoned as readers' tastes changed. (Falconer 1969, 73, who mentions the disappearance of "play on words or on typography, exotic names and adjectives, jests, and paradoxes").

In short, from the perspective of research on the writing process, publication raises the status of a text to that of a *written document exerting constraints on its virtual rewritings*. These rewritings have to be interpreted in the light of these constraints.

2.3   What should be the *object* of linguistic analysis?

In order to describe the various writing processes, genetic critics can scrutinize the creation of discourse both on and off line (e.g. using Eye and Pen software). They can also study production by examining the traces left by certain writing technologies (e.g. pen and paper, or keyboard and hard drive) or even accounts of those practical experiences. Evidence of the various phases of editorial genesis (book corrected by its author, poster, correspondence between author and publisher, etc.) includes the different editions available in libraries. Although the latter do not contain autographical traces of rewritings, by comparing the texts, geneticists interested in this aspect can see that there are several versions of the same work.

Comparing these versions brings all kinds of differences to light. In the following section, we show how data processing can facilitate these comparisons, thereby promoting the study of hitherto neglected features. First, however, we need to discuss the terms used to think about and describe these differences. How can they be articulated with the notion of rewriting used in genetics to talk about writing processes?

Philology offers us the notion of *variant*, which is often used in the genetic context, but its philological origin can lead to misinterpretation.[6] The notion of *variation* seems more relevant, even though it is not used in the genetic field. As in music, where it is used in relation to the notion of *themes,* variation presupposes the existence of an identity that serves as its basis. Whereas variants are mutually exclusive, the term variation implies the coexistence of two objects, regarded as different from one angle, but identical from another (Ferrer 2011).

In the language context, the linguistic analysis of variation involves the comparison of two sequences (A and B), extracted from two versions of what is considered being the same text. We can say that there are *linguistic variations* between Version A and Version B if the comparison between the two enables us to pinpoint both identical sequences and, in *n* point(s), different sequences. *Variation* is thus defined as the relationship between Sequence x_i from Version A and Sequence y_i from Version B.

---

6.   The philological variant presupposes that there is only one authorized and pertinent text, and that this text is an original one. By contrast, for genetic criticism, the finished text does not yet exist and is an end in itself.

Compact basis (b_i) refers to the left and right co-texts in Versions A and B that flank the x_i and y_i sequences.

2.4 From variation to *rewriting*

For geneticists, concerned with the analysis of writing processes, *rewriting* is a written reformulation by which the previous state (x) of a linguistic sequence is invalidated by its replacement with a new one (y). From this point of view, rewriting can be regarded as a substitution process whose elements are chronologically oriented (Lebrave 1983). We can access it in at least three different ways: by observing the gestures involved in the operation, either on or off line (Leblay & Caporossi 2014; see also Leblay and Caporossi this volume); by observing the graphic traces of these gestures (e.g. crossings-out); and by studying the products of the operation. The latter implicitly brings together rewriting and variation. For a variation to be used as a basis for genetic investigation, it needs to correspond to a writing gesture. However, not all variations between texts are the outcome of rewriting. Hence, for a text comparison to take place in a genetic investigation (formulating hypotheses about the writing processes), Versions A and B need to be (1) regarded as the same text, (2) attributable to the same production entity, and (3) chronologically oriented.[7] Ideally, (4) Version B should also result directly from Version A, with no other documents between the two (the precision of genetic hypotheses depends on this constraint). Under these conditions, Versions A and B can be viewed as the input and output of a machine whose internal workings are concealed, rather like black boxes. The geneticist attempts to model the writing processes by comparing Sequences x_i and y_i. This modelling naturally depends first and foremost on the metalanguage used to interpret x_i and y_i, as well as on the *laws* governing their transformation in terms of form and content.

3. Analyzing Balzac's work in the light of editorial genetics

Balzac's work provides an exemplary case study for editorial genetics. Unlike other writers of his time, who mainly revised and rewrote their work while it was still in manuscript form (Flaubert is a prime example), Balzac regularly made major modifications between successive editions of a same text (up to eight different versions of *La Peau de chagrin*, something that is extremely rare in literary practice.

---

7. Although the relation of variation is reciprocal (if A is a variation of B, B is a variation of A) the relation of rewriting is not (if B is a rewriting of A, A cannot be a rewriting of B).

Let us summarize the different phases of the genetic process, as critics have already described it.[8] The *dossier génétique* relating to Balzac's works (collected at the end of the 19th century by Viscount Spoelberch de Lovenjoul) contains only a few preliminary documents (notes, drafts, plans, scenarios), but a welter of manuscripts. Their analysis shows that after some hesitations (as revealed by many stalled beginnings), writing seemed to progress quickly, as evidenced by the small number of deletions or insertions (to the extent that one could imagine that these texts are just copies of previous versions that have disappeared). The first rewriting phase often occurred during the correction of the galleys and proofs, in some cases reflecting a desire to expand the text with large, and even spectacular, insertions. The second phase took place after publication, in a highly systematic way (for instance, Balzac started correcting the Furne edition of the *La Comédie humaine* (*The Human Comedy*), first published in 1842, in his personal copy).

Stressing the specificity of this practice, Stéphane Vachon notes that "Balzac's writing process include[ed] the continuous re-publication of his works and recreat[ed] his manuscripts by multiplying the working editions" (Vachon 1997, 72). As we said in the Introduction, we do not consider Balzac's rewriting habits to cancel the difference between the preparatory documents and the published editions, but this difference no longer corresponds to a strict divide between process and product. Moreover, it should be noted that the rewriting process operates differently when it takes place directly in the proofs as opposed to between the different editions of a text: the former often involves an extension of the text, whereas the latter is a kind of perpetual re-adaptation, maintaining a certain compactness of text, especially for the Furne edition of *La Comédie humaine*. With the notable exception of *La Peau de chagrin*, which was augmented with the addition of several narrative episodes, the new version of each text is nearly always slightly shorter than the previous one that served as the basis for the rewriting process.

This is the case of *La Bourse* (*The Purse*), a short story that summarizes the main features of the editorial genetic process in Balzac's work. The original version of this story was published in the second edition of *Scènes de la vie privée* (*Scenes of Private Life*) by Mame et Delaunay-Vallée in 1832. The second version was published in 1835 by Madame Béchet, in Volume IX of *Études des mœurs au XIXème siècle* (*Studies of Manners in the 19th Century*), then with les *Scènes de la vie parisienne* (*Scenes of Parisian Life*). The third version corrected by the author (the 1839 Charpentier edition of les *Scènes de la vie parisienne* contained the same text as the previous one) is the one that was published in the Furne edition of *La*

---

8.   See Stéphane Vachon's article entitled "Les enseignements des manuscrits d'Honoré de Balzac. De la variation contre la variante", *Genesis* 11, 1997.

Comédie humaine (Volume I, 1842), so the story returned to its original place in *Scènes de la vie privée*. The fourth version incorporated the handwritten corrections that Balzac had inserted in his personal copy of the Furne edition for a new edition that was only published after his death.

4. The automatic analysis of literary variants

Despite having different goals, researchers belonging to the literary genetics and philological communities all have to deal with different versions of a given text that can vary immensely and thus generally contain a great many variants. The manual analysis of these variants requires a huge amount of work first to list and classify them, then to analyze their changes over time. It is a highly repetitive and tedious task, which leads to a large number of errors when the work is done manually. Computers are thus extremely useful for automatically listing and classifying these variants.

Literary genetics takes its name, of course, from biological genetics. The parallel goes further, as a text can be seen as a sequence of words or even characters, just as a strand of DNA is a chain of four different nucleotides represented by the letters A, T, G and C). Two DNA sequences can be compared on the basis of four formal operations: deletion, insertion, substitution (aka mutation), and frame shift (Lewis 2005). In the same way, two texts can be compared using these four operations: the author may have added some words and removed others, and a portion of text may have been replaced or moved. It should, however, be noted that from a strictly formal point of view, *deletion* and *insertion* can account for every type of change, as a substitution or a move can be formalized as a deletion followed by an insertion. However, the notion of *move* is more informative, as it expresses the fact that the same piece of text has been removed from one place in order to be added in another place. Similarly, *substitution* indicates that one sequence of letters (or words) has been replaced by another at a precise point in the text. Different parameters have to be defined for these operations, such as the minimum length of what can be classified as a move (e.g. changes in punctuation are generally regarded not as moves but as deletions or insertions of commas, full stops, etc.).

A number of tools have been developed to track changes in texts and classify these changes according to the four operations described above. These tools can generally be configured so as to tailor the parameters to the context or the author. One of the best known piece of software is Edite/Medite, developed under the supervision of Ganascia and Lebrave at ITEM (Bourdaillet et al. 2009).

Edite/Medite requires the documents to be formatted in XML. All the changes between two versions of a text can then be automatically calculated and classified in one of the above four categories. A graphical interface has independently been

Figure 2. Screenshot of the graphical interface developed by Yannick Saraillon to complement Edite/Medite software

developed to enable users to navigate within the text, view two versions of the text side by side, and gain access to the complete list of changes corresponding to the four operations, among other things. Figure 2 provides a screenshot of this interface.

Edite/Medite is clearly extremely useful for manipulating different versions of a given text. However, researchers working on such texts, be they geneticists, philologists or linguists, rapidly feel the need for further functionalities. The four operations are purely formal and do not take into account the linguistic content of the sections of text under consideration. A linguistic analysis of these fragments would be highly useful, making it possible to access the changes from a different, more content-oriented, point of view (see Brunner and Pordeus Ribeiro this volume).

The following section contains a proposal to go beyond the current state of the art. Our goal is to define a method for automatically analyzing the variations observed in literary texts. We begin by defining a linguistic model (What kinds of facts do we want to observe? What classification would be useful in our context?). We then examine the extent to which this model can be implemented, taking Balzac's short story *The Purse* as an example.

4.1   The linguistic classification of literary variants

The links between linguistics and literary texts are complex. Whereas linguistics mainly involves the meticulous analysis of a finite set of sentences to test linguistic features on the basis of tiny, systematic variations, establishing *minimal pairs* to check whether these variations are linguistically driven, literary texts come to us *as they are* (Fuchs 1983; Culioli 1983). For example, Culioli reminds us that a literary text is not a representative sample of a linguistic phenomenon, and cannot be extended or directly manipulated, contrary to what linguists frequently do. A text

is the result of a complex creative process, but it does not afford us access to the operations that gave rise to the actual observable result.

From this point of view, textual genetics gives us a unique opportunity to gain access to variations, as the different versions of a given text contain traces of the changes that the author (or editor) have made to the text, thereby providing clues to the creation process.

4.1.1 *What kind of linguistics?*

There are obviously many ways of doing linguistics, so some rather naive but nevertheless useful questions are: Which type of linguistic analysis is most useful for the analysis of literary texts? Which linguistic theory is most appropriate? And first of all, what should we expect from a linguistic analysis in this context? To what extent is linguistic different from stylistics? In other words, should we regard stylistics as the branch of linguistic analysis that is best suited to literary texts?

Stylistics is generally assumed to "establish principles capable of explaining particular choices made by individuals and social groups in their use of language" (Wikipedia). From this point of view, stylistics differs from mainstream linguistics. The first task is to observe and describe the differences between two versions of a given text. At this stage, the reasons why the author's choices vary from one version to the other, the interpretation of the text, and the aesthetic dimension of text variation are put to one side.

Our point of departure is thus quite simple. We want to stay as close to the text as possible, which means that the analysis must describe the modifications in the text from a neutral point of view, untrammeled by theory. We stay away from the interpretative dimension (status of the author, stylistic value of any change, etc.) as far as possible. The interpretative analysis takes place in the second stage, based on the descriptive approach proposed here. The advantage of this is that the stylistic analysis is based on comprehensive observations, and not on isolated examples, as is all too often the case.

We therefore need to develop a general descriptive framework for the linguistic analysis that corresponds to our specific area of interest (literary variants from a genetic analysis perspective) as well as to the more general domain of literary text analysis. We propose dividing this analysis into four different *levels* (words, morphology and syntax, sentences and whole text).

1. Lexicon (richness and diversity; contexts of use)
2. Grammatical morphemes, more specifically:
   a. Determiners
   b. Tense and mood of the verbs
   c. Modals

3.   Sentences (length and complexity)
4.   Text (structure; organization and use of discourse markers)

These levels are, of course, quite generic and rather conventional. However, they ensure that the analysis is comprehensive and does not only take one aspect of the text into account, which is what often happens.

The next step consists in identifying existing tools and/or developing new ones to automate the analysis.

4.2   Automating the analysis

Lexical analysis requires a system that is capable of compiling lists of words and comparing them. This analysis can be run on either the word forms themselves or a lemmatized version of the text (in order to differentiate between types and tokens). More sophisticated analyses can be conducted using advanced tools like the Sketch Engine, which provides a detailed analysis of the context in which words appear in a text (Kilgarriff et al. 2004).

4.2.1   *Lexical analysis*
We propose to calculate the specificity of each word form, and rank words according to their positive or negative specificity. *Specificity* corresponds to the traditional definition put forward by Lafon (1980).

4.2.2   *Morphological and syntactic analysis*
Morphosyntactic analysis requires a system that can analyze a text and add morphological features to word forms. The quality of analyzers is generally satisfactory, but performances on literary texts can vary dramatically, depending on the nature of the text. Manual correction of the analyzer's output is required to achieve a near-perfect analysis.

Syntactic analysis requires a syntactic parser, but parsing is far from perfect, and manual correction can be overwhelming. However, in the case of variants between two versions of a text, sometimes only a local syntactic analysis is required, rather than the analysis of whole sentences.

4.2.3   *Implementation*
So far, our strategy has consisted in using existing tools as far as possible, rather than developing new ones. Of course, when nothing is available to perform the analysis, the development of new tools becomes necessary. The preliminary experiments described in this chapter all involved existing, off-the-shelf tools.

The lexical analysis was performed by the TXM toolbox (developed as part of the Textometrie project: http://textometrie.ens-lyon.fr/) (Heiden 2010). This tool

allows users to examine the vocabulary of a given text and to compare the word lists of different texts, as and when required (see previous section).

As for the morphological analysis, we used TreeTagger (Schmid 1994), which is also integrated into the TXM platform to compare lemmatized vocabulary lists. TreeTagger can further be used to compare the morphosyntactic features of two versions of the same text. For example, we can check how verb tenses, modals or determiners are used, as these are known to be frequent sources of change between different versions. Other features can be studied separately, as long as they form part of the morphosyntactic analysis that is automatically performed. It should be noted that if TreeTagger was used in this study, this was mainly on account of its ease of use, as an integral part of TXM. We have since developed our own analyzer, which perform better than TreeTagger for French, and we will be using this new analyzer in subsequent studies.

Concerning parsing, we did not use an actual syntactic analyzer, as we decided that the sequences of tags produced by TreeTagger represented the first step towards syntax. This proved to be sufficient in this context, but a real parser might be useful in other studies. However, parsers are far from perfect, and this should be set against the cost of manual error correction.

4.3   Experiment

In this section, we describe an experiment featuring a short story by Balzac.

4.3.1   *The corpus*

Instead of studying the genesis of a text from the early drafts to first published version, we chose to focus on the notion of *editorial genesis* which, as we have seen, refers to the study of the different published versions of a text. We felt that this was more relevant in our case, as we were not so much interested in the creative process itself as in the practicalities of our analytical method. For example, Balzac is known for having published several versions of most of his texts (there are at least eight extant versions of *La Peau de chagrin*, some featuring major changes, others just a handful of minor corrections).

We analyzed the first two versions of the short story *La Bourse* (*The Purse*), published in 1832 and 1835, although there are also two later versions, published in 1839 and 1842. A truly comprehensive analysis would, of course, take all four versions of the text into account.

With this type of text, geneticists and/or linguists look for different and even contrasting things. What major changes did the text undergo? Are there any regularities/patterns in the changes Balzac made to the text? By the same token, are there any isolated, remarkable changes? This, in our opinion, is what makes the

analysis of literary texts especially difficult: analysts want to have access both to the generalities and to the specificities of the text. Analysis tools should allow both.

4.3.2 *Semi-automatic analysis*

TXM automatically generates vocabulary lists, and offers different ways of comparing these lists, based on frequency and specificity (see Figure 3).

A quick look at the results leads to some interesting observations:

1. Systematic changes, which are easy to spot even without a tool of this sort. For example, the name of the main protagonist was *Jules* in 1832 and *Hippolyte* in 1835;
2. Less systematic changes, which are probably meaningful but difficult to spot when they are distributed across the text. For example, Balzac frequently changed the adverbs he used (*chèrement* and *alternativement* are replaced with *particulièrement* and *complaisamment*). This kind of observation is especially hard to interpret, and requires careful study of the text to see whether it is meaningful or not;
3. Nonsystematic changes owing to orthographic variations. The spelling of some French words was still quite fluid at the beginning of the 19th century, as we can see here, where *mouvemens* becomes *mouvements* and *vêtemens* becomes *vêtements*, while *savant* becomes *savans* and *sergent* becomes *sergens*;
4. Some isolated words or semantic families also undergo changes from one edition to the other (e.g. between the 1832 and 1835 editions, *créancier* disappears and *rançon* appears). These changes can be spotted automatically, but their interpretation requires expert analysis.

1832 version	1835 version
Jules	Hippolyte
mouvemens	mouvements
vêtemens	vêtements
savant	savans
sergent	sergens
…	…
créanciers	rançon
…	…
chèrement	particulièrement
alternativement	complaisamment
…	…

Figure 3. Word-form lists

These changes, ranked by specificity (Lafon 1980; Lebart et al. 1997) as well as by frequency, are highly informative. When specificity is taken into account, it is

not the absolute number of changes that counts, but this number in relation to the frequency of the word (i.e. a relatively small number of changes can be meaningful if they concern an infrequent lexical form). In TXM, punctuation marks can be regarded as lexical items. The comparison of the two versions of Balzac's text ranked by specificity yields two very interesting results (see Figure 4).

Token	Freq	1832	1835	Spec.
Jules	80	80	0	−23.7
…	93	87	6	−18.7
!	101	67	34	−3.0
:	233	141	92	−2.8
A	7	7	0	−2.1
Alors	44	30	14	−1.9
.	953	515	438	−1.8
et	634	340	294	−1.2
car	27	18	9	−1.2
y	66	39	27	−1.0
…	…	…	…	…
Hippolyte	51	0	51	15.7

Figure 4. List of the most specific changes, from one version to the other

1. Balzac made numerous and meaningful changes to the punctuation. Strong punctuation marks (especially exclamation marks, suspension marks, semi-colons and, to a lesser extent, full stops) are less common in the 1835 version of the short story, meaning that this version contains fewer sentences. As a direct consequence, sentences are longer and more complex in the 1835 version. This aspect of the work has never been directly addressed by critics, as far as we know.
2. Discourse connectives are also less frequent in the 1835 version, where Balzac made more use of juxtaposition, placing sentences side by side without any explicit connections between them.

Observations concerning isolated lexical items, as well as punctuation marks, can therefore be meaningful at the sentence level, as we discuss in greater detail below.

The morphological analysis performed with TreeTagger failed to reveal any meaningful changes at this level. The use of modals, verb tenses and determiners seemed to remain relatively stable from one version to the other (which is not the case for all Balzac's novels). We think it is also important to spot instances of stability, as the absence of change can be just as meaningful when interpreting changes in the text.

As for the sentence and discourse levels, we have already observed that sentences are longer and more complex in the 1835 version. The linguistic tricks used

by Balzac to make the sentences longer include greater use of personal pronouns and relative clauses.

Automatic analysis can highlight various systematic patterns in discourse particle use. For example, in the 1832 edition, Balzac frequently used a semicolon followed by the French conjunction *car* (for), which is not "stylistically correct" (esp. according to critics of the time). All occurrences of this construction had disappeared in the 1835 version: Balzac either deleted the word *car*, so as to obtain two separate sentences, or else removed the semicolon, leaving the conjunction on its own.

(1) *Cette mission lui plut ; <u>car</u> l'amour fait son profit de tout, et rien ne séduit plus un jeune homme que de jouer le rôle d'un bon génie, auprès d'une femme.* (1832)

   Cette mission lui <u>plut</u>. L'amour fait son profit de tout, et rien ne séduit plus un jeune homme que de jouer le rôle d'un bon génie auprès d'une femme. (1835)

The same can be observed for the conjunction *donc* (therefore). A brief look at word forms shows that *donc* remained stable, whereas *Donc* (with the uppercase D) disappeared entirely, meaning that in 1835, it was no longer used to start sentences. From a stylistic point of view, it is supposed to be better to integrate conjunctions within the sentence, rather than just putting them at the beginning.

(2) *Schinner avait choisi ses amis parmi les hommes les plus honorables et les plus distingués.* (1832)

   Schinner avait <u>donc</u> choisi ses amis parmi les hommes les plus honorables et les plus distingués. (1835)

This kind of change is probably a consequence of comments Balzac received in the press: he was often portrayed as a writer with a poor style, using colloquial and improper expressions.

5. Conclusion: Which process gives access to a genetic and linguistic analysis of variation?

Genetics shares the variation problem with other textual approaches. However, unlike other points of view that can lead to two texts being treated as variations of the same text (considering every form of discursive re-elaboration, including translation, plagiarism, imitation, vulgarization, transposition, etc.), the genetic outlook on textual variation (a) explores the modalities of passing from Versions A to B, and (b) does so by adopting the point of view of the text's reader-writer.

In this chapter, we have looked at the conditions and constraints that influence this rewriting activity, particularly when it takes place after the text's publication, thus generating several versions of the *same* text.

Seeking to understand a process, even though it only can observe different products, the genetic approach can be defined as a *poetics of transition between states* (Lebrave 2009), where linguistic analysis is applied to *n* discrete textual variations in order to reconstruct the process whereby Text A *morphs* into Text B.[9] However, the reconstructed process may not literally correspond to the transformation of A into B. In reality, the observed variations merely provide material for a *differential semantics* that consists in comparing the effects of varying sequences on their supposedly identical co-texts. The reconstructed process concerns the motives that lead the reader-writer to prefer the second sequence to the first, and to move from the one to the other. Variation linguistics is not intended to provide writing research with a readymade instrument for analyzing transformation. Rather, it is about the *poetics* or, more accurately, the *hermeneutics* of the *transition between states*. It describes the precise nature of the transition from one state to another in linguistic terms, and this description enables us to formulate hypotheses about the reasons behind this change by the writer, be it consciously or unconsciously, in his or her particular context, depending on how far we can reconstruct his or her linguistic skills.

Our contribution is thus to move from written linguistics to writing linguistics, from the linguistics of product to the linguistics of process.

References

de Biasi, Pierre-Marc. 2005 [2000]. *La Génétique des textes*. Paris: Nathan.

Bourdaillet, Julien, Jean-Gabriel Ganascia, and Jean-Louis Lebrave. 2009. "Topologie et génétique textuelles: un dialogue médié par la machine." *Lexicometrica* 9 special issue: unnumbured (available online: http://lexicometrics.univ-paris3.fr/numspéciaux/special9/bourdaillet.pdf).

---

9.   If we examine the traces left *by* the writing (and not *in* the writing itself), and if we consider them from a linguistic point of view, the "deepness of relationship between all forms of textual variation" (Lebrave 2009, 18) allows to assert that the *linguistic treatment* of variation and the observable traces in the manuscripts are identical. One should either way seek to understand the differences between what we call *textual variations*.

Culioli, Antoine. 1983. "Préface." In *La genèse du texte: les modèles linguistiques*, ed. by Jean-Louis Lebrave, Catherine Fuchs, Almuth Grésillon, Jean Peytard, and Josette Rey-Debove, 9–12. Paris: CNRS Editions.

Falconer Graham. 1969. "Le travail du style dans les révisions de *La Peau de chagrin.*" In *L'Année balzacienne*: 71–106.

Ferrer, Daniel. 2011. *Logiques du brouillon. Modèles pour une critique génétique*. Paris: Seuil.

Fuchs, Catherine. 1983. "Eléments pour une approche énonciative de la paraphrase dans les brouillons de manuscrits." In *La genèse du texte: les modèles linguistiques*, ed. by Jean-Louis Lebrave, Catherine Fuchs, Almuth Grésillon, Jean Peytard, and Josette Rey-Debove, 73–102. Paris: CNRS Éditions.

Grésillon, Almuth. 1994. *Éléments de critique génétique. Lire les manuscrits modernes*. Paris: PUF. DOI: 10.1515/arbi.2000.18.1.1

Grésillon, Almuth. 2007. "'Nous avançons toujours sur des sables mouvants.' Espaces et frontières de la critique génétique." In *La création en acte. Devenir de la critique génétique*, ed. by Paul Gifford, and Marion Schmid, 29–40. Amsterdam/New York: Rodopi.

Heiden, Serge. 2010. "The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme." In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation* (PACLIC 24), ed. by Ryo Otoguro, Kiyoshi Ishikawa, Hiroshi Unemoto, Kei Yoshimoto, and Yasunari Harada, 389–398. Sendai: Tohoku University.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. "The Sketch Engine." In *Proceedings of EURALEX*, ed. by Geoffrey Williams, and Sandra Vessier, 105–116. Lorient: Université de Bretagne Sud.

Lafon, Pierre. 1980. "Sur la variabilité de la fréquence des formes dans un corpus." *Mots* 1: 27–165. DOI: 10.3406/mots.1980.1008

Lebart, Ludovic, André Salem, and Lisette Berry. 1997. *Exploring Textual Data*. Dordrecht: Kluwer Academic Press. DOI: 10.1007/978-94-017-1525-6

Leblay, Christophe, and Gilles Caporossi. 2014. *Temps de l'écriture. Enregistrements et représentations*. Paris: L'Harmattan.

Lebrave, Jean-Louis. 1983. "Lecture et analyse du brouillon." *Langages* 69: 11–23. DOI: 10.3406/lgge.1983.1139

Lebrave, Jean-Louis. 2009. "Manuscrits de travail et linguistique de la production écrite." *Modèles linguistiques* 30: 13–21. DOI: 10.4000/ml.330

Lewis, Ricki. 2005. *Human Genetics: Concepts and Applications*. Boston, MA: McGraw Hill.

Mahrer, Rudolf, and Valentine Nicollier Saraillon. 2015. "Les brouillons font-ils texte? Le cas des plans pré-rédactionnels de C. F. Ramuz." In *Faire texte. Unité*(*s*) *et* (*dis*)*continuité*, ed. by Jean-Michel Adam, (to be published). Besançon: Presses Universitaires de Franche-Comté.

Peytard, Jean. 1993. "D'une sémiotique de l'altération." *Configurations discursives, Semen* 8. Retrieved on 8 April 2014. URL: http://semen.revues.org/4182

Schmid, Helmut. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees." In *Proceedings of International Conference on New Methods in Language Processing*, ed. by Harold Somers, and Danny Jones, 44–49. Manchester: NeMLap.

Souchier, Emmanuël. 2007. "Formes et pouvoirs de l'énonciation éditoriale." *Communication et langage* 154: 23–38. DOI: 10.3406/colan.2007.4688

Vachon, Stéphane. 1997. "Les enseignements des manuscrits d'Honoré de Balzac. De la varia-
tion contre la variante." *Genesis* 11: 61–80. DOI: 10.3917/balz.010.0199

Vachon, Stéphane. 2009. "Perspectives en génétique balzacienne." In *Balzac, Flaubert. La genèse
de l'œuvre et la question de l'interprétation*, [Proceedings of the Second International Con-
ference, 14–16 December 2007, Nagoya], ed. by Kazuhiro Matsuzawa, 35–45. Nagoya,
Japan: Graduate School of Letters, Nagoya University.

A graph theory approach to online writing data visualization

Christophe Leblay & Gilles Caporossi

ITEM-TERs and University of Turku, Finland / ITEM-TERs and HEC Montreal, Canada

There are currently several systems for collecting online writing data using keystroke logging. Each of these systems provides reliable and very precise data. Unfortunately, with the exception of very brief recordings, such huge amounts of data are generated that it is virtually impossible to analyze them. In this chapter, we describe a representation technique based on graph theory that allows the writing process to be understood from a fresh viewpoint. This application was originally intended to represent the data provided by ScriptLog, but the concepts can be applied in other contexts, too.

Keywords: keystroke logging; writing process; visualization; graph theory

1. Introduction

The process of writing (activity), unlike its end product (text), has two dimensions: space and time. Although there is obviously a relationship between process and product, it is not possible to reconstruct the former by analyzing the latter. It is therefore interesting and important to find ways of studying the writing activity per se, bearing in mind that word processing differs from handwriting.

Recent approaches to the study of writing based on online recordings contrast with those based on the analysis of paper versions, in that the latter focus on the page space and the former on the temporal dimension.

Writing models based on online recordings were first developed in the 1980s, in the wake of Matsuhashi (1987)'s pioneering work. Adopting a bipolar division, Matsuhashi suggested distinguishing between the conceptual level (semantics, grammar and spelling) and the sequential plan (planning and phrasing).

Virtually all the research that followed concerned the software used for recording, with Ahlsén and Strömqvist (1999), Wengelin (2006), and Doquet and Leblay (2014) focusing on ScriptLog software applications, Sullivan and Lindgren (2006) on JEdit applications, Van Waes and Schellens (2003) and Van Waes and

Leijten (2006) on Inputlog software, Jakobsen (1999, 2006) on TransLog software, and Chesnet and Alamargot (2005) on Eye and Pen software.

While not denying all the work done on text revisions treated as a product (final text), these recent approaches all indicate that writing is primarily a temporal activity. The multitude of software approaches developed for the online recording of the writing activity reflects a clear interest in the study of the writing process. Be it from a cognitive psychology or a didactic point of view, analyzing the writing activity as a process is very important for researchers. However, although the nature of the recording depends on the type of software used, the resulting *logfiles* all share similar characteristics. In particular, they are all exhaustive, and the large amounts of data they contain are difficult for human researchers to analyze. Depending on the point of view and the topic of the research, the data therefore have to be processed and filtered. They need to be converted into a more legible format for the researcher. They may become less exhaustive in the process, but we are left with the information that is most relevant to the analysis.

An important topic in raw data transformation is visualization. Visualization allows researchers to form intuitions and develop a clearer understanding of the underlying process of creation. As the goal of the representation changes with the research topic, we cannot expect a single representation to be adapted to every possible study.

Although the technical format of the data depends on which software is used, all systems (except in the case of handwriting) record the same elementary events (i.e. keyboard keystrokes or mouse clicks). An example from ScriptLog is shown in Figure 1.

time	type	from	to	key
0.00	10	1	0	<START>
4.21	7	0	0	L
4.46	7	1	1	I
4.75	7	2	2	E
5.05	7	3	3	U
5.26	7	4	4	
5.70	7	5	5	I
5.86	7	6	6	D
8.08	7	7	7	...
8.36	7	8	8	A
8.53	7	9	9	L
8.81	7	10	10	
12.28	5	11	11	<DELETE>
12.45	5	10	10	<DELETE>
12.61	5	9	9	<DELETE>
12.78	5	8	8	<DELETE>

Figure 1. Excerpt from a logfile produced by ScriptLog

Here, each line is associated with an elementary event. Technically speaking, this record represents the whole of the writing process, but without any prepro-

cessing it is extremely unwieldy, and this preprocessing will differ according to the task being carried out by the researcher.

For instance, studying the length and location of pauses does not require the same level of information as studying the text revision process. In both cases, the same basic information is used, but the researchers have to apply the aggregation to a different level, according to their needs. Data preprocessing is sometimes unavoidable, and given the large amount of data produced by the system (a logfile corresponding to a 15-minute recording may contain up to 2,000 lines), this preprocessing should ideally be automated, in order to avoid errors.

2. Visualization techniques

Although visualization is an important part of writing process analysis, only a few visualization techniques currently exist.

One of these techniques is so-called *linear representation*, in which every character that is written is displayed. If a portion is deleted, for instance, it is crossed out instead of being removed, in order to show the process of text production and not just the final product. Cursor movements using arrows or the mouse are also identified, so that it is possible to track the text construction process. This kind of representation emphasizes the spatial dimension, and the temporal dimension is difficult to follow (e.g. the user has to interpret numbers indicating cursor movements, which is not very convenient). An example of linear representation is given in Figure 2. This type of representation has the advantage of displaying the text, but it can be difficult to decipher when the process of creation is particularly complex.
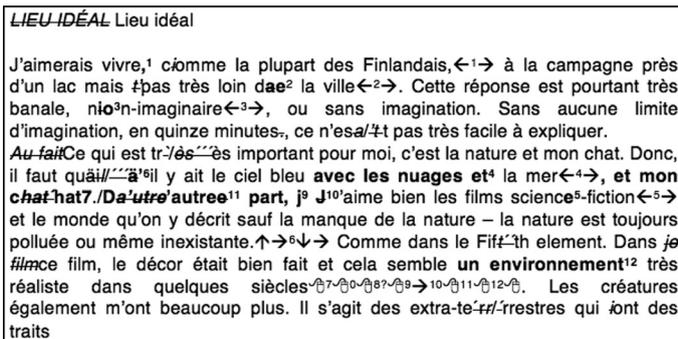
Figure 2. Linear representation of a short, 15-minute text (Leblay 2009)

Another way of visualizing the creative process is to focus on just a few values representing the text produced so far. The *Fil de la plume* graph (Chenouf et al. 1996) displays the position of the cursor, as well as the total length of the text, as

a function of time, and indicates zones where text that has already been written is modified by the writer, as we can see in Figure 3. This type of representation, which relies on a geographic information system, is known as *GIS representation*, and is used in various software, such as Inputlog (Van Waes & Leijten 2006).

Although this approach represents both spatial and temporal dimensions, a major weakness of GIS representation is that the position of visible text may cease to be correct when insertions or deletions occur upstream. If the position of a character is altered, it becomes difficult to figure out which parts of the text are involved in any subsequent modifications. Another drawback is that users cannot tell which points in the graph correspond to which places in the text.
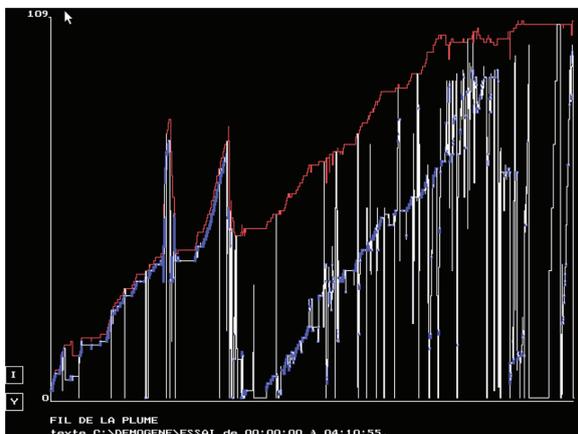
Figure 3. *Fil de la plume* GIS representation (Chenouf 1996)

3.   An alternative: The graph representation

In this section, we describe a new representation technique (Leblay & Caporossi 2014) that allows for the visual identification of basic operations such as insertion and deletion, but also makes it possible to isolate portions of the document according to the processing activity performed by the writer. Each time this new representation technique has been presented to psychologists or linguists, it has been given a very positive reception.

To sidestep the problem of the written text changing position as a result of revisions, we have come up with a slightly different approach whereby the characters are given a relative, rather than an absolute position, when they are first written. This solution has proved to be more suitable for representing the dynamic aspect of the writing activity. Keystroke sequences are merged to form a textual entity that is represented by a node in the graph. If two nodes interact, either chronologically or spatially, they are connected by an edge (or link) showing this relation.

Graphs are mathematical tools based on nodes or vertices that can be connected by links or edges. Application fields may be more or less concerned by graph theory. For instance, some graph-theoretical results directly apply to chemistry (Caporossi et al. 1999a and 1999b). For other fields, such as transportation, scheduling and communication, the algorithms underlying graph theory and networks may be more relevant. Since the 1990s, graphs in the shape of concepts maps have also been used for representational purposes in the human sciences (Novak 1990). Here, we describe how graph representation can be used to visualize data pertaining to the writing process. Examples of graph representations of the writing process are provided in Figures 4 (novice) and 5 (expert).
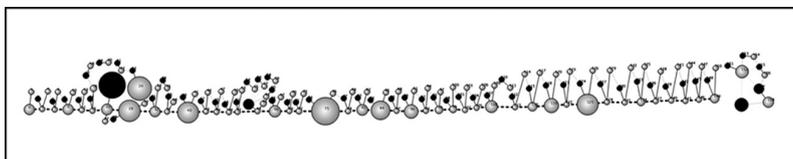
Figure 4. Graph visualization: an example of a novice writer (global view)
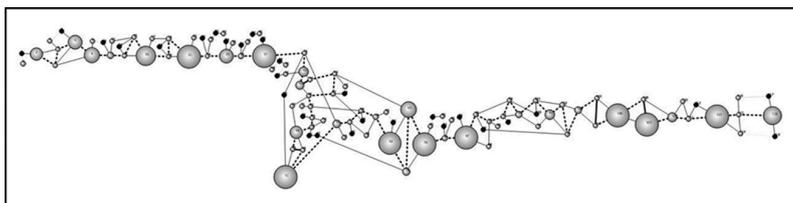
Figure 5. Graph visualization: an example of an expert writer (global view)

The size of a vertex depends on the number of elementary operations it represents. In the case of the novice writer, we can see that there are only a few large nodes, indicating a higher frequency of errors or typos. We could have displayed the section of text corresponding to each node, thus providing a virtually linear representation, but decided not to here, in order to keep the graphs as simple as possible.

The structure of the graph is also very informative: it is almost linear for the novice, whereas for the expert, the central portion is much more complicated. This complex portion, in the middle of the graph Figure 5, represents a section of the text that was rewritten and changed at a higher level and clearly not from just the lexical point of view. Analysis of that portion of the graph reveals that the author modified the text in four successive passes.

There are three different types of key or mouse events: (i) additions or insertions of characters or spaces; (ii) deletions of characters or spaces; and (iii) cursor moves using arrows or the mouse. Spatially and temporally contiguous sequences are merged and represented by the nodes.

3.1   Nodes

The size of a node reflects the number of elementary events it represents, and its color the nature of these events. A light color represents an addition, and black a deletion. The nodes are numbered according to their order of creation.

3.2   Links

The nodes are connected by links (or edges) representing spatial or temporal relations. The nature and width of these edges indicate the type of relation. A solid line represents a chronological link (solid lines link Node 0 to the last node, running through all the nodes in chronological order). All the other links between the nodes correspond to spatial relations. The link between an addition node and its deletion counterpart is narrow, and the spatial link between nodes that form part of the final text is broad. The content of the nodes with broad links therefore corresponds to the final version of the whole text (these broad links form a path representing the final text; the spatial dimension).

3.3   Analysis of graphic patterns

These graphs can be analyzed in a number of different ways, and here we concentrate on the most useful ones. We begin by identifying patterns that correspond to some of the classic operations in the writing process. From a technical point of view, these operations correspond to easily recognizable subgraphs. It is useful to be able to identify these subgraphs, in order to analyze the graph as a representation of the writing process.

3.3.1   *Additions and insertions*
Text can be added in three ways, two of which are very similar:

i. Adding text to the end of the node that is currently being written is not represented by any particular pattern, but the node increases in size;
ii. Inserting text in the node that is currently being written (i.e. not at the end) causes the node to split, as illustrated in Figure 6;
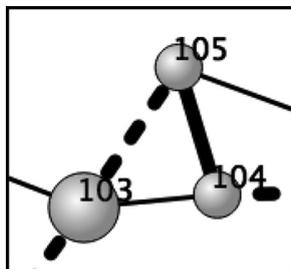
Figure 6. Insertion in the current node: Nodes 103 and 104 were merged until Node 105 was inserted

iii. Inserting text in a node that has already been written causes the node to split, and the resulting configuration is shown in Figure 7.
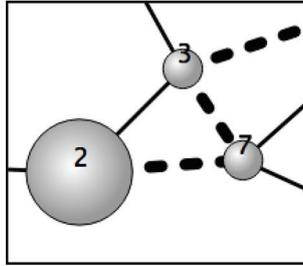
Figure 7. Insertion: Nodes 2 and 3 were merged until Node 7 was inserted

From the graphic and linguistic standpoints, *insertions* correspond to internal additions (ii and iii) while *addition* corresponds to an extension of the text (i).

3.3.2   *Deletions*

Like additions, deletions result in different subgraphs, depending on whether they erase the end of the most recent node, an internal part of the most recent node or an internal part of a node that has already been written:

i.   The case of an immediate deletion (e.g. after a typing error) is shown in Figure 8. We can see that this was the most frequent operation for the *novice* (Figure 4), meaning that the text was scarcely modified once it had been written;
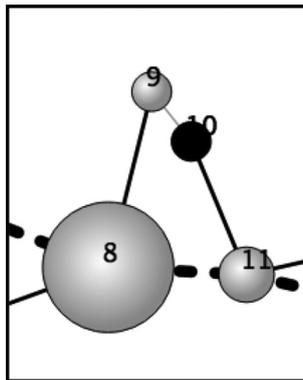
Figure 8. Immediate deletion: Nodes 8 and 9 were merged until the portion of text corresponding to Node 9 was deleted (Node 10)

ii.   A deletion in the most recent node (not at the end) is shown in Figure 9;
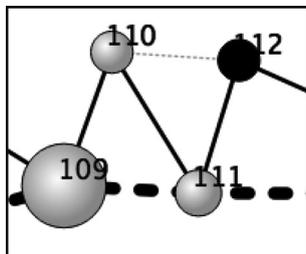
Figure 9. Deletion in the most recent node (but not at the end): Nodes 109, 110 and 111 were merged until the portion of text represented by Node 110 was deleted (Node 112)

iii. Delayed deletion results in the subgraph shown in Figure 10.
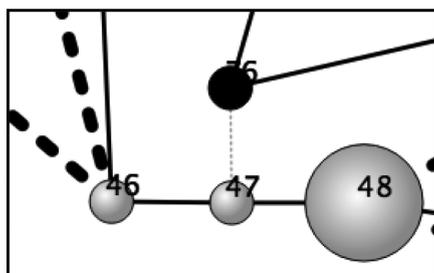
Figure 10. Deletion: Nodes 46, 47 and 48 were merged until the portion of text represented by Node 47 was deleted (Node 76)

3.3.3   *Substitutions*

Although more complex operations can be regarded as sequences of the simple operations described above, they nevertheless give rise to particular subgraphs that are easily recognized. For instance, replacement can be viewed as deletion immediately followed by insertion at the same place. Figure 11 represents the subgraph corresponding to a replacement in a node that has already been written. We can see that substitutions usually occur when writers are correcting their text, this activity usually being associated with *expert* writers rather than *novices*.

4.   Summary and future research directions

In this chapter, we described a new technique for representing written language production that offers a solution to the problem of shifting text positions. This technique allows the researcher to easily identify each portion of the document that is modified by the writer. According to linguistics researchers working on the writing process, this type of representation is easier to understand than those that were
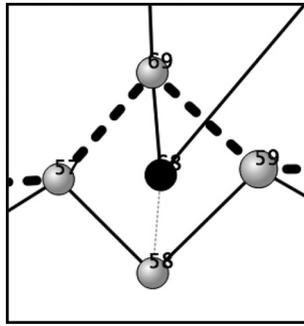
Figure 11. Substitution: Nodes 57, 58 and 59 were merged until the portion of text corresponding to Node 58 was deleted (Node 68) and replaced by an insertion (Node 69)

previously available. One major advantage is the ability to visualize modification patterns from both a spatial and a temporal point of view in the same representation. Intuition also seems to be stimulated more by a graph representation than it is by linear or GIS representations.

Several important aspects of graph representation in writing require further investigation, especially the temporal dimension. For instance, nodes corresponding to long pauses (the definition of the minimum duration of a pause being defined by the user) could be inserted, or the time and duration corresponding to each node could be indicated.

It is also important to distinguish between the various levels of text improvement as defined by Faigley & Witte (1981), by distinguishing surface modifications (correction of typos, orthographic adjustments, etc.) from text-based modifications (reformulation, syntactic modifications, etc.). A first step in this direction would be to highlight nodes containing more than one word, possibly with a space before and after visible characters. A second step would be to better define the nature of the transformation represented by a given node. This would require computational linguistics tools.

The graph is currently drawn by hand. We therefore need to devise an algorithm that can automatically place vertices in such a way that (i) patterns are easy to recognize and (ii) the spatial aspect is preserved as much as possible, so that the writing process remains easy to follow.

References

Ahlsén, Elisabeth, and Sven Strömqvist. 1999. "ScriptLog: A Tool for Logging the Writing Process and its Possible Diagnostic Use." In *Augmentative and Alternative Communication: New Directions in Research and Practice*, ed. by Filip Loncke, John Clibbens, Helen Arvidson, and Lyle Lloyd, 144–149. London: Whurr Publishers.

Caporossi, Gilles, Dragos Cvetkovic, Ivan Gutman and Pierre Hansen. 1999a. "Variable Neigh-borhood Search for Extremal Graphs 2. Finding Graphs with Extremal Energy." *Journal of Chemical Information and Computer Sciences* 39: 984–996. DOI: 10.1021/ci9801419

Caporossi, Gilles, Ivan Gutman, and Pierre Hansen. 1999b. "Variable Neighborhood Search for Extremal Graphs 4. Chemical Trees with Extremal Connectivity Index." *Computers and Chemistry* 23(5): 469–477. DOI: 10.1016/s0097-8485(99)00031–5

Chenouf, Yvanne, Jean Foucambert, and Michel Violet. 1996. *Genèse du texte*, vol. 2 [INRP research report].

Chesnet, David, and Denis Alamargot. 2005. "Analyse en temps réel des activités oculaires et grapho-motrices du scripteur: intérêts du dispositif Eye and Pen." *L'année psychologique* 105(3): 477–520. DOI: 10.3406/psy.2005.29706

Doquet, Claire, and Christophe Leblay. 2014. "Temporalité de l'écriture et génétique textuelle: vers un nouveau langage?" In *Actes du Congrès mondial de linguistique française vol. 8*, ed. by Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer, and Sophie Prévost, 2767–2781. Berlin. ⟨http://www.shs-conferences.org/articles/shsconf/abs/2014/05/shsconf_cmlf14_01204/shsconf_cmlf14_01204.html⟩ DOI: 10.1051/shsconf/20140801204

Faigley, Lester, and Stephen Witte. 1981. "Analysing Revision." *College Composition and Com-munication* 32: 400–414. DOI: 10.2307/356602

Jakobsen, Amt Lykke. 1999. "Logging Target Text Production with Translog." In *Probing the Process in Translation. Methods and Results*, ed. by Gyde Hansen, 9–20. Copenhagen: Samfundslitteratur.

Jakobsen, Amt Lykke. 2006. "Research Methods in Translation: Translog." In *Computer Key-Stroke Logging and Writing: Methods and Applications*, ed. by Kirk Sullivan and Eva Lind-gren, 95–105. Amsterdam: Elsevier.

Leblay, Christophe, and Gilles Caporossi. 2014. *Temps de l'écriture. Enregistrements et représen-tations.* Louvain-La-Neuve: Academia-L'Harmattan (coll. Sciences du langage: Carrefours et points de vue). DOI: 10.4000/questionsdecommunication.6988

Leblay, Christophe. 2009. "Les invariants processuels. En deça du bien et du mal écrire." *Pra-tiques* 143/144: 153–167. DOI: 10.4000/pratiques.1430

Leijten, Mariëlle, and Luuk Van Waes. 2005. "Writing with Speech Recognition: The Adaptation Process of Professional Writers." *Interacting with Computers* 17(6): 736–772. DOI: 10.1016/j.intcom.2005.01.005

Lindgren, Eva, Kirk Sullivan, Urban Lindgren, and Kristyan Spelman Miller. 2007. "GIS for Writing: Applying Geographic Information System Techniques to Data-Mine Writing's Cognitive Processes." In *Writing and Cognition: Research and Applications*, ed. by Gert Rij-laarsdam (series ed.) and Mark Torrance, Luuk Van Waes and David Galbraith (vol. eds.), 83–96. Amsterdam: Elsevier.

Matsuhashi, Ann. 1987. "Revising the Plan and Altering the Text." In *Writing in Real Time*, ed. by Ann Matsuhashi, 197–223. Norwood, NJ: Ablex Publishing Corporation.

Novak, Joseph D. 1990. "Concept Maps and Vee Diagrams: Two Metacognitive Tools for Science and Mathematics Education." *Instructional Science* 19: 29–52. DOI: 10.1007/bf00377984

Strömqvist, Sven, and Henrik Karlsson. 2002. *ScriptLog for Windows – User's manual*. Technical Report. University of Lund: Department of Linguistics and University College of Stavanger, Centre for Reading Research.

Sullivan, Kirk, and Eva Lindgren (Eds.). 2006. *Computer Keystroke Logging and Writing. Methods and Applications*. Amsterdam: Elsevier.

Van Waes, Luuk, and Peter Jan Schellens. 2003. "Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers." *Journal of Pragmatics* 35 (6): 829–853. DOI: 10.1016/s0378-2166(02)00121–2

Van Waes, Luuk, and Mariëlle Leijten. 2006. "Inputlog: New Perspectives on the Logging of On-Line Writing Processes in a Windows Environment." In *Computer Key-Stroke Logging and Writing: Methods and Applications*, ed. by Kirk Sullivan, and Eva Lindgren, 73–93. Amsterdam: Elsevier.

Wengelin, Åsa. 2006. "Examining Pauses in Writing: Theories, Methods and Empirical Data." In *Computer Key-Stroke Logging and Writing: Methods and Applications*, ed. by Kirk Sullivan, and Eva Lindgren, 107–130. Amsterdam: Elsevier.

PART IV

Writing practices in context

Writing, literate activity, semiotic remediation

A sociocultural approach

Paul A. Prior
University of Illinois at Urbana-Champaign

This chapter sketches a sociocultural framework for understanding writing, particularly exploring the way notions of *literate activity* (Prior 1998) and *semiotic remediation* (Prior, Hengst, Roozen & Shipka 2006; Prior & Hengst 2010) maintain a distinct interest in both written artifacts and associated actions that are dialogically dispersed across people, tools, times and places. Drawing in particular on traditions grounded in the work of Vygotsky and Voloshinov, this approach argues for seeing writing as chronotopically laminated trajectories. Theoretical and methodological implications of this approach are illustrated by reviewing a line of research that has investigated writing as situated, mediated, and dispersed. The chapter concludes with key implications of this approach for the research, teaching, and practice of writing.

Keywords: writing theory; sociocultural theory; writing research; semiotics; chronotopes

1.   Introduction

This chapter emerges from the Crossroads workshop on the process-product interface that led to this volume. That workshop and this volume have been grounded in the observation that there are multiple disciplinary approaches to tracing writing processes and to analyzing written texts, but a lack – whether within process approaches, within textual approaches, or between process and textual approaches – of the kind of sustained dialogue that is needed to construct a full and coherent account of writing. As Cislaru (in the introduction to this volume) states:

> …this book aims to contribute to the development of an interpretive approach to writing and its dynamics. It gives an overview of the state of research on the process-product interface through a range of viewpoints on process, product, and the links between them. Collectively, its chapters explore the possibility of establishing a coherent path from the real-time dynamics of the writing process to the product anchored in its formal dimensions and its pragmatic functions. The result is a look

at how results and concepts from different domains may support each other in the development of a mixed approach to the process-product interface.

The idea of a process/product interface might index a fuller scheme – product, process, and social context – that has been offered to narrate a certain arc in writing theory, research, and pedagogy (see, e.g. Nystrand 2005; Galbraith & Rijlaarsdam 1999). In that scheme, research (mostly cognitive) on writing processes replaced an earlier (mainly pedagogical and rhetorical) product tradition, and then was itself soon eclipsed by approaches (anthropological, rhetorical, postmodern) that focused on the social contexts of writing. Prior and Thorne (2014) critiqued this scheme and its underlying narrative, proposing instead a fine-grained, multidimensional mapping to articulate particular research projects. Drawing on Prior and Lunsford (2007), that mapping was organized around five key senses that co-exist in the word *writing*:

1. as an inscriptional artifact (a text),
2. as an individual cognitive and embodied capacity to act (the literate competence, dispositions, repertoires of the writer),
3. as the situated activity writers engage in (the chain of actions undertaken by writers and their respondents in a focal process – whether fleeting or lasting many years – of producing a text),
4. as a set of technologies or mediational means (whether digitized word processing, calligraphy inscribed on vellum, or use of a wooden stick to write in the dirt), and
5. as a mode of social organization (e.g. Smith's [1974] notion of documentary reality).

Given this scheme (text-person-activity-mediation-society), a study of writing might investigate one or more of these potential objects of inquiry, using one or more methods and pursuing one or more broad goals.

The Crossroads workshop underscored that theoretical and research attention (linguistic, discursive, rhetorical, text-genetic) to textual products continues to be active and productive, as does research on micro processes of composing, as does observational, ethnographic, and cultural examination of writing practices and histories. In fact, it is not unusual for studies to combine attention to two or all three of these dimensions of literate practice. Olive and Cislaru (this volume), for example, represent a process-product blending as they analyze bursts of (elicited) inscription in terms of both the timing of production and the linguistic-discursive features of the textual bursts. Perrin's research (2014; Grésillon and Perrin this volume) on the multimodal composing and delivery of TV news programs in the Swiss Broadcasting Corporation highlights with particular clarity and depth the value of combined attention to tracing products, processes, and social contexts in seeking a "coherent path from real-time dynamics of the writing process to the product anchored in its formal dimensions and its pragmatic functioning" (Cislaru this volume, Introduc-

tion). This chapter considers how a sociocultural approach to writing (Prior 2006) can address the integration of text, cognition, and social practices.

2. A sociocultural perspective

Sociocultural (also referred to as cultural-historical activity theory or CHAT) approaches to writing are often grounded in traditions that have developed from Vygotsky's (1987) approach to the sociocultural development of mind and the dialogic semiotics of Voloshinov (1973) and Bakhtin (1986). This sociocultural framework sees human activity as always mediated by other people, tools, and practices (Wertsch 1991; Cole & Engeström 1993; Scribner 1997), so texts must be understood as artifacts-in-activity. Given Vygotsky's (1987) emphasis on the genesis of tools (material and psychological) and people (learning/development) and Voloshinov's insistence that language (indeed any type of cultural sign) is "a purely historical phenomenon" (1973, 83), sociocultural approaches emphasize concrete chains of history and the complex ways temporality is folded into people, objects, environments, and practices. This focus on concrete histories challenges structuralist (e.g. Saussurean, schema-theoretical) accounts that posit synchronous, homogenous systems governing discourse and society (see critiques in Prior 1998; Bloomaert 2010).

Working in this sociocultural tradition, Lemke noted the importance of understanding how multiple temporal scales of activity are integrated:

> Each scale of organization in an ecosocial system is an integration of faster, more local processes (i.e. activities, practices, doings, happenings) into longer-timescale, more global or extended networks. It is *relative timescale* that determines the probability and intensity of interdependence …, and it is the circulation through the network of *semiotic artifacts* (i.e. books, buildings, bodies) that enables coordination between processes on radically different timescales.
> (Lemke 2000, 275)

Similarly, in research on navigation in navy ships, Hutchins (1995) highlighted *heterochronicity*, the way a mix of elements from multiple past activities undertaken by varied people for varied purposes across varied settings are folded into a specific local stretch of distributed cognition and activity. In sociocultural approaches, this weaving, unweaving, and reweaving of historical trajectories takes the place of abstract social and cognitive structures, a view well articulated by the flat ontology and sociology of Latour's (2005) actor-network theory (ANT). This chapter sketches a sociocultural framework for understanding writing, particularly considering how alternate units of analysis like *literate activity* (Prior 1998) and *semiotic remediation* (Prior, Hengst, Roozen & Shipka 2006; Prior & Hengst 2010) entail examining *laminated chronotopic* trajectories

of activity (Prior & Shipka 2003; Prior & Schaffner 2011) that are folded into and felt in both processes of writing and situated engagements with texts.

3.   From writing to literate activity

Cultural images of writing typically picture an individual inscribing a material text (usually on paper or screen). Such images fit Goffman's (1981) observation that prototypical models of communication conflate the roles of animator (speaker/ inscriber), author (who composes words, meanings, purposes), and principal (whose ideas and interests are being represented), assuming all three reside fully in the individual producer. Goffman noted many exceptions (e.g. reported speech, ghostwriters, diplomatic communications) to this prototypical model, stressing instead the multiple footings and laminations typical of communicative encounters. For sociocultural theory, these three roles in production are by necessity always distributed, and similar dialogic complexity in reception is central to any *utterance* (Bakhtin 1986; Voloshinov 1973), as utterances are situated wholes that always involve production, reception, distribution, and use. I did not make the pencil and paper I first wrote the last sentence with, nor did I create the words, their orthography, or the intertextual affordances of many other texts that may infuse that sentence with a range of meanings. The discourses and tools I use speak through me in some measure, however much I try to make them express my sense. Indeed, sociocultural theory takes everything – the writer, the language, the tools of inscription, the social purposes, genres, and uptakes of texts – as socially and historically made, hence as distributed and laminated. Understood as a blend of texts, persons, activities, mediational means, and social formations/practices, writing appears as temporally and spatially stretched out trajectories rather than as punctual events in a narrow and isolated here-and-now.

From this perspective, whether writing refers to processual acts of inscription, to texts produced by such acts, or both, it can only be a synecdoche for longer, broader histories of semiotic activity. From the start, cognitive process research (e.g. Flower & Hayes 1981) highlighted *reviewing*, the amount of immediate reading and rereading of emerging text observed in tightly focused acts of inscription on its own (that is, in the typical lab task of elicited writing without other texts or people present to consult). In fact, the composing of any text has cognitive and social dimensions that reach beyond any immediate act of inscription, if indeed they are connected to such an act: a person walking on a beach or a couple talking and sharing a beer at a pub can be engaged in a writing process. Following from these kinds of observations, Witte (1992) argued that the unit of analysis for research on writing could be little less than *unlimited semiosis*, as Eco (1976) interpreted Peirce's (1998) triadic semiotics. The idea of unlimited semiosis resonates with Bakhtin's (1986) proposal to take dialogic *utterances* as the

basic units of his translinguistics and his assertion of the unfinalizability of utterances – the way they remain open to re-use, re-interpretation, and re-purposing.

In my early research (e.g. Prior 1991; 1995; 1998), I explored connections between writing and disciplinary enculturation in graduate seminars. The research involved observation and recording of seminar meetings, collection of students' drafts and final papers (including professors' written comments and, when given, grades), semi-structured and text-based interviews, and historical contextualizations, for example, of the history of seminars or of a specific discipline. Analysis of these data identified the complex co-authoring of texts, ideas, discourses, and social formations that emerged through chains of events that involved a mix of reading, writing, talking, observing, and acting.

To account for these data, I proposed in Prior (1998) that an appropriate unit of analysis for writing research would be *literate activity*, "a confluence of many streams of activity: reading, talking, observing, acting, making, thinking and feeling as well as transcribing words" (xi) in whatever medium, "activity that is not only multimodal, but also temporally and spatially dispersed and distributed across multiple persons, artifacts, and sites" (Prior 1998, 137). I argued that sociocultural approaches to writing must grapple with the fact that texts and acts of inscription "are no more autonomous than the spray thrown up by white water in a river, and like that spray, literate acts today are far downstream from their sociohistoric origins" (*ibidem*, 138). Drawing on Bakhtin's (1981) notion of utterance as *chronotopic* (as a phenomenon always situated in and indexing times-spaces), I argued that writing must be understood "as situated, mediated, and dispersed," as an aspect of literate activity that is "not located in acts of reading and writing, but as cultural forms of life saturated with textuality, that is strongly motivated and mediated by texts" (*ibid.*, 138).

As I engaged in analysis of the graduate seminars, however, I also realized that my research designs (centered on graduate students writing in seminars) had imagined a relatively fixed chronotopic scene (the site of a seminar, the time of a semester). One case in particular made the problem of this chronotopic fixity clear. Although I had invited all 60 students in a study of four graduate seminars to keep process logs, in the end only one did so. Lilah was taking an American Studies course taught by a geography professor, who I called Kohl. My frame tacitly assumed that she was writing a paper for Kohl's American Studies seminar. However, her log constantly referred to papers from two other courses, making it plain that Lilah was navigating the three courses in a combined way. For example, although she really did not find Kohl's course or Kohl himself that compelling, she liked the ethnographic orientation Kohl pushed in his seminar. She decided to write her papers for both American Studies and a History seminar on local celebrations of Cinco de Mayo (the May 5th celebration commemorating a Mexican victory over the French at Puebla in 1862). In spite of her lack of enthusiasm for Kohl, Lilah carried his ethnographic framework into her history paper, although it turned

out the history professor did not find that kind of evidence very persuasive. Lilah's process log also highlighted the way her life entwined with her academic thinking. For example, going out with her husband to a Mexican restaurant in a Mexican neighborhood played into her papers on Cinco de Mayo and her reflections on the way local celebrations had come to be more about tacos than the battle of Puebla.

Prior (1998) and Prior and Shipka (2003) articulate the blend of Bakhtin's (1981) account of chronotopes (represented and embodied) with Goffman's (1981) account of the lamination of discourse. Once we take on board this understanding of activity as situated, mediated, and dispersed, fundamentally dialogic and heterochronic, then it follows that any cultural act or object is layered with multiple histories – with a fuzzy, emergent set of affordances for meaning and action – that any activity can only be a *laminated assemblage* (Prior & Schaffner 2011).

The complex *chronotopic lamination* of Lilah's case (Prior 1997; 1998) led me to seek a broader, more open method to trace chronotopic trajectories of writing (cf. Karsten 2011). To that end, Jody Shipka and I (Prior 2004; Prior & Shipka 2003; Shipka 2011; Shipka & Chewning 2007) designed a process drawing and interview protocol that asked undergraduates, graduate students, and professors to draw representations of their work on some particular text or project. Brent Johnson, an undergraduate college student in kinesiology, focused on an assignment for Jody's first-year composition course. The assignment was to write an educational autobiography; he wrote about how much he had learned from watching movies and how watching movies indexed the important times and relationships
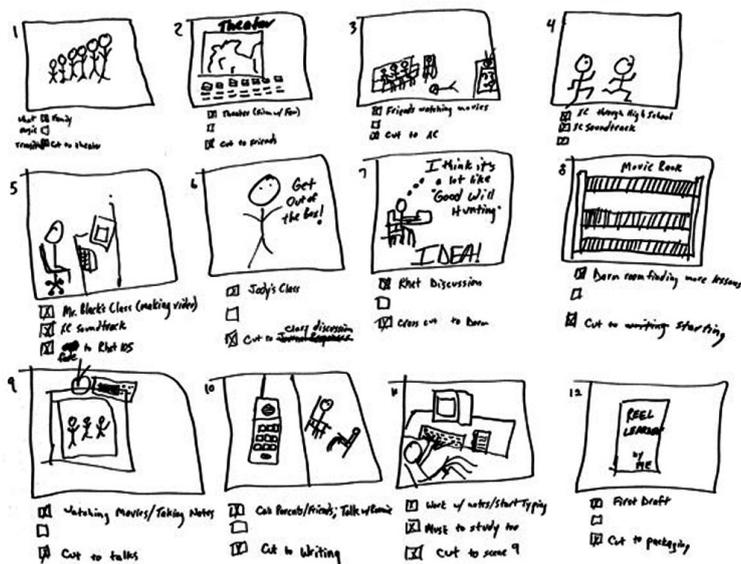

Figure 1. First twelve (of 18) panels from Brent Johnson's drawing of the writing process for one of his first-year composition projects

of his life. His paper, "Reel Learning" (playing with the homonyms "real" and film projector "reel"), argued that his most important lessons had almost always happened outside, or in non-core parts, of school.

Brent drew his overall process as storyboards, not a form we had invited, but one he chose because of his interest in film. Figure 1 displays the first 12 of the 18 panels he drew. What is striking is that we do not see Jody's class – the class where the autobiography was assigned – until the sixth panel. In panels 1–5, Brent represents being born into his family, going to movie theatres, watching films at home with friends, running races in high school, and taking a high school class where he learned to shoot and edit videos. Brent's temporal framing of his process brilliantly captures that writing processes are chronotopically dispersed, not bounded, that the experiences of our lives can always be recruited into composing.

In another interview, Michelle Kazmer, who was doing a Ph.D. in Library and Information Science (and is now an associate professor in the School of Information at Florida State University), drew the process of writing her dissertation proposal. In the center of the page there was a sequence of three images (see Figure 2). Michelle indicated that the first image on the left represented her in bed, very demoralized about writing at this stage and feeling she should just give up on the PhD. She then narrated the next two images, focusing on a critical conversation she and her boyfriend ("curly-haired person") had at the Esquire, a local bar:

> This is the Esquire – it's kind of hard to tell because it's a little table and that's-they're pints there at the Esquire and this is us talking back and forth. And, um, finally I said, "I have not been here for three and a half years to walk out of here without a degree. This is stupid! You know, I just, I can't sit in front of the computer and just go, 'heeh, heeh, heeh.' You know? I just have to write something. It doesn't have to be the best dissertation proposal ever, it just has to be good enough to pass and, you know, that's what people are always telling me, right? So, curly-haired person and the beers and the Esquire and, you know, I finally get to that point where it's like, "NO!" Okay, fine. There's typing on the screen now.

The image of the conversation at the bar is followed by an exclamation mark and the third image of Michelle typing away at the screen.

Figure 2. Extract from Michelle Kazmer's drawing of the process of writing her dissertation proposal, images 9-11 of 16 on the page

We designed the drawing protocol to capture the contours (the shape) of complex chronotopic trajectories of composing. When we think about writing processes, I propose that we be sure to seek accounts that include things like Kazmer at the bar (Figure 2) or Johnson watching movies and talking to his mom on the phone (panel 10 in Figure 1) as well as fast shifts between planning and reviewing during inscriptional events, that we seek, in other words, literate activity – chains of reading and inscribing, talking and observing, acting and making involving a temporally and spatially dispersed set of scenes and cast of characters.

4.   Semiotic remediation and the problem of the text

Writing as text-person-activity-mediation-society involves multiple semiotic resources (not just linguistic signs) and multiple semiotic traversals in the process. The most basic traversal is the constant movement between external and inner (cognitive and embodied) semiotics. Vygotsky (1987) noted that externalization (e.g. speech) involves transformations. Converting the internal semiotics of sense into the externalized signs of language and image involves what Hutchins (1995) has called the propagation of representations across media (human minds and bodies, tools, inscriptions, etc.), each of which have particular affordances. Voloshinov (1973) evoked these semiotic transformations rather poetically:

> The process of speech, broadly understood as the process of inner and outer verbal life, goes on continuously. It knows neither beginning nor end. The outwardly actualized utterance is an island arising from the boundless sea of inner speech, the dimensions and forms of the island are determined by the particular situation of the utterance and its audience.   (Voloshinov 1973, 96)

Prior, Hengst, Roozen, and Shipka (2006; see also Prior & Hengst 2010) have proposed *semiotic remediation* as a unit of analysis, arguing for a dialogic approach to all semiotic practices-in-the-world, to "the diverse ways that humans' and nonhumans' semiotic performances (historical or imagined) are re-represented and reused across modes, media, and chains of activity" (Prior et al. 2006, 734). This dialogic perspective on all sign systems (gestures and embodied activity as well as language and visual symbols or designs) is informed by diverse disciplinary discussions of mediation, recognizability, repurposing, and affordances of media. *Remediation* points to ways that all activity is (re)mediated – not mediated anew in each act – taking up the materials at hand (materials with a history), putting them to present use, and thereby producing altered conditions for future action. It also highlights that all activity is composed of multiple semiotic materials. As a unit of analysis, semiotic remediation extends the cognitive image of unlimited semiosis into a practice- and material-oriented perspective: it is designed to disrupt static notions of isolated products and bounded processes, beginning instead with distributed activity and cognition as givens.

Since Emig's (1971) seminal work on composing, writing studies – particularly cognitive and dynamics research into composing processes (e.g. Flower & Hayes, 1981; Olive, Alves & Castro 2009; Perrin 2014) – has viewed writing as the moment-to-moment production of texts, just as speaking is the moment-to-moment production of talk. However, the problem of the text and its sense remains. No utterance (written, oral, multimodal) can achieve its sense and function in a moment. Its relevance, production, interpretation, and use all require attention to the histories that lead to it, the unfolding events of its use, the imagined projections of its future, and ultimately the way it is in fact understood, taken up, replayed and reused in near and perhaps more distant futures. If production of written utterances is equated to production of spoken utterances, how do we understand texts that emerge out of long histories of production, texts that are composed and often lengthy? Such utterances (e.g. Tolstoy's *War and Peace*) not only have a history, as even a simple "Salut, Marie, ça va?" must have a history; they have a history of focused composition.

I have argued (Prior 2009) that we reject Bakhtin's (1986) problematic account equating written texts to oral language through the notion of turn-taking and instead consider the nature of *composed utterances*, utterances that call on us to analyze the chains of utterances that are woven together; the various ways that the composed document/performance overtly or covertly indexes its specific history of composition; and the ways that production, reception, and use take this history into account. Of course, the notion of composed utterance also applies to much talk. For example, Irvine (1996) analyzes how insult poetry for Wolof wedding ceremonies is co-composed prior to the event by sponsors, others in the community and a *griot* (a low-ranking female bard); how the *griot* delivers and leads communal repetition of the insults during the event; and how what Irvine calls *shadow conversations* (those conversations that are not here-and-now but are felt here-and-now) are critical to the cultural production, uptake, interpretation, and reuse of the insults. Political speeches, film and stage drama, religious ceremonies, sales pitches, language drills, sermons – a lot of talk fits into the category of composed utterance, often with texts woven significantly into the history. Roozen's analysis (Prior, Hengst, Roozen & Shipka 2006) of semiotic remediation in the trajectories of an amateur comedy skit offers a detailed glimpse into the complexity of composed performances, tracing the way compositional events brought together multiple people who co-composed the text/performance in interaction. Such composed performances index not simply some authorial vision, but also the emergent, interpretive work of the actors animating their lines and of the director, stage crew, audiences, and others shaping the contexts of the performed, embodied skit.

To better understand situated semiotic practices and the role of new technologies in literate activity, I began a study of a university Art and Design group as they

revised and redesigned an interactive, web-based art object called *IO*, an interactive website that mixed words, sounds, and images. The study focused on the work of two professors, Joseph Squier and Nan Goggin, and two of their graduate student research assistants, Tony and Eunah. The data included video and audio recordings of 14 group meetings over 11 months (each lasting at least an hour); several video recordings of individuals working on the project outside of group meetings; interviews with Joseph, Nan, Tony and Eunah; and a collection of texts, including screen captures and electronic files. The *IO* group continued to work on the redesign for at least a year after my data collection ended.

Figure 3 displays a text that emerged as a key mediator of the work of the group in its first year and illustrates how attention to chronotopic trajectories illuminates the folding of activity into composed utterances. The text is fairly simple, consisting of printed text (words, numbers, and lines) overlain with handwritten annotation (lines and numbers, mostly paired sequences separated by commas and representing Cartesian coordinates for the screen interface). That simplicity belies the complex origins and use of this text.
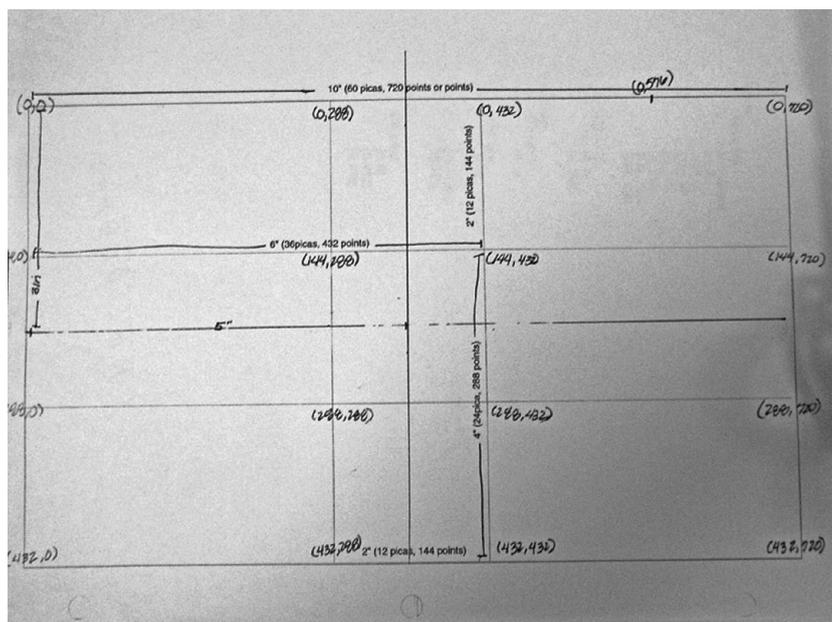

Figure 3. A printed, handwritten, drawn representation of the IO interface

Prior (2010) offers a detailed analysis of processes and artifacts in this group's work that I summarize briefly here to explore the semiotic remediation around this text. The document in Figure 3 represents a template for the screen interface

of *IO*, a template used to direct placement of images and text from the database as they loaded onto the screen. For example, an image that measured 6 inches by 4 inches might be loaded onto the screen with its top left corner at the 144, 0 coordinates (an option highlighted by the handwritten lines starting at 144,0 on the document). A 2-inch by 4-inch image could be loaded into the six boxes along either the left or right sides of the interface. The central 2-inch column would become the default space for loading text.

At an early meeting in February 2001, Joseph, Nan, Christian (a dance professor who worked briefly on the project), and Tony were planning the digital infrastructure for the revised *IO*, assessing how they could integrate a set of new programs (PHP, an open source database program; MySQL, an open source inquiry language program; and Flash 5, a proprietary multimedia application that would be the front end of *IO*). They needed to make critical decisions about how to structure the database, access it, and present images and words on the screen. A videotape of the meeting in the studio displayed an important sequence of interactions at the whiteboard as Nan and Tony, who took the lead in programming the database, talked about how to store and call up the images. The drawing/text on the whiteboard on the left of Figure 4 represents the Flash template for a revised *IO* screen interface. It was drawn, written, talked, and gestured into existence in less than three minutes of somewhat muddled interaction. The final representation that day involved at least 29 separate actions that touched the surface of the whiteboard, movements made by Nan and Tony, who used two different colored markers to sketch the boxes (Nan), revise the boxes (Nan), add letters (Tony), and then add arrows from letters into the diagram (Nan). The drawing and discussion of this emerging template was being coordinated with references to a PHP data entry screen on a laptop on the table, with another drawing on the whiteboard (representing a database architecture), and with many gestures in the air and over the diagram. As my fuller analysis (Prior 2010) details, Nan and Tony's interactions over this box diagram

Figure 4. A video screen capture (left side) of Nan drawing a line on the whiteboard in February as Tony watches and of Tony in November (right side) placing the document from Figure 3 over the computer screen as Nan and Eunah watch

involved a lot of miscommunication as Nan wanted Tony to be thinking about how the database related to the design of future *IO* screens whereas Tony was focused on another issue, how users could contribute images to *IO* to achieve what Joseph imagined as an intelligent art object that would learn. Inscription at the whiteboard emerged through sequential, co-present interaction, inscription as embodied and social activity rather than as artifact (although inscriptions became artifacts-in-interaction and could have a longer duration, as a version of the diagram did in this case).

I first saw the paper version of the whiteboard inscription in a September meeting seven months later. Tony, Eunah, Nan and Joseph were re-starting the project after a summer lull. Nan raised again the question of how Tony had set up image types (and their sizes) in the database. She then referred to "the grid" while making a box gesture with both hands that closely echoed a gesture she had made over the diagram on the whiteboard in February. She then stood up and came back with the document shown in Figure 3. For the next few minutes, she and Tony gestured on and over the paper and the screen as they discussed how many image types needed to be set up in the database.

This piece of paper (and other pencil drawings of interface elements) came out again in interactions over the next three months. In a November meeting, as Tony, Nan, and Eunah were assessing a new screen interface design, they discussed a problem with where the images were loading. Tony held the paper up to the screen (see image on the right side of Figure 4) to clarify the problem, which involved how a dark border Eunah had added around the interface meant that they needed to mathematically adjust the Cartesian coordinates for loading images to the screen.

Talking about this document in a December interview, Nan noted in passing that the dimensions of the revised paper template (which were noticeably not like the nearly identical squares drawn on the whiteboard in February) were designed to align with the mystical golden section proportions that Pythagoras had identified and that have since been popular in Western art and architecture (see Elam 2001). In sum, the textual product in Figure 3 represents a laminated enfolding of multiple historical trajectories. Its sense and functions were rooted in not only the long history of golden section proportionality in the West, but also in the whiteboard interactions of the group in February. It was drawn in shadow-conversation anticipation of its being brought out and used to coordinate talk and action as it was in the September discussion of the database and the November problem solving at the screen. In short, the linguistic and visual structure of this document (and other texts) can only be unpacked and understood in relation to its chronotopic trajectories and their semiotic remediation.

5.   Conclusion

Writing research needs methods that can drill down into moment-to-moment text production; that can illuminate the linguistic and semiotic affordances of a material inscription but also its webs of intertextual and interdiscursive affordances (e.g. by use of text genetic methods, Grésillon and Perrin this volume); and that can locate the text, its production, reception and use in chronotopic trajectories that give it meaningful and functional affordances. For full accounts of writing, larger units of analysis like literate activity and semiotic remediation are, thus, necessary. Working from the perspective of sociocultural and semiotic practice, Agha (2007) defines language use rather radically as "events of semiosis in which language occurs" (Agha 2007, 6). Paraphrasing Agha, I would argue that literate activity must be understood as *events of semiosis in which writing is implicated*. I say "implicated" here because of the kind of complex temporalities the drawing protocol makes visible, because Kazmer at the bar and Johnson watching movies in high school are events in which writing does not "occur" but in which writing is implicated.

Working in the institutional space of the newsroom, Perrin (2014; Grésillon and Perrin this volume) has been able to capture an impressively deep picture of composing processes; however, his analysis offers an attenuated image of the chronotopic trajectories, the other times and spaces outside of that institution, that shape the journalists' actions. In contrast, Roozen's (2009) account of the rich intersections of Kate's writing for school with her textual and multimedia production (online and off) of fan fiction captures an impressive image of the heterogeneous trajectories of Kate's literate and semiotic activity; however, Roozen only gestures at Kate's moment-to-moment practices of writing (and the same can be said for the drawing protocol research I have done).

Accounts of writing can be built by combining the findings and frameworks of different research projects, but I also believe that attention to chronotopic trajectories and closely situated evidence of episodes of focal composition can be put together in one study. Bowen's (2011) research on the digital literate activity and lives of elderly individuals; Fraiberg's (2010) research on multilingual, multimodal, and cultural practices in and around a high-tech, start-up company in Israel; Sheridan's (2008) accounts of literate activity around a feminist youth project in the community; and Berry, Hawisher, and Selfe's (2012) examination of transnational literate lives, all illustrate designs that move between close attention to composing and wider, more open-ended chronotopic trajectories.

Olinger's (2014) sociocultural research on disciplinary writing styles suggests a key implication of this perspective for teaching and practice. As she argues, we

need to work from a deep recognition of the situatedness and heterogeneity of disciplinary genres, to reimagine writing styles more as tropes of identity and stance rather than as hard-and-fast rules of discourse.

To teach students to engage in such dynamic and emergent practices, Shipka (2011) offers a detailed account of her activity-based, multimodal framework for composing. Asking students to make and defend choices about the whole rhetorical situation (goals, materials, genres, ways of composing, and contexts of reception), Shipka's pedagogy rejects "the highly decontextualized skills and drills, linear, single-mode approach to writing instruction" (Shipka 2011, 85) regardless of whether the mode that is privileged is a traditional paper essay or some multimedia digital form. The detailed examples of her pedagogical designs and students' practices in response to her invitations to act rhetorically are well worth close reading. Shipka (2011) sums up the pedagogy in these terms:

> …when called upon to set their own goals and to structure the production, delivery, and reception of the work they accomplish in the course, students can: (1) demonstrate an enhanced awareness of the affordances they employ in service of those goals; (2) successfully engineer ways of contextualizing, structuring and realizing the production, distribution, delivery and reception of their work; and (3) become better equipped to negotiate the range of communicative contexts they find themselves encountering both in and outside of school.
> (Shipka 2011, 103–4)

Grounded in sociocultural research on writing, literate activity, and semiotic remediation rather than in traditional cultural views of writing processes and products, Shipka's pedagogy is one that emphasizes the development of the *semiotic agility* (Prior 2010) that is, in fact, the hallmark of our literate and semiotic practices in the world.

References

Agha, Asif. 2007. *Language and Social Relations*. Cambridge: Cambridge University Press. DOI: 10.1177/09579265100210050605

Bakhtin, Mikhail. 1981. *The Dialogic Imagination: Four Essays*. Trans. Caryl Emerson and Michael Holquist. Ed. Michael Holquist. Austin, TX: University of Texas Press. DOI: 10.2307/40136453

Bakhtin, Mikhail. 1986. *Speech Genres and Other Late Essays*. Trans. Vern W. McGee. Austin, TX: University of Texas Press.

Berry, Patrick, Gail Hawisher, and Cynthia Selfe. 2012. *Transnational Literate Lives in Digital Times*. Logan, UT: Computers and Composition Digital Press/Utah State University Press. Available at http://ccdigitalpress.org/transnational

Bloomaert, Jan. 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press. DOI: 10.1177/09579265110220050605

Bowen, Lauren. 2011. "Resisting Age Bias in Digital Literacy Research." *College Composition and Communication* 62(4): 586–607.

Cole, Michael, and Yrjo Engeström. 1993. "A Cultural-Historical Approach to Distributed Cognition." In *Distributed Cognitions: Psychological and Educational Considerations*, ed. by Gavriel Salomon, 111–138. New York: Cambridge University Press.

Eco, Umberto. 1976. *A Theory of Semiotics*. Bloomington, IN: Indiana University Press. DOI: 10.1017/s004740450000484x

Elam, Kimberly. 2001. *Geometry of Design: Studies in Proportion and Composition*. New York: Princeton Architectural Press. DOI: 10.1162/leon.2003.36.5.413

Emig, Janet. 1971. *The Composing Processes of Twelfth Graders*. Urbana, IL: National Council of Teachers of English. DOI: 10.2307/814097

Flower, Linda, and John Hayes. 1981. "A Cognitive Process Theory of Writing." *College Composition and Communication* 32(4): 365–387. DOI: 10.2307/356600

Fraiberg, Steven. 2010. "Composition 2.0: Toward a Multilingual and Multimodal Framework." *College Composition and Communication* 62(1): 100–126.

Galbraith, David, and Gert Rijlaarsdam. 1999. "Effective Strategies for the Teaching and Learning of Writing." *Learning & Instruction* 1999(9): 93–108. DOI: 10.1016/s0959-4752(98)00039–5

Goffman, Erving. 1981. *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press. DOI: 10.1017/s0047404500009623

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge: MIT Press.

Irvine, Judith. 1996. "Shadow Conversations." In *Natural Histories of Discourse*, ed. by Michael Silverstein and Greg Urban, 131–159. Chicago, IL: Chicago University Press. DOI: 10.1017/s0047404598224043

Karsten, Andrea. 2011. "Chronotopes in Writing: Excerpts from a Case Study." *Tätigkeitstheorie – Journal für tätigkeitstheoretische Forschung in Deutschland (Activity Theory – Journal of Activity-Theoretical Research in Germany)* 5: 87–120.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor Network Theory*. Oxford: Oxford University Press. DOI: 10.1108/eoi.2008.27.3.307.2

Lemke, Jay. 2000. "Across the Scales of Time: Artifacts, Activities, and Meanings in Ecosocial Systems." *Mind, Culture, and Activity* 7(4): 273–290. DOI: 10.1207/s15327884mca0704_03

Nystrand, Martin. 2005. "The Social and Historical Context of Writing Research." In *Handbook of Writing Research*, ed. by Charles MacArthur, Steve Graham, and Jill Fitzgerald, 11–27. New York: Guilford Press.

Olinger, Andrea. 2014. "On the Instability of Disciplinary Style: Common and Conflicting Metaphors and Practices in Text, Talk, and Gesture." *Research in the Teaching of English* 48(4): 453–478.

Olive, Thierry, Rui Alexandre Alves, and São Luís Castro. 2009. "Cognitive Processes in Writing during Pause and Execution Periods." *European Journal of Cognitive Psychology* 21(5): 758–785. DOI: 10.1080/09541440802079850

Peirce, Charles. 1998. *The Essential Peirce: Selected Philosophical Writings, Volume 2 (1893–1913)*. Ed. The Peirce Edition Project. Bloomington, IN: Indiana University Press.

Perrin, Daniel. 2014. *The Linguistics of Newswriting*. Amsterdam: John Benjamins. DOI: 10.1075/aals.11

Prior, Paul. 1991. "Contextualizing Writing and Response in a Graduate Seminar." *Written Communication* 8(3): 267–310. DOI: 10.1177/0741088391008003001

Prior, Paul. 1995. "Tracing Authoritative and Internally Persuasive Discourses: A Case study of Response, Revision, and Disciplinary Enculturation." *Research in the Teaching of English* 29(3): 288–325.

Prior, Paul. 1997. "Literate activity and Disciplinarity: The Heterogeneous (Re)production of American Studies around a Graduate Seminar." *Mind, Culture, and Activity* 4(4): 275–295. DOI: 10.1207/s15327884mca0404_5

Prior, Paul. 1998. *Writing/Disciplinarity: A Sociohistoric Account of Literate Activity in the Academy*. Mahwah, N.J.: Erlbaum. DOI: 10.4324/9780203810651

Prior, Paul. 2004. "Tracing Process: How Texts Come into Being." In *What Writing Does and How It Does It: An Introduction to Analysis of Texts and Textual Practices*, ed. by Charles Bazerman, and Paul Prior, 167–200. Mahwah, NJ: Lawrence Erlbaum.

Prior, Paul. 2006. "A Sociocultural Theory of Writing." In *The Handbook of Writing Research*, ed. by Charles A. MacArthur, Steve Graham, and Jill Fitzgerald, 54–66. New York: Guilford Press. DOI: 10.1111/j.1467-873x.2008.00423.x

Prior, Paul. 2009. "From Speech Genres to Mediated Multimodal Genre Systems: Bakhtin, Voloshinov, and the Question of Writing." In *Genre in a Changing World*, ed. by Charles Bazerman, Adair Bonini, and Deborah Figueredo, 17–34. Fort Collins, CO: WAC Clearinghouse and Parlor Press. Available at: http://wac.colostate.edu/books/genre/

Prior, Paul. 2010. "Remaking IO: Semiotic Remediation in the Design Process." In *Exploring Semiotic Remediation as Discourse Practice*, ed. by Paul Prior, and Julie Hengst, 206–234. Houndmills, UK: Palgrave MacMillan. DOI: 10.1057/9780230250628

Prior, Paul, and Julie Hengst. 2010. "Introduction: Exploring Semiotic Remediation." In *Exploring Semiotic Remediation as Discourse Practice*, ed. by Paul Prior, and Julie Hengst, 1–23. Houndsmill, UK: Palgrave MacMillan. DOI: 10.1057/9780230250628

Prior, Paul, Julie Hengst, Kevin Roozen, and Jody Shipka. 2006. "'I'll Be the Sun': From Reported Speech to Semiotic Remediation Practices." *Text and Talk* 26(6): 733–766. DOI: 10.1515/text.2006.030

Prior, Paul, and Karen Lunsford. 2007. "History of Reflection and Research on Writing." In *Handbook of Research on Writing: History, Society, School, Individual, Text*, ed. by Charles Bazerman, 81–96. New York: Taylor & Francis.

Prior, Paul, and Spencer Schaffner. 2011. "Bird Identification as a Family of Activities: Motives, Mediating Artifacts, and Laminated Assemblages." *Ethos: Journal of the Society for Psychological Anthropology* 39(1): 51–70. DOI: 10.1111/j.1548-1352.2010.01170.x

Prior, Paul, and Jody Shipka. 2003. "Chronotopic Lamination: Tracing the Contours of Literate Activity." In *Writing Selves, Writing Societies: Research from Activity Perspectives*, ed. by Charles Bazerman, and David Russell, 180–238. Available at the WAC Clearinghouse http://wac.colostate.edu/books/selves_society

Prior, Paul, and Steven Thorne. 2014. "Research Paradigms: Beyond Product, Process, and Social Activity." In *Handbook of Writing and Text Production*, ed. by Eva-Maria Jakobs, and Daniel Perrin, 31–54. Berlin: Walter de Gruyter. DOI: 10.1515/9783110220674.31

Roozen, Kevin. 2009. "Fan Fic-ing" English Studies: A Case Study Exploring the Interplay of Vernacular Literacies and Disciplinary Engagement." *Research in the Teaching of English* 44(2): 136–169.

Scribner, Sylvia. 1997. *Mind and Social Practice*. Cambridge: Cambridge University Press. DOI: 10.1002/(sici)1099-0720(199802)12:1%3C93::aid-acp498%3E3.3.co;2-k

Sheridan(-Rabideau), Mary. 2008. *Girls, Feminism, and Grassroots Literacy: Activism in the Girl Zone*. Albany, NY: SUNY Press. DOI: 10.1177/1750481309337217

Shipka, Jody. 2011. *Toward a Composition Made Whole.* Pittsburgh, PA: University of Pittsburgh Press.

Shipka, Jody and Bill Chewning. 2007. "Live Composition: Four Variations of a Telling." In *Re-Situating and Re-Mediating the Canons: A Cultural-Historical Remapping of Rhetorical Activity*, ed. by Paul Prior, Janine Solberg, Patrick Berry, Hannah Bellwoar, Bill Chewning, Karen Lunsford, Liz Rohan, Kevin Roozen, Mary Sheridan, Jody Shipka, Derek Van Ittersum, and Joyce Walker. *Kairos: A Journal of Rhetoric, Technology, and Pedagogy* 11.3. Available at http://kairos.technorhetoric.net/11.3/topoi/prior-et-al/index.html

Smith, Dorothy. 1974. "The Social Construction of Documentary Reality." *Sociological Inquiry* 44(4): 257–268. DOI: 10.1111/j.1475-682x.1974.tb01159.x

Voloshinov, Valentin. 1973. *Marxism and the Philosophy of Language*. Trans. Ladislav Matejka and Irwin Titunik. Cambridge, MA: Harvard University Press.
DOI: 10.1017/s0047404500008071

Vygotksy, Lev. 1987. *Problems of General Psychology: The Collected Works of L.S. Vygotsky, Volume 1*. Trans. Norris Minick. New York: Plenum. DOI: 10.1017/s0142716400008341

Wertsch, James. 1991. *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge: Harvard University Press. DOI: 10.1177/092137409100400104

Witte. Stephen. 1992. "Context, Text, Intertext: Toward a Constructivist Semiotic of Writing." *Written Communication* 9(2): 237–308**.** DOI: 10.1177/0741088392009002003

Writing acts and writing performativity

Producing and disseminating leaflets

Fanny Delbreilh

EHESS Paris *CentreAnthropologie de l'écriture* / Anthropology of Writing Center
Teaching assistant – Université Sorbonne nouvelle Paris 3

This chapter is about public writings which pertain to collective action, common objects of a protest graphic culture: leaflets. From a pragmatic viewpoint, I try to analyze the way these written objects operate.

In order to do this, I aim to explore the notion of "writing act" developed by Béatrice Fraenkel. Through a contemporary case study, mainly based on an ethnographic fieldwork on a French feminist protest march in 2011 and on casual sources about protest practices, I try to scrutinize the writing act related to leaflets and to understand the specificities of written performativity. More precisely, I aim to point out how taking into account the material aspect of these situated writings, the handling gestures and bodily commitment allows to show how the stages of production and dissemination of the leaflets are an integral part of a writing act related to these writings, and how their performativity is intimately related to their materiality.

Therefore, first I show how the stage of material production of the leaflets is a collective and sustained activity, and a collective enunciation. Then, situated observations and photographic enquiry point out how the actors make use of real skills when handing out leaflets. Subsequently, I try to understand to what extent this gesture of handing out leaflets is part of the performative written enunciation and I aim to show how getting the addressee(s) to take the leaflets is a fundamental stage in the performativity of these distributed writings. Finally, by assigning the leaflets' utterances to the precise situations in which they are handed out, I try to point out how the writing acts of leafleting have different meanings, values and stakes in diverse situated actions.

Keywords: writing act; written performativity; materiality of writing; leaflets; distributed writings; protest writings; protest march

1. Introduction

1.1 Studying leaflets

Many writings of different kinds (posters, banners, placards, stickers, graffiti, pamphlets…) take an essential part in collective, political and protest action; and writing practices are a crucial element of what historian Charles Tilly has called the « repertoire of collective action » (Tilly 1986, 541; Tilly 2008).[1] Moreover, protest writings and the uses of writing can be regarded as specific objects of anthropological or historical study in line with major works in history about writing and reading practices.[2]

Leaflets are part of a protest – and activist – graphic culture, in which they seem to be common objects, in two senses: ordinary, mundane and well-known writings on the one hand; writings shared by very different groups, political and social movements, on the other hand. Furthermore, they are used in diverse situations: from election or information campaigns to gatherings, protest marches, sit-ins, strikes, as well as more occasional interventions; from lawful and institutionalized situations to clandestine and dangerous ones, depending on countries and times. Finally, the ways of disseminating leaflets also vary: leaflets can usually be handed out in public spaces, but also delivered in mailboxes, left on windshields or other available places; they can be posted up on walls and different surfaces, thrown, dropped, and so on.

My dissertation-in-progress is about these common, versatile and barely studied public writings, especially about leaflets produced in France during the 1950s and 1960s. In this work, I scrutinize from various angles the way these written objects operate: how can one act with these writings? How can these writings act? In order to achieve this, my research focuses on three points: (i) first, a pragmatic analysis of a corpus of leaflets from archival collections; (ii) secondly production and dissemination practices through case studies; (iii) and finally ideas about these writings' capacity of action over time through the testimonies of several actors. In doing so, I hope to examine the relevance of the notion of writing act(s) related to leaflets.

In this paper, I focus on the notion of writing act mainly through a contemporary case study.

---

1. For further discussions on this notion see Tarrow (2010) and Offerlé (2008).

2. These main works are historian Roger Chartier's ones in modern history (Chartier 1994; 1995, and 1998 for the main works in English). As Barton and Papen point out, "historical studies are prominent within francophone research on writing" and "the work of historians has had greater influence on studies of contemporary practices than is the case in the Anglophone world" (Barton & Papen 2010, 15). Protest writings are particularly studied in contemporary historical works of Philippe Artières (Artières & Rodak 2008; Artières 2013).

1.2   The notion of writing act

Béatrice Fraenkel derives the notion of "writing act" from that of "speech act" proposed by philosopher of language John Austin (Austin 1975). She has developed the notion in three main works: two are in French (Fraenkel 2006; 2007) and one is in English (Fraenkel 2010).[3] For this paper, I will rely on both the English and French texts.

In her 2006 article, Fraenkel clarifies the paradoxical status of the writing act in Austin's view. She shows that although he considers them as model of acts, at the same time, he reduces their written characteristics to orality. Besides, Béatrice Fraenkel studies features of a writing act model based on written legal acts, following the approach of Austin who is constantly nourished by legal references, as she points out (Fraenkel 2006, 78). In conclusion, she proposes to free oneself from this legal model and to identify "ordinary acts" (Fraenkel 2006, 89). In order to achieve this, first, she follows Austin's analytical choice, i.e. starting with ordinary verbs referring to day-to-day acts: "*copier, enregistrer, signer, étiqueter, afficher etc.*", namely "copying, registering, signing, labelling, posting up" (Fraenkel 2006, 90). Secondly, she proposes that:

> when we do them [ordinary acts], we are doing three things simultaneously: making an artifact (a copy, a register, a label, a signature, a poster), producing an utterance and doing an act that changes the course of big and small things.
> (Fraenkel 2006, 90)[4]

Thirdly, she invites us to pay close attention first to "written objects, polygraphical and falling within chains of writing" (Fraenkel 2006, 90).

Therefore, I would like to raise the question of whether there is an ordinary writing act related to leaflets: one that would be, for instance, different from putting up. And if so, how can we describe it? What are its features?

This leaves us facing straight away certain difficulties concerning the possible verb referring to this act. First, if in French a leaflet is named *un tract*, the verb *tracter* is a neologism used by activists and collective movements, but not necessarily by a majority of French speakers. In English, the verb *to leaflet* and the form *leafleting* exist but are not particularly common. Secondly, as for the verb *posting up* (or *afficher* in French), *leafleting* as much as *tracter* is indeed making a leaflet; however the meaning of the two verbs insists on the diffusion or publication's stages.

---

3.   She also uses it in a chapter of a book in English (Fraenkel 2011). Unless noted otherwise, I translate to English the quotations from French Fraenkel's papers (more particularly 2006 and 2007).

4.   In French: "en les [des actes courants] effectuant on se livre à trois choses en même temps: fabriquer un artefact (une copie, un registre, une étiquette, une signature, une affiche), produire un énoncé et poser un acte qui modifie le cours des choses, petites et grandes".

This leads to another question: is the stage of material production of the artifact known as "leaflet" (composing, laying-out, printing) outside of the writing act known as *leafleting?* Yet, the utterance is "produced" during this preparatory stage. It is possible to discern here one of the main characteristics of written acts and written performativity: a totally different temporality from that of speech acts. Two specific legal cases cited by Fraenkel can be useful to understand this better. Firstly, the act of making a will (*tester* in French) raises the question: when is the scene of the testamentary act performed? When the dying person is dictating their will and signing it, or when the will is being read? (Fraenkel 2006, 75). Similarly, one can think, in French law, of the procedure of writs being served by bailiffs' clerks (named *signifier* in French), i.e. delivering a missive from the court to the addressee in person. As Collard shows (Collard 2010) and Fraenkel sums up: "Between the moment of the act making and that of its being delivered to the addressee in person, who can say when the performative act occurs?"(Fraenkel 2006, 84).[5] We could, likewise, simply ask: when is the writing act of *leafleting* performed?

Identifying a writing act related to leaflets, naming it, understanding its performativity is not that easy. A way to modestly progress in these aims is to accurately describe the making and disseminating stages by taking into account the material aspect of writing.

1.3   Materiality of writing

By the material aspect of writing I do not think here of either a linguistic sign's materiality, nor of a graphical aspect of writing, but of its characteristics as an artifact, an object. As cited before, when carrying out a writing act, one is "making an artifact": registering it is making a register, copying it is making a copy. More generally, we can assume that the material aspect of writing is fundamental, regardless of whether one thinks in terms of pragmatics or not. Writing is also "an artisanal activity":

> We are so familiar with the objects that we write – letters, exercise books, notebooks, messages – that we find it difficult to see writing as a craft skill. However, it is clear that when I write *in* a notebook, for example, not only do I fill it but I also *create it*. I produce it as a written object. The same goes for all our writing activities: we are constantly producing written objects without giving them a thought.   (Fraenkel 2010, 39)

So writings are hybrid artifacts, combining graphical and linguistic characteristics but also material and spatial ones (Fraenkel 2001a and 2001b; Denis & Pontille 2010b).

---

5.   "Entre le moment de fabrication de l'acte et celui de sa remise en mains propres, qui dira où s'opère l'acte performatif?"

Paying such attention to material aspects results in specific methodological choices and analytical shifts. It implies, first, to watchfully and finely look at small scales: at precise producing gestures and activities; at handling gestures and bodily commitments of actors with these objects. We also have to closely analyze the written objects themselves (supports, shapes, sizes, proper materials…). Yet, contrary to some representations of stable artifacts, like all materials and objects, written objects can be fragile, vulnerable and need attention, care and maintenance (Fraenkel 2011, 310–312; Denis & Pontille 2011 and 2010b). We thus have to look closely at maintenance, conservation and care practices or, on the contrary, at erasing or destructing ones. Methodologically, one can find these detailed analyses in both ethnographic inquiries (observing, describing, photographing small and seemingly insignificant gestures or objects), in the work on corpus of writings (handling, touching, measuring, turning over written objects) and in the study of records of writing activities and situations in documents, stories and archives (taking notice of precise endogenous terms and regular naming patterns of actions).

I aim to point out how taking into account the material aspect of situated writing, handling gestures and bodily commitment allows to show how the producing and disseminating stages are an integral part of a writing act related to leaflets, and how the performativity of writing is intimately related to its materiality.

2. Producing leaflets

In this paper, I rely on two types of data. The first kind is extracted from different casual sources (normative or prescriptive texts, records of actions, scenes) about protest practices in the 1960s and 1970s. Secondly, the main data comes from a collective ethnographic fieldwork on a feminist protest march (the feminist protest march, 5th March 2011, for Women's Day), conducted in 2011.[6] The study consisted on the one hand of ethnographical situated observations and a photographic

---

6. This collective study was executed as part of the methodological workshop "Writings of Feminist Protest" organized by Béatrice Fraenkel and Claire Bustarret in 2010–2011 at EHESS Paris. The enquiry was on two marches: a night march organized by *Rage de nuit* collective the 27th November 2010 in Paris; the feminist combined protest march for the Women Day, the 5th March 2011. Workshop's participants all together took part in this enquiry that led to the one-day workshop "*Manifester avec l'écrit* [Protesting with Writing]", the 6th may 2011, where spoke Béatrice Fraenkel, Sophie Pène, Isabelle Bretthauer, Valentina Tomasini, Nora Labo, Arnaud Dubois, Maud Valegeas and myself. Much of this chapter is based on this talk and also on another specific work with Claire Bustarret who particularly studied banners and placards during the fieldwork.

inquiry (photographs of situations where written objects are carried by actors or somehow present) to examine the uses and activities specific to writing practices in protest marches; and on the other hand of the systematic gathering of leaflets on site, to study their material, graphical and linguistic characteristics.

I did not directly observe the production stages of the leaflets for the 5th March 2011 protest march but the systematic description of the material aspects of the corpus of writings and the support of writing scenes from other collective actions can already point out several important characteristics. Indeed, manufacturing these printed sheets of paper is a collective and sustained activity.

2.1 The writing temporalities

Other protest writings are sometimes created on site, just before or during the march. For instance, on the 5th March 2011, Claire Bustarret documented several scenes in which activists were manually writing placards with markers just before the march started or in which a mother was setting up handmade sandwich boards and armbands on her children's bodies as well as her own.[7] Sometimes banners are even completed just before the demonstration starts, like in a record of a *MLF* feminist activist about her first protest march:

> As I speak a little Russian, I was in charge of sticking pre-cut, fancy red satin Cyrillic letters the right way. (…) For the first time, that 8th of March I had the impression of being where I 'had to be'.   (*Génération MLF 1968–2008* 2008, 240 quoted in Fraenkel 2011b, 16; my translation)[8]

On the contrary, leaflets are made before the march. They are made elsewhere, but they generally are *for* the specific march or collective action.

The leaflets gathered during the 5th March protest march are all printed sheets of paper. So, the making of these writings consists of roughly at least three steps: composing the text and laying it out, which could be accomplished simultaneously or not, and printing. Leaflets are often rapidly produced but these stages themselves can be of different durations depending on the case. For instance, concerning the combined leaflets (leaflets produced by several political and social collectives grouping together around common claims) like two made for this joint protest march and studied by Maud Valegeas (see Figures

---

7.   Claire Bustarret talked about banners and placards in the 5th March 2011 protest march in a symposium organized by our research center during the *Writing Research Across Borders Conference*, on 21th February 2014, in Nanterre University.

8.   In French: "Parce que je parlais un peu russe, on m'a chargée de coller dans le bon sens des lettres cyrilliques prédécoupées dans un joli tissu de satin rouge (…) Ce 8 mars pour la première fois j'ai eu l'impression d'être où "je devais être"."

1 and 2), choosing words and appending signatures is usually quite a long and decisive process. Meetings are organized, first versions are composed and sent by e-mail or distributed during the meeting, discussions occur, amendments to the text are made during the meeting or by e-mail, and when there are debates, signatures can be added, removed, added again. For the leaflet made by young organizations for the 5th March (see Figure 2), the process took twelve days.[9]
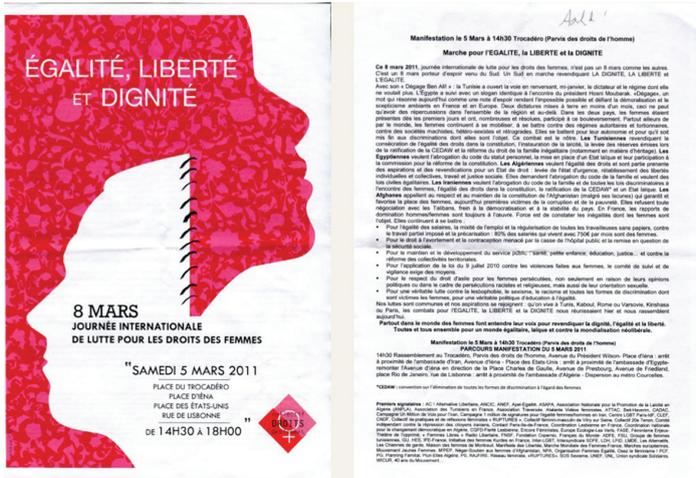

Figure 1. The combined leaflet from the *Collectif Droits des Femmes* [Women's Rights Group. A5 front and back, glazed paper, black, white, red and pink. Scan by the author

As we can see and as Fraenkel proposes, the temporality of writing is not an isolated or an immediate moment: "Writing time is an excessively 'socialized' time, continuous, joined together with other writings and other acts" (Fraenkel 2006, 83).[10] Or, to say it otherwise, the temporal frame of writing is well specified by the notion of "*chaînes d'écriture*" or chains of writing (Fraenkel 2006, 83; Fraenkel 2001b).

This temporal characteristic is, as shown by our examples, intimately linked to another one: "the capacity of incorporating 'several hands', of accumulating them over time while safeguarding a certain unity" (Fraenkel 2006, 83).[11] Indeed, leaflet composition and leaflet making is a collective activity.

---

9. It is Maud Valegeas that finely presented this process of composing and signing the two leaflets reproduced here, during talk on the one-day workshop "*Manifester avec l'écrit* [Protesting with Writing]", the 6th May 2011. I am here only repeating some of her acute analyses.

10. "Le temps de l'écrit est un temps 'socialisé' à l'excès, continu, solidaire de part en part d'autres écrits et d'autres actes".

11. "sa capacité à intégrer 'plusieurs mains', à les cumuler dans le temps tout en sauvegardant une certaine unité".

L'égalité femmes/hommes : un problème réglé ?

Notre génération a grandi avec l'idée que les luttes féministes avaient atteint leurs buts. Pourtant aujourd'hui, même les quelques acquis réels sont remis en cause. Face au retour des idées réactionnaires notamment portées par le Front National et contre l'endurcissement des politiques libérales, une relève féministe est nécessaire ! Des associations et des collectifs se créent un peu partout : la nouvelle génération féministe est en marche !

L'égalité au travail : encore un long chemin à parcourir...

Les écarts de salaires entre les hommes et les femmes et les hommes sont de 27% en moyenne et les femmes occupent plus de 80% des emplois à temps partiel. Ces chiffres trahissent différents problèmes : les femmes sont orientées dès le plus jeune âge, à cause des filières sexuées, vers les métiers les plus précaires et les moins bien payés, elles sont poussées aux temps partiels pour favoriser leur exploitation gratuite dans le cadre domestique (tâches ménagères, éducation des enfants), et elles sont également victimes de nombreuses discriminations : à travail et diplôme égal, elles gagnent toujours 10% de moins que les hommes. Avec la crise et les plans d'austérité, les femmes sont encore plus durement touchées par la précarité aujourd'hui. Et le seront encore plus demain, du fait du recul de l'âge départ à la retraite amoindrissant nettement leurs pensions. La crise actuelle aggrave ces inégalités.

Le droit à disposer de son corps

Sans un réel service public, les droits à l'avortement et à la contraception ne sont que théoriques. Pourtant le gouvernement ferme de nombreux Centres d'Interruption Volontaire de Grossesse, menace de diminuer les subventions du Planning familial et supprime des postes d'infirmières et d'assistantes sociales sur les lycées et les facs. À quoi sert un droit qui ne peut être utilisé ? Le coût de la contraception et de l'avortement rend particulièrement difficile pour les jeunes et les femmes les plus précaires la maîtrise de leur sexualité et le droit à disposer de leur corps.

Pour l'émancipation des femmes

La prise en charge par la société de ces questions n'est pas assumée, et cela est lié à la mise à disposition du corps de femmes. Les violences physiques et psychologiques, le harcèlement et les viols, découlent des systèmes patriarcaux et du maintien de la domination masculine, qui entrave l'émancipation des femmes, qui passe par une destruction des systèmes de genre faisant correspondre à chaque sexe un rôle et une identité. La lutte des femmes rejoint donc la lutte des Lesbiennes, Gays, Bi, Trans et Intersexes.

Face à cette situation, la lutte pour les droits des femmes est plus que jamais d'actualité. Reconstruire un mouvement féministe massif est toujours aussi urgent. Le 8 mars est la journée internationale de lutte pour les droits des femmes. Mobilisons-nous pour l'émancipation de chacun·e, la lutte pour l'égalité entre femmes et hommes et la défense des droits des femmes !

MANIFESTATION Samedi 5 mars 2011 à 14h30
Parvis des droits de l'homme / M°Trocadéro

Collectif féministe Paris 3, Collectif féministe Paris 8, Collectif Tirésias Paris 4, SUD étudiant, NPA Jeunes, MJCF.

UN 8 MARS SOUS LE SIGNE DE LA SOLIDARITE INTERNATIONALE!

Depuis 1910, le 8 mars est la journée internationale de lutte et de manifestations pour les droits des femmes. Cette année, cette journée a lieu dans un contexte de soulèvements révolutionnaires des peuples du monde arabe notamment en Egypte et en Tunisie où les dictateurs, Ben Ali et Moubarak, ont dû quitter le pouvoir! Les féministes de ces pays ont participé pleinement à ces processus d'émancipation. Là-bas comme partout, les schémas de genre doivent être abolis ! Nous exprimons notre soutien et notre solidarité internationale à tou·te·s celles et ceux qui luttent pour leur émancipation et leurs droits.

› A travail égal, salaire égal !
› Avortement et contraceptions libres et gratuits! Non a la fermeture des civg et des plannings !
› Non a la fermeture des crèches publiques !
› Pour une loi cadre contre les violences faites aux femmes !
› Pour le droit à disposer de son corps, moyens de protection pour toutes les sexualités, notamment lesbiennes ! Pilules du lendemain dans les lycees et les facs, centres de planification et crèches publiques dans les universités !

Figure 2. The combined leaflet from several "youth" organizations. A4 back, black and white Scan by Maud Valegeas

2.2 Polygraphy. A collective enunciation

Even when the leaflets do not stem from several organizations grouped together around common claims but from a single group, composition can be a collective activity. An example of a writing scene extracted from an activist newspaper allows us to see this. The scene is from an article out of the Maoist paper *Servir le*

peuple, 1st April 1968, entitled "*Alès, les ouvriers de la céramique en grève* [Alès, Ceramic Workers on Strike]". It took place during the Workers Assembly held just after a protest march in the city. The local trade union's secretary wanted to stop the strike:

> That's when workers get angry, they all agree, *they want to go on, they go on.*

> Throughout the action, the secretary will fight each initiative, each right idea coming from the workers. So, the workers decide *together* to design a leaflet to popularize their struggle, and they get to work. Leaning over a worker's shoulder, the secretary of the local trade union reads what the latter has written: "Yet our claims are quite negligible whereas the boss's profits are tremendous".

> "Tremendous? What do you know about it?"

> Surprised, the worker says everybody knows about it. But the secretary protests.

> "That's impossible. You can't say tremendous. First we don't know anything about it. Second, we'll cut ourselves off from the boss. He can sue us for libel. We have to say 'substantial'. Then the worker wrote: "Situation in Alès is really disastrous", which causes a new fit from the secretary. "Disastrous? It's too strong a word to use here! It is not that disastrous! We must say 'In a phase of economic recession'!" But all the workers point out to him that economic recession is precisely bourgeois language for 'Disastrous workers situation'.

> In the meantime, another worker, who did not let himself get disturbed by all this chitchat, intervenes and reads his leaflet, which is enthusiastically accepted by all the others.[12]

This scene presents a set of actions related to leaflet making: the group decision to "make a leaflet", the "work" of composing and writing, the reading of a temporary version, the debate about the words used, the reading aloud of another proposition, the collective validation. Composing a leaflet is a stake and several actors are part of it – the group of workers, the local trade union's secretary, the first composing worker, and the second one. This writing practice is presented, in this activist newspaper relating a strike, as a collective action: they decide together, compose together, and validate the text together. But we can see how choosing words is the subject of debate and struggle and is part of the strike itself.

Other actors are not mentioned here, although their participation is crucial in the making of leaflets, in this 1960s scene. These are the ones that lay out, and type or print, then reproduce (by duplicating, roneo…) the chosen text. In our collective enquiry about the feminist protest march, these stages are also a

---

12. *Servir le peuple* 1st April 1968, no. 19. *Archives Nationales*. 78 aj/29. I translate.

blind spot but a systematic codicological description of the gathered leaflets can highlight some regularities.[13]

2.3 Printed sheets of paper

Out of our corpus of 30 writings, apart from one, the medium of all these written objects is the unique sheet. Thus, the typographical and codicological unit of leaflets is neither the notebook nor the page, which are the codicological units of books, booklets or pamphlets, but the sheet. Besides, all of the leaflets are printed, which distinguishes them from the tools and writing techniques used for making other protest written objects, like banners and placards, often handmade. Similarly, the formats of the leaflets are standardized ones (mainly in A4, for 16 leaflets, 7 in A5 and 4 in A6), contrary to some "do-it-yourself" formats of placards. (see Figures 1, 2 and 8).

An interesting characteristic is, among all these printed writings, the opposition between the leaflets made by personal or homemade techniques and the leaflets professionally or semi-professionally made. The first ones are, indeed, printed on ordinary paper, common in office supplies, whereas the second ones are on glazed paper, which is variably thick and rigid, and may sometimes bear the references of the professional printer marked in small letters in the top or bottom corners. It can therefore be said that several hands are involved in the making of these leaflets: how can their involvement in the responsability of the utterance be characterized?

Moreover, what we observed by paying close attention to the material aspect of these writings compelled to explore this issue in greater detail. On the one side because at the outcome of this making process, once the leaflet is made, this "product" is not stable. Indeed writings are objects, and being so, they are fragile, vulnerable; they can change and be materially altered. They can fade, fly away in the wind, be burnt, be torn up to shreds, cut out… They need to be maintained and cared for (Fraenkel 2011, 310–312; Denis & Pontille 2011 and 2010b). Thus, once the product is made, the process is still going on. On the other side, as leaflets are sheets, and ones made in multiple copies, this leads to physical and material constraints that require real know-how. This also conduces to consider differently the part of these characteristics in the performativity of these flying sheets.

---

13. Codicology, or science of codex, is the study of books and particularly manuscripts as material objects. Codicology and paleography are essential to history of writing, particularly of medieval and modern Europe, and the study of writing and reading practices (McKenzie 1986; Chartier 1994 and 1995). But these "auxiliary sciences" of history prove to be really useful to study contemporary practices. For more references and analysis about codicology see Fraenkel (2001a, 124) and Mbodj-Pouye (2010, 127 and 139–141).

3.   Handing out leaflets: A technique and know-how

Situated observations and, above all, photographs are a really efficient way to meticulously describe the various aspects of the leafleting technique: body postures, small and big gestures, body positioning in space, movements. This reveals that the actors – demonstrators and activists – make use of real skills when handing out leaflets.

3.1   Storing

Leaflets constitute, indeed, *piles* of sheets. Before the protest march, they need to be stored and transported where the march takes place. During the demonstration, they also need to be stockpiled in a place, accessible throughout all the march, for the demonstrators to come and restock fresh leaflets to be handed out. Some groups have a car at their disposal for this purpose, present all throughout the march, where they stock leaflets, other written objects like banners and other objects used during the demonstration.[14] Other demonstrators can store the leaflets in various containers, often various bags, carried on their bodies (see Figure 3).

Figure 3.   An activist carrying a bag full with leaflets. Paris, 5th March 2011. Photograph by the author

---

14.   We observed the same use of a car during the Night protest march organized by *Rage de nuit* collective, on the 27th November 2010, at night.

3.2 Holding: Tools and bodies

Once the leaflets have been collected, their material characteristics set other con-
straints for the people who hand them out. Indeed, leaflets are numerous and fly-
ing sheets of paper: they need to be both *held*, in order not to fall, fly away or
scatter, and *kept* and *carried,* for the activists to hand out the greatest number
without having to come back to the car or storing places too often, without risking
too long of an interruption of the distribution. To do so, the demonstrators adopt
several tactics. Either they tightly hold the pack of leaflets, often by pressing them
on their body (see Figure 4). They may also install a tool on their bodies where
leaflets are stored and where they regularly draw fistfuls of sheets, like this Parisian
demonstrator, who is visibly experienced, carrying all through the march a big
blue bag around her neck (see Figure 3).

Figure 4. Participants holding packs of leaflets. Paris, 5th March 2011. Photograph by the
author

3.3 Moving

Then the activists move to go and give the leaflets to addressees. By observing their
positioning with regards to that of the cortege, one can notice several scenarios.

First, the people that hand out leaflets can be positioned in a specific space, i.e.
on the sides of the cortege. They are walking in parallel at a different pace and they
devote a large part of their demonstration time exclusively to leafleting. Secondly,
they can be *inside* the cortege, and from time to time come away from it to go and
give leaflets to the people on the sides, then return to the cortege. Finally, they can

be *inside* the cortege but on the edges, close to the passersby, handing leaflets while parading (see Figure 5). Thus, within a protest march, one can observe very parallel and different rhythms of action, and of writing action: a kind of choreography related to writings becomes apparent.

Figure 5.   On the left, the cortege, and several activists leafleting on its sides. Paris, 5th March 2011. Photograph by the author

Therefore handing out leaflets during a protest clearly appears to involve the body, and to use it in accordance with very specific ways, requiring, and constituting at the same time special skills. Leafleting truly belongs to what Stany Grelet calls "the techniques of struggle" and "the technologies of protest" (Grelet 2005, my translation), following in an inspiring way the propositions of anthropologist Marcel Mauss to closely examine the "techniques of the body", and those of his student Haudricourt's to meticulously describe and analyze techniques and technologies (Mauss 1973; Haudricourt & Jean-Brunhes Delamarre 1955; Haudricourt 1987).

Fieldwork also reveals the complexity of the relations between the written objects' carriers and their utterances. The activists who carry banners and brandish placards are not necessarily those who produced the utterances inscribed on them. In the same way, the demonstrators who hand out leaflets are not necessarily those who composed the text of the leaflets: so they are not proper enunciators or speakers. However, to what extent is handing out leaflets part of the performative written enunciation and a form of responsibility for the utterance?

4.   Getting the addressee(s) to take the leaflet(s): The performativity of distributed writings

The observation and description of the precise leafleting gestures and actions shows the importance and the role of the material aspect of the leaflets in their performativity. Indeed leaflets are writings that, during a protest march, the activists properly *hand out*, i.e. give or deliver *by hand* to an individual addressee, and the activists' aim is clearly to make the addressees *take* the leaflets.

4.1   Handing out

To make the addressees take the writings, the activists, once again, make use, often carefully, of specific skills, relying on gestures, moves and oral interactions. First, they move towards the aimed addressee; they extend the hand that holds the leaflet and look at the person whom they are handing it out to; very often, they complete these gestures with initial verbal contact (salutation: "Hello!"; ask: "Do you want to know who we are?"; presentation of the writing: "The leaflet disseminated in Iran"). The aim of these words is to make the addressee take the leaflets, and to potentially engage in a longer conversation about the protest march. Once the addressee takes the leaflet, they leave (see Figure 6).

These efforts are not necessarily successful: the passersby often refuses to take the leaflet, or does not take it straightaway, as in the scene below, related by some French feminist activists in their account of the historic feminist protest march of the 20th November 1971 in Paris in the feminist newspaper *Le torchon brûle* n. 3.

> (…) On the whole our leaflets are well received by women, passers by, storekeepers. A lot of women express their agreement: "The youth is less dumb than we are." In a café, a fifty-year old woman is outraged: "What are they doing, all these men, in the protest? *They* don't need to have an abortion!"
>
> But one woman flatly refuses the leaflet, saying she's not interested.
>
> – But what about the others?
>
> – Let them sort things out!
>
> After five minutes of an increasingly hostile discussion, a comrade says to her:
>
> – And yet, if women don't take matters in their own hands, men won't do it for us.
>
> – *That* is really true that! (Then she took the paper).[15]

---

15.   *Le torchon brûle* 3, February 1972, reproduced in App et al. 2011, 21–27. The extract is page 23–24. I translate.

Figure 6.   An activist is handing out a leaflet. Paris, 5th March 2011. Photograph by the author

As one can see, taking the leaflet, "taking the paper" in this precise situation is not a trifling gesture at all for this woman. It is both a mark of interest or disinterest for the leaflet and maybe for the demonstration ("she's not interested") and, in this case, a mark of agreement with at least an exchange of words, and maybe ideas. The scene also clearly shows the activists' commitment in *one* act of handing out *one* leaflet, and how each addressee counts.

4.2   Giving the greatest number of leaflets. Reaching addressees

Moreover, the demonstrators who hand out leaflets also try to give out the greatest number of them, like the experienced activist we have already seen who systematically hands out leaflets to all passers by positioned on the sides of cortege, making a kind of circle (see Figure 7). Fieldwork observation also shows how activists try to give out the greatest number of leaflets to the greatest number of people, avoiding handing out a leaflet twice to the same person. So, the leaflets are indeed *distributed* writings, i.e. public writings, in multiple copies, circulating, and aimed at reaching an audience, but, specifically, at reaching *individual* or *distinct* addressees in an audience.

Figure 7.   The cortege is at the left of the image (marked by the arrow). The same activist as in Figure 3 is handing out leaflets to people who are watching the march. We are on the *Champs-Elysées*. Paris, 5th March 2011. Photograph by the author

4.3   The performativity of leaflets

Managing not only to give out the leaflet but also to make one take the writing is essential, because that is one of the conditions first, for the leaflet to be *read*, and then for the discourse developed in the leaflet to act. To say it otherwise, the activity of making an addressee take the leaflet is one of the felicity conditions of the acts related to the leaflet, in an analogous (but not similar) way to the conditions to "secure uptake" mentioned by Austin (Austin 1975, 138). Thus, taking

into account the materiality of writing and the situatedness of writing uses leads to comprehendfelicity conditions in a different manner: no longer as "stabilized conventional procedures and principles", but as "concrete situations of accomplishment of performative enunciations" (Denis 2006, 14).[16]

Besides, from the participants' point of view, leafleting during a protest march is not meaningless. Leafleting is a way to mark their social, political, or ideological belonging and membership to a collective, to notify it and to perform it. This is also a way for them to indicate that they *at least* agree with the messages displayed in public spaces by this precise collective. Handing out leaflets and, even, brandishing or just holding them, without having written them, even without having read them, is a form of commitment to these leaflets, a form of responsibility. Frequent situations, both historical and contemporaneous, of verbal, physical and legal clashes during or about leafleting situations remind us of this. This point deserves a more in-depth investigation, though we can bear in mind the fact that leafleting is a public practice, and leaflets are public writings, which are framed and recognized by the Law. In certain times or countries, handing out or even having a leaflet can lead to serious sanctions.

The relationships between the written objects' carriers and the utterances appears thus to be more complex than superficial links. The situated observation I carried out led me to push this issue further. Indeed, observing the gestures, bodies and postures during the protest march highlights to whom and in which space the leaflets are handed out. So, in the final part of this paper, I would like to try and assign utterances to these precise situations. If one of the aims of leaflets and leafleting is to convey messages across public spaces, leafleting can be seen as doing several and different things.

5.   Leafleting: Situated acts, various stakes

5.1   Presenting oneself and explaining the collective action

Most of the gathered leaflets have common discursive characteristics. Generally the name and/or logo of the collective taking charge of the utterance is present. Graphically, it is usually placed at the top of the sheet, in the header, or at the bottom, in the signature, or both (see Figures 1, 2 and 8).

---

16.   I translate: "les conditions de félicité (…) ne sont plus figées sous la forme de procédures et de principes conventionnels stabilisés (…). Elles sont appréhendées en tant que situations concrètes d'accomplissement des énonciations performatives."
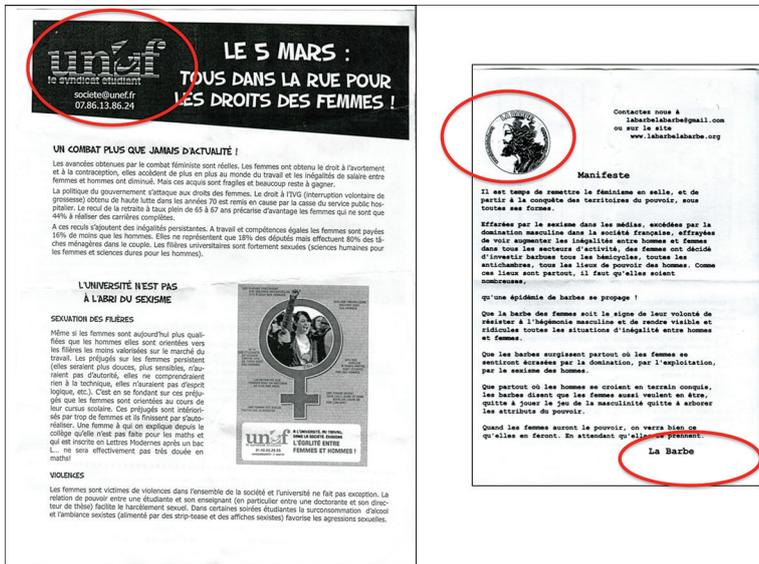
Figure 8. Left: the leaflet from the student trade union *UNEF*. A4, front, black and white. Right: front of the leaflet from *La Barbe*; A5 black and white. Scans by the author

Then, the enunciators do several things, often intertwined: they analyze the situation which protest march is reacting to, they explain the reasons and motives for the collective's participation, and finally, they claim or suggest a number of measures, actions that should be taken. So, when they are handed out to the passersby, leaflets and leafleting seem aimed both at presenting the collective, explaining its positions as well as making sense of the march, either on the spot (see Figure 9), or after the fact, i.e. at giving it a meaning.

Yet fieldwork reveals that a lot of leaflets are handed out *amongst* the demonstrators, and in many ways: leafleting shows contact and also struggles between collectives.

5.2 Making presence and actions known

Leafleting amongst demonstrators occurs at different times. It may happen before the beginning of the march: on the *Parvis des Droits de l'Homme* (or Trocadero esplanade), some activists who hand out leaflets are walking around, among static, more or less shapeless groups of demonstrators, and give out leaflets to the people they are passing by. Leafleting amongst the participants also frequently happens during the march, within the cortege or on its sides (see Figure 6). In both situations, handing out the leaflets seems to have a different aim: showing to other collectives one's effective presence in the protest, and also making one's point of view

Figure 9. Passersby on the *Champs-Elysées*, watching the cortege and reading leaflets. Paris, 5th March 2011. Photograph by the author

known. This partly evokes one of the aspects often highlighted in the research on protest marches, i.e. their self-centered character (Favre 2006).

Most of the leaflets handed out amongst participants are linked to the specific 5th March demonstration and are from collectives which are part of it. But in a different way, some leaflets handed out within the cortege are not related to the motives of the current march. They are announcements of other upcoming collective actions (often gatherings, marches and meetings), not specifically about women or feminist actions, and calls to take part in them. In these situations, leafleting carried out by activists and aimed at participants in the march aims to establish contact with them, to inform them about their actions, but also to potentially recruit participants.

5.3   Displaying exteriority or disagreement. Disrupting the demonstration

Finally, observation reveals some leaflets handed out by groups who do not parade and do not take part in the march. During the 5th March demonstration, two cases got my attention.

First of all, roughly halfway through the demonstration's route, one could see a bus shelter covered with posters of a collective (*Voie prolétarienne partisan*). Posters are put up in order to be seen from the cortege (see Figure 10). Some people brandish placards, looking at the cortege; one brandishes a newspaper; two others hand out leaflets. One is rather static and gives out leaflets to the demonstrators passing nearby; the other takes several steps towards the cortege, and regularly goes

into it and gives out leaflets to demonstrators. When reading the leaflets, one can see that these writings explain the collective's position on the oppression of women. Here, leafleting is clearly aimed at the cortege, and not at passersby. And a real site for leafleting and disseminating the writings is set up, before the cortege's arrival. During the big joint marches and demonstrations, like those on the First of May in France, one can often see these kinds of installations, all along the main cortege, often made by political parties or trade-unions: the activists use the configuration of ordinary public space, and temporarily transform them, with posters, flags, and often tables and chairs, into stands, or points of sale and of dissemination, and also gathering points. But, on the 5th March 2011, even if the protest march was a joint demonstration and a rather institutionalized one, the site described just above was the only example of this kind of installation. The stake for this collective seems to be seen, to make its presence and its positions known, maybe to recruit people, but by materially marking its exteriority and by not taking part in the march.

Figure 10. Activists handing out leaflets and brandishing newspapers. The cortege goes from left to right (see the arrows). Paris, 5th March 2011. Photograph by the author

The second situation is even more puzzling: in a location set about two-thirds of the march's route, a rather large group of people gets settled on a small square, totally static, on the left of the main cortege, but close enough for demonstrators to "crash into" the group. They hold a banner directed toward the cortege, with a slogan handwritten on it, but barely visible from a distance; they are silent and they use a very particular way of handing out or, more exactly, not handing out leaflets. Indeed, in front of the banner, some persons hand leaflets high up, in a rigid bodily

position, without moving, and without giving the leaflets to the demonstrators, like a kind of human display rack. The people who want to have the leaflets *have to* go and grasp them. Besides, the text, soberly printed on white paper, explains the reasons for their disagreement and for their refusal to take part in the march. After initial reactions of surprise and small stops, the wave of demonstrators gradually bypasses the group and resumes walking. So, in this case, the act of leafleting or rather brandishing leaflets and not handing them out, which is part of a larger action of writing, is aimed to *display* – to the demonstrators first and also, albeit not clearly, to the passersby – a refusal to be part of the collective action. It also disrupts spatially, bodily and temporally the protest march, which precisely presents itself as a joint and united collective action.

The material description of leaflets, of the handling gestures, bodies, positioning in space and actions around writings, thus reveals that the protest march does not boil down to an interaction between the demonstrators and a vague « audience », or between the cortege and the passersby. It is also constituted of moments of contact, exchanges, clashes, disagreements and antagonisms between groups. Moreover, at the outcome of this analysis, one can see that, even if the leaflets are made for the event, they do not appear only as writings on or about the event. Indeed, the different writing acts of leafleting, and the specific actions around these writings, contribute to shape the event as well, in all its singularity.

6.   Conclusion

In this study of leaflets and leafleting in a feminist protest march, I have tried to show the relevance and specificities of the notion of writing act. As we have seen, and in line with Fraenkel's propositions, a writing act is characterized by some aspects unique to writing and to written enunciation: a temporality totally different from the instantaneousness of speech acts, i.e. a continuous and distended time; a collective or multiple enunciation, or polygraphy; a production of written objects. Seriously taking into account this last and major feature, by using accurate and detailed observations and descriptions as methodological tools, indeed shows that *leafleting* is not a writing act performed at a given moment by only one enunciative instance. The production (or making) and disseminating stages are both part of a leafleting act. Moreover, the study of the gestures and tactics to hand out and make the addressees take the writings specifically shows how the material characteristics of the leaflets – and their consequences in terms of gestures, bodily commitment, techniques, and maintenance practices – are crucial to understanding how the performativity of these "writing acts" operates and is ensured, but also the performativity of the speech acts.

The notion of writing act is still an exploratory one, but, here, the distinction between writing act and speech act has to be stressed. As Béatrice Fraenkel explains, the writing act is not a twin or a double of the oral speech act but it appears as an act in its own right: "a writing act is added to a speech act and this writing act is not simply an act of scription, because it assigns a specific value to the utterance" (Fraenkel 2007, 103, my translation). A good example to better understand this point is her analysis of the road signs such as 'Dog' or 'Keep off the Grass', carrying on Austin's analysis of these types of utterances. First, these road signs are indeed "warnings", i.e. well known performative and precisely exercitive speech acts. Secondly, these signs are artifacts located in specific places and so they "take on their full performative force only when they are displayed in an appropriate place". But Fraenkel goes further by explaining that:

> (…) they do much more than ensure optimum conditions for the effectiveness of these 'performatives'. They also modify the places where they are found: the house which displays a 'Beware of the Dog' sign becomes a forbidden, protected place, just as the notices 'Keep off the Grass' or 'No Posters' modify the status of the grass or the wall.   (Fraenkel 2010, 38)

The road sign 'Keep off the Grass' is, thus, both a speech act of warning and a writing act of "labelling" (Fraenkel 2010, 38). In this line, analytically distinguishing between the speech acts of the leaflet's utterances and the writing acts of leafleting seems to be a fruitful hypothesis. But, compared to the sign 'Keep off the Grass', the leaflets appear to be more complicated, or at least different, to examine: first because they are discourses and not short and concise utterances, secondly because they are not placed signs – i.e. not exactly stable writings but stabilized ones (Denis & Pontille 2010a; 2011) – but rather circulating writings.

And, precisely, this last point appears to me as a crucial one. Indeed, the starting question of my research about leaflets is about the notion of *écritures exposées* proposed and analyzed by several major historical works in epigraphy (Fraenkel 1994), also called "exhibited writings" (Petrucci 2006, 191), "displayed writings" (Fraenkel 2011), and belonging to more general "public lettering" (Petrucci 1993). Characterized by their legibility, visibility and publicness (Fraenkel 2011, 306), this family of writings encompasses a set of various, day-to-day and well known urban writings: tags, graffiti, political and advertising posters, roadsigns, signage systems, but also solemn writings specifically studied by Petrucci. Therefore, do leaflets belong to *displayed writings*, as they seem to do?

Leaflets can be and are often posted up, stuck on walls, and thus operate as displayed writings. Besides, among the family of protest writings, leaflets almost always go hand in hand with other displayed writings such as political posters, with which they are systematically linked. But a closer look at artifacts, bodies, techniques, and situations, like the ones I have presented here, clearly shows some

specificities of these writings. Instead of being displayed or exhibited, leaflets appear to be *distributed* writings in the literal sense – which can be temporarily characterized as public, free, mobile and multiple writings, (aimed at) materially reaching individual addressees among an audience, one by one.

And, as soon as one pays attention to them, a set of numerous usual practices and ordinary writings arise. They have very different graphic, discursive and linguistic characteristics, and can be distributed in various situations: for instance advertising handbills and flyers, free newspapers, but also visiting or business cards, booklets, exam papers, folders, files… Lots of objects, not necessarily written, are also distributed daily, handed out and circulated in public places (Scollon 1997). In order to avoid typologies, as well as analogies or large categories that could flatten down the performative force and the specific modes of action of these writings, it is necessary to maintain a close look at both precise utterances, graphic features, handling gestures, bodily commitment, places, and objects. Focusing both on the material characteristics of "graphic artifacts" (maps, petitions, files, lists), and on the actors, gestures and concrete ways in which these objects circulate among people and places is a fruitful way to better understand how written objects perform and shape forms of urban governance, Matthew S. Hull argues in *The Materiality of Bureaucracy in Urban Pakistan* (Hull 2012). Thus, a modest way of understanding how writings shape our world is to try and assess what things we are doing when we are concretely disseminating, distributing and circulating writings.

Acknowledgements

References

App, Corinne, Anne-Marie Faure-Fraisse, Béatrice Fraenkel, and Lydie Rauzier. 2011. *40 ans de slogans Féministes 1970/2010*. Donnemarie-Dontilly: Editions iXe.

Artières, Philippe. 2013. *La banderole: histoire d'un objet politique*. Leçons de Choses. Paris: Autrement.

Artières, Philippe, and Pawet Rodak. 2008. "Ecriture et Soulèvement. Résistances graphiques pendant l'état de guerre en Pologne (13 Décembre 1981–13 Décembre 1985) [Writing and Uprising: Graphic Resistance during the Period of Martial Law Imposed in Poland (December 13, 1981, to December 13, 1985)]." *Genèses* 1 (70): 120–39.

Austin, John Langshaw. 1975. *How to do Things with Words*. Cambridge, Massachusetts: Harvard University Press.

Barton, David, and Uta Papen. (Eds.). 2010a. *The Anthropology of Writing: Understanding Textually-Mediated Worlds*. Continuum. London; New York, NY: Continuum International Pub. Group.

Barton, David, and Uta Papen. (Eds.). 2010b. "What is the Anthropology of Writing?" In *The Anthropology of Writing: Understanding Textually Mediated Worlds*, ed. by David Barton, and Uta Papen, 3–32. London ; New York: Continuum.

Chartier, Roger. 1994. *The Order of Books, Readers, Authors, and Librairies in Europe between the Fourteenth and the Eigteenth Centuries*. Cambridge: Polity Press.
DOI: 10.1177/026569149602600208

Chartier, Roger. 1995. *Forms and Meanings. Texts, Performances, and Audiences from Codex to Computer*. Philadelphia PA: University of Pennsylvania Press.
DOI: 10.1177/092137409700900107

Chartier, Roger. 1998. "Writing the practices." *French Historical Studies* 21 (2): 255–64.
DOI: 10.2307/286629

Collard, Damien. 2010. "Ce que 'signifier' veut dire: le cas des clercs du bureau de signification de Paris." In *Le travail des huissiers. Transformations d'un métier de l'écrit*, ed. by Béatrice Fraenkel, David Pontille, Damien Collard, and Gaëlle Deharo, 59–74. Toulouse: Éditions Octarès.

Denis, Jérôme. 2006. "Préface: Les nouveaux visages de la performativité." *Études de communication* 29: 8–24.

Denis, Jérôme, and David Pontille. 2010a. "Placing Subway Signs: Practical Properties of Signs at Work." *Visual Communication* 9 (4): 441–62. DOI: 10.1177/1470357210382189

Denis, Jérôme, and David Pontille. 2010b. "Performativité de l'écrit et travail de maintenance [Performativity of Writing and Maintenance Work]." *Réseaux* 5 (163): 105–30.
DOI: 10.3917/res.163.0105

Denis, Jérôme, and David Pontille. 2011. "Materiality, Maintenance and Fragility. The Care of Things."*'How Matter Matters', Third International Symposium on Process Organization Studies*. http://halshs.archives-ouvertes.fr/halshs-00602914.

Favre, Pierre. 2006. "Les manifestations de rue entre espace privé et espaces publics." In *L'atelier du politiste. Théories, Action, Représentations*, ed. by Pierre Favre, Olivier Fillieule, and Fabien Jobard, 193–213. Paris: La Découverte.

Fraenkel, Béatrice. 1994. "Les écritures exposées." *Linx* 31: 99–110. DOI: 10.3406/linx.1994.1327

Fraenkel, Béatrice. 2001a. "La résistible ascension de l'écrit au travail." In *Langage et Travail. Communication, Cognition, Action*, ed. by Anni Borzeix and Béatrice Fraenkel, 113–142. Paris: CNRS Editions. DOI: 10.1016/s0038-0296(02)01273-6

Fraenkel, Béatrice. 2001b. "Enquêter sur les écrits dans l'organisation." In *Langage et Travail. Communication, Cognition, Action*, ed. by Anni Borzeix and Béatrice Fraenkel, 231–261. Paris: CNRS Editions. DOI: 10.1016/s0038-0296(02)01273-6

Fraenkel, Béatrice. 2006. "Actes écrits, actes oraux: la performativité à l'épreuve de l'écriture [Written Acts and Speech Acts: Performativity Facing Writing Practices]." *Études de communication* 29: 69–93. DOI: 10.4000/edc.369

Fraenkel, Béatrice. 2007. "Actes d'écriture: quand écrire c'est faire [Writing Acts: How to do Things with Writing]." *Langage et Société* 3 (121–122): 101–12. DOI: 10.3917/ls.121.0101

Fraenkel, Béatrice. 2010. "Writing Acts: when Writing is Doing." In *The Anthropology of Writing: Understanding Textually Mediated Worlds*, ed. by David Barton, and Uta Papen, 33–43. London ; New York: Continuum.

Fraenkel, Béatrice. 2011a. "Catastrophe Writings: in the Wake of September 11." In *Visible Writings, Cultures, Formes, Readings*, ed. by Marija Dalbello, and Mary Shaw, 305–18. New Brunswick, New Jersey, London: Rutgers University Press.

Fraenkel, Béatrice. 2011b. "'Nos drapeaux sont des foulards à fleurs', Les manifestations du MLF." In *40 ans de slogans Féministes 1970/2010*, ed. by Corinne App, Anne-Marie Faure-Fraisse, Béatrice Fraenkel, and Lydie Rauzier, 5–17. Donnemarie-Dontilly: Éditions iXe.

Génération MLF 1968–2008. 2008. Paris: Editions des femmes.

Grelet, Stany. 2005. "L'amour de l'art. Pour une technologie de la contestation." *Vacarme* printemps (31): 15–20. DOI: 10.3917/vaca.031.0015

Haudricourt, André-Georges. 1987. *La technologie, science humaine: recherches d'histoire et d'ethnologie des techniques*. Paris: Editions de la Maison des sciences de l'homme. DOI: 10.2307/3106044

Haudricourt, André-Georges, and Mariel Jean-Brunhes Delamarre. 1955. *L'homme et la charrue à travers le monde*. Géographie Humaine. Paris: Gallimard. DOI: 10.7202/1002763ar

Hull, Matthew S. 2012. *Government of Paper: The Materiality of Bureaucracy in Urban Pakistan*. Berkeley and Los Angeles ; London: University of California Press. DOI: 10.1017/s0010417513000492

Mauss, Marcel. 1973. "Techniques of the Body." *Economy and Society* 2 (1): 70–88. DOI: 10.1080/03085147300000003

Mbodj-Pouye, Aïssatou. 2010. "Keeping a Notebook in Rural Mali: A Practice in the Making." In *The Anthropology of Writing: Understanding Textually Mediated Worlds*, ed. by David Barton, and Uta Papen, 126–44. London; New York: Continuum.

McKenzie, Donald F. 1986. *Bibliography and the Sociology of Texts*. London: British Library.

Offerlé, Michel. 2008. "Retour critique sur les répertoires de l'action collective (XVIIIe-XXIe Siècles)." *Politix* 21 (81): 183–204. DOI: 10.3917/pox.081.0181

Petrucci, Armando. 1993. *Public Lettering: Script, Power, and Culture*. Translated by Linda Lappin. Chicago: University of Chicago Press.

Petrucci, Armando. 2006. "Crowd Writings." In *Crowds*, ed. by Jeffrey T. Schnapp, and Matthew Tiews, 191–194. Stanford: Stanford University Press.

Scollon, Ron. 1997. "Handbills, Tissues, and Condoms: A Site of Engagement for the Construction of Identity in Public Discourse." *Journal of Sociolinguistics* 1 (1): 39–61. DOI: 10.1111/1467–9481.00003

Tarrow, Sidney. 2010. "Charles Tilly and the Practice of Contentious Politics: From France to England [Not quite] Back Again." *Histoire@Politique. Politique, culture, société. Revue électronique du Centre d'histoire de Sciences Po*, no. 10. DOI: 10.3917/hp.010.0051

Tilly, Charles. 1986. *The Contentious French*. Cambridge, Massachusetts: Harvard University Press. DOI: 10.1017/s0147547900005251

Tilly, Charles. 2008. *Contentious Performances*. Cambridge; New York: Cambridge University Press. DOI: 10.1086/654741

Professional writing as institutional norm and personal style

Inspection reports in French elementary schools

Claire Doquet & Céline Poudat

Université Sorbonne nouvelle Paris 3, CLESTHIA (EA 7345), *Linguistics of writing* section / Université de Nice Sophia Antipolis, UMR 7320 BCL

The present study is based on 85 inspection reports in primary schools, composed during the 2010–2011 academic year by six National Education Inspectors in an academy in Brittany, France. Our main objective is to describe the genre of the report in a text statistics framework, using both extralinguistic (enunciative situation, pragmatics) and linguistic characteristics. Among the results we obtained, we noted that the reports were significantly grouped together by author. In spite of significant compositional and generic constraints, strengthened by the report model recommended for use by the Academy, some authors clearly distinguished themselves from the others and this allowed us to consider "personal styles" of writing for these administrative reports.

Keywords: discourse analysis; text statistics; professional genres; inspection reports; personal styles

1. Introduction

Professional writing has been theorized by anthropologists as *creation*: in contrast to its oral counterpart, written discourse is inseparably constituted by the object, which, in the most concrete sense of the term, remains its medium. "Every act of writing", affirms Béatrice Fraenkel (2008, 62) "contributes to the construction of an object, and this object takes up its place where the action occurs". Whatever its nature and form, be it a book, report, note, file or review, professional writing is an integral feature of many people's work activity, from its conception to its final distribution. Created as part of one's work, it is stored at the workplace by its author or recipient. The discourse that is produced is indissociable from the situation in which it is written (see also Delbreilh this volume; Prior this volume), and forms "part of the act of creating an object, an act relevant to written language practices" (Fraenkel 2008, 62).

The study described here was carried out as part of the ECRITURES (*Writings*) research project financed by the French National Research Agency (ANR) and conducted by the CLESTHIA laboratory at University Sorbonne nouvelle.[1] The aim of this research project is to characterize professional writing, focusing on the writing of reports in two different professional spheres:

– descriptive reports of children at risk;
– reports on elementary school inspections carried out in France by the Ministry of National Education.

It is the second category of reports that formed the subject of the present study. The final versions of these reports underwent textometric analyses using Le Trameur (http://www.tal.univ-paris3.fr/trameur/) and Dtm-Vic (http://www.dtmvic.com) software, which will be used to examine the draft versions at a later date. This chapter therefore describes one of a series of studies on the written product considered in terms of its formal conditions and pragmatic ends.

We begin by describing the inspection report from the point of view of the writing category to which it belongs, as well as the way in which it constitutes an *object* (cf. Fraenkel 2008) in the professional relationship between inspectors and the teachers in their local education authority. We then set out the characteristics of the corpus we studied and report the results of our analyses.

2. Inspection reports: Overview of the genre, discursive and institutional constraints

Written in different locations, signed by people other than their authors, addressed to three different entities, archived, and regularly consulted, elementary school inspection reports have all the characteristics of an object that occupies a position and plays a role in the professional environment. As such, its composition can be seen as an act of creation – a *creating* in the sense employed by Fraenkel (2008). This *creating* is framed by a specific form that gives the report-as-object a unique appearance. It thus takes on an official character that might be there to compensate for its general lack of efficiency.

2.1 Professional writing that is circulated

Ministry of National Education inspectors (NEIs) are responsible for a specific local education authority (LEA), a geographical zone of varying size that generally

---

contains between 40 and 60 nursery and/or elementary schools. They have author-ity over the teachers in their LEA, and inspecting these teachers is without doubt viewed as their most important task.

In principle, teachers must be inspected every four years. In reality, inspec-tions generally take place every five or six years. An inspector's visit is thus a note-worthy event for a teacher, as fewer than a dozen inspections will take place over the course of his or her career. The term *inspection*, with the semes of hierarchy and authority bestowed on it by its most frequent contexts (e.g. inspection by the police), reinforces the important nature of the event.

Today, inspections take place as follows:

1. The teacher fills out a preliminary document, which is sent in advance to the NEI. In this document, the teacher fills in sections about his or her profes-sional activity, questions, aspirations, problems, and so on. This document allows the inspector to prepare the visit, for which notice is given at least two weeks in advance.
2. Inspection visit.
    – First phase: observation of classroom teaching. The inspector spends between 1½ and 3 hours in the teacher's classroom. The inspector takes a seat, armed with the tools of the trade (pen and paper and/or com-puter), at a table on which the teacher has placed his or her various teaching aids (schedule for the term and/or year, weekly class timetable, schedule for the different disciplines, daily lesson plans, etc.). During this phase, the NEI observes the activity of the teacher and the students, looks through the aids that have been provided, in order to gain addi-tional information, and takes notes or even begins to write the report, if he or she is working directly on a computer.
    – Second phase: interview with the teacher. This interview is fairly ritual-ized. The NEI begins by listing the points that attracted his or her atten-tion. These points are discussed in detail, with the discussion possibly broadening out thereafter. The NEI will have also noted a number of aspects that the teacher is expected to explain or comment on.
3. Creation of the report: a report is drafted following each inspection visit, containing at the very least a description of the classroom observation and the interview, together with evaluative comments and advice, and finally the grade given to the teacher by the inspector.

Once it has been drawn up, the inspection report follows a complex path before being reproduced in triplicate and stored in the offices of the LEA's inspectorate (local level), the offices of the Administrative Director of the National Education

Services (ADNES, state level), and finally by the teacher who was inspected. In summary, once it has been written by the NEI, the report makes a round trip between the NEI and his or her secretary, then goes to the ADNES, the inspectorate, the teacher, and back to the NEI, before being sent to its three recipients.


1. Writing of the report
3. Proofreading + signature + suggested grade on the attached slip

6. Signing of the report
8bis. Reception and filing of a copy of the report

NEI

Teacher

Secretary

ADNES

2. Formatting of the report
5. ADNES receives the proofread report
7. Reproduction in triplicate of the report signed by the teacher
8. One copy is filed by the Inspectorate, the other two are mailed

4. Reading of the report + assignment of a definitive grade
8bis. Reception and filing of a copy of the report

Figure 1. The path of the inspection report

This complex path highlights the particular enunciative situation of report's composition: written by the NEI, it is supplemented by the ADNES, who confirms – or questions – the grade that has been suggested, and who may, if he or she so wishes, add further recommendations for the teacher. The fact that the report is addressed to three different parties raises various issues surrounding its composition:

–   With regard to the teacher, the report has a formative and even transformative goal, as it is supposed to flag up good practice, but also less good practice, by formulating observations and advice. Given the lack of disciplinary power that NEIs have over teachers, this advice tends to be couched in very moderate terms, designed not to upset or annoy the teacher.
–   The ADNES is the NEI's superior, so the NEI must be sure to present him- or herself in a good light in the report. Thus, depending on how the visit is justified, the NEI may presents him- or herself either as a conscientious employee who obediently follows the orders given by the ADNES, or else as a rebel with very little regard for his or her superiors.

2.2   The framework of the hierarchical discourse: The form of the inspection report

Depending on the LEA, the NEI may either be free to decide on the form of the report, or else be required to use a template. This was the case for the LEA we studied. A working group set up two years before these reports were compiled established a specific template for the reports that all the NEIs were supposed to follow. They did so in an attempt to harmonize reports from different LEAs. In practical terms, it also constitutes an adaptation of the existing professional discourse, thereby contributing to its legitimacy. Following the example of official judicial documents (Gasse-Granjean & Tock 2003), the template uses visual marks to create a visible hierarchy of the enunciative roles.

Mrs. / Mr.:　　　　　　　　　　　　　　　　　　　Date of inspection　　/　/
School:
Class:

1. General context, work conditions, and structure of the class.

Monitoring of the students

Attendance: attendance rate		
In September:		The entirety of the month prior to the visit:
Course: number of students		
Passing:	Failing:	Retained at the beginning of the year (repeating):
Individualized projects		
PPRE:	PPS:	PAI:	Number of AVS:

2. Conformity of the institutional tools

Conformity of the pedagogical organization

	Yes	*No*	*Observations*
Attendance record			
Schedule			
School report book/ record of competences			

3. Pertinence of professional writing

	Yes	*No*	*Observations*
Daily lesson plans			
Planning			

4. Sessions observed (pedagogical goals, proposed situations, results obtained.../analyses and advice)

5. Work and product of the students (observations, analyses, role of writing, diversity of expectations, visual displays, tools for the course.../analyses and advice)

6. Role of evaluation in the teaching practices (tools of observation and monitoring, pedagogical adaptation, use of tools, references to performance indicators, help and support tools, and student results.

7. Interview (Professional investment, analysis of teaching practices, training, review of previous reports, career, areas for improvement...)

Bibliographic information.

Figure 2. Report template recommended for use by the LEA we studied

The first three sections contain identifying and quantitative data that are not analyzed here, although some were entered in the metadata. Only Sections 4–7 lead to the composition of a text. This text may be either a description of what was observed (in standard font) or a series of remarks and recommendations (in italics). After completing this first sheet, the NEI provides a general assessment of the performance he or she observed.

The report structure described above imposes a significant constraint on writers and, whether consciously or not, many refuse to conform to it: the reports we examined contained regular and sometimes quite substantial deviations from this template. This observation is an early indication of the unique status of these reports, in that they are both institutional objects and products to which the writers personally contribute. Among the constraints on composition, in addition to the structure, we can cite the need to use *academic* terms, which serve to place the reports in the institutional context to which they belong. For example, talking about the *actual work of the students* or the *portfolio* anchors the report in the contemporary institutional discourse. Despite this institutional obligation of subordination and discourse use, NEIs can invest themselves in the composition of the reports not only structurally (by modifying the template) but also in the choice of wording. Our data showed that some NEIs had writing *styles* that were peculiar to them. These styles are connected in part to professional variables, such as the NEIs' length of service, in that they may have trained at a time when report writing recommendations and, indeed, the very profession of inspector, were very different from what they are today. These styles may also, of course, depend on variables that are more individual – and thus more difficult to define –, such as the degree to which the NEIs contribute to their own personal development (keeping up with all the latest research, or even conducting research themselves) and their personal relationships with teachers (many NEIs are married to teachers, and most are former elementary teachers).

2.3   Assessing teachers and fostering their development. The report's (well-nigh) impossible mission

The inspection report is, without doubt, primarily the record of a teacher's assessment at a given point in his or her career.[2] The writer of the report, the NEI, is the immediate superior of the teacher being inspected.

However, the inspection is of very relative importance: teachers can legally refuse to be inspected and, even though the inspection results in the awarding of a

---

2.   The French term *évaluation* is extremely polysemic and loaded with connotations in the teaching world. The École Supérieure de l'Éducation Nationale (ESEN) website provides an insight into the notion from the inspectorate's perspective: http://www.esen.education.fr/fr/ressources-par-theme/evaluation/epistemologie-et-methodologie-de-l-evaluation/

grade that is supposed to affect the professional advancement of all state employees, the NEI's liberty in the assignment of this grade is checked by a grid that shows the range of possible grades according to the teacher's length of service. Additionally, the infrequency of the inspections (once every three to four years at the most) does not allow for teachers to be properly monitored.

This relative powerlessness with regard to the work of the teachers leads NEIs to view these visits and the ensuing reports as an opportunity for dialogue and advice rather than as a means of taking disciplinary measures. In his research on the assessment of elementary school teachers, Jean Ferrier (1999) used the term *inspections* for the century 1880–1980, but preferred to talk about *assessments* for the period from 1980 onwards. In 1983, the then Education Minister Alain Savary noticeably modified the nature of the NEIs' work by (1) making it compulsory for them to give advance warning of their visit, whereas previously the visits had been unannounced, (2) placing the individual inspection within a more general system that also assesses the work context, and (3) giving teachers a right of reply and the possibility of contesting the grade.[3] In July 1990, this reform was taken one step further by statutory guidance specifying that "the evaluation primarily concerns the students and the staff. The individual assessment of each teacher is placed in the more general context of the establishment."[4] As a result of a movement where teachers refused to undergo inspections in the 1970s, the inspection moved from being punitive to being evaluative. In France, as in other European countries, the assessment of work done by individuals has increasingly moved toward the assessment of system efficiency (see Paquay, 2004). Although we can still talk about the individual report as an *object* with specific functions within the professional context, it is regularly called into question in debates about teacher assessment, and its status has changed as the report process itself has been adapted and modified.[5]

3.   Corpus and methodology

The above description of inspection reports shows just how complex they are. It is difficult to attribute a precise role to them. There is considerable tension between the ambiguity of the report's author (the empirical writer is not the one who signs it) and its evaluative and even punitive character. Then there is the gulf between

---

3.   Memorandum of 13 December 1983 regarding the modalities of inspections carried out by NEIs.

4.   Decree of 18 July 1990 and memorandum of 4 July 1990.

5.   A list of official publications, books, articles and reports can be found on the École Supérieure de l'Éducation Nationale (ESEN) website: http://www.esen.education.fr/fr/ressources-par-theme/evaluation/evaluation-des-personnels-enseignants/

the NEI's power – the mark of this power potentially being the report (the officially documented and administrative record of an *in situ* assessment of the teacher's performance) – and the relative ineffectualness of the report.

3.1   Corpus

We collated 120 reports of inspections conducted in primary education institutions (nursery or elementary schools). All these reports had been produced during the 2010–2011 academic year by six NEIs in the same LEA in Brittany. Unfortunately, one third of these reports had to be excluded from the study, and the corpus thus comprised 82 reports, totaling 94,462 words.[6]

The NEIs were chosen in accordance with two criteria:

–   urban/rural/special district (education and services for students with special needs (e.g. adaptation and schooling of handicapped children, ASH) in nursery schools, etc.);
–   length of service (less than 5 years/more than 5 years as an NEI).

Their consent to take part in this study was sought by their superiors. Four of the six NEIs were part of a working group that had spent two years developing a new template for the reports. This template had been distributed to all the NEIs, who were all supposed to use it (see above).

The data were collected by the LEA. Following an initial meeting in June 2011, the LEA made available 20 reports per NEI. There was no possible bias, as the NEIs had no knowledge that their reports would be used as material for a linguistic study at the time of writing. This first meeting led to the distribution of questionnaires that were quickly returned by all six NEIs. These allowed us to collect demographic metadata (length of service, professional responsibilities, etc.) and to gain some insight into their conception of the inspection report, its role as a supervisory device for the teachers, and so on. The questions probed the NEIs' conceptions and were organized into three areas:

–   The role assigned to the inspection report: recipient (teachers vs. superiors) and principle functions of the report (assessment vs. training vs. exercise of authority);
–   The most important elements to observe during a classroom visit;

---

6.   About 25 of the reports were inspections of principals or headmasters. These take a specific form: the section about activity in the classroom is generally shorter, and there is an additional section on administrative duties. We could not, therefore, include them with the reports about teachers, who are not school administrators. Other reports were rejected because of technical problems during collection (inappropriate format difficult to rectify).

–   The procedure for writing the report: technical aspects (composition tools), the order in which the different sections are filled in, the NEI's professional stance with regard to the teacher (e.g. choice of elements to mention in the report, reasons behind this choice).

As the inspection reports contained confidential data, they had to be rendered anonymous prior to the textometric investigations:

–   replacement of the name of the school with "rural school" or "urban school" and deletion of the name of the municipality;
–   deletion of the inspection day (we only retained the month and year);
–   replacement of the names of the six NEIs with letters from A to F;
–   replacement of the names of the teachers with identifiers made up of the letter corresponding to the NEI and a single digit (e.g. the teacher Mr A3 was the subject of the third report written by NEI A). The sex variable was communicated, enabling us to maintain the abbreviations Mr, Miss or Mrs).

In general, each report contained around 10,000 characters (range: 5000–12,000).

3.2   Methodology

In order to describe the report genre, we chose to adopt an exploratory and empirical approach. In other words, we endeavored to let the corpus speak for itself and allowed its structure and its specificities guide us, in order to test our hypotheses and arrive at our interpretations and conclusions.

In this sense, we set ourselves apart from previous studies, which were either more local and more qualitative (e.g. Cauterman et al. 2007), or else more deductive (Si Moussa 2000). This author applied quantitative methods to content analysis, such that the units that were retained and potentially annotated were the fruit of a research hypothesis, rather than being inferred from the structure of the corpus. Our analyses, on the other hand, took the textual material and corpus system as their starting point. In other words, our study was constructed on the basis of the associations and contrasts that gave the corpus its quantitative structure.

3.2.1   *Processing the corpus*
Once the corpus had been formed, it was annotated in XML in order to facilitate the extraction and import of data in formats suitable for the analytical tools we were using. The annotation was partially automated by means of scripts.[7]

In this way, both the overall structure and individual sections of each document were marked out, as were the formatting marks (bold, italics, underlined): italics had regularly been used to structure the document.

These annotations therefore allowed us to take note of variations in the structure of the reports as a whole, as well as in the separate sections making up the reports.

3.2.2   *Analytical methods*

We chose to adopt an exploratory perspective, using inductive methods to highlight the organization of the reports and the most specific and most meaningful linguistic units (words, morphosyntactic categories) that would help us to describe them. This comprehensive examination thus revealed the most important characteristics as far as the structure of the documents was concerned – characteristics we subsequently examined in greater detail, implementing more conventional and qualitative linguistic corpus methods, such as concordance and co-occurrence.

In order to bring the structure of the corpus to light, we applied a set of methods and tools developed to analyze large bodies of text as part of the *Textometry* project, based on the systematic quantification of lexicometry. Developed at the Ecole Normale Supérieure de Saint Cloud in the 1980s, and originally applied to political corpora, lexicometry involves the use of a set of methods and statistical measurements to explore and compare (sub-)corpora, considering that the norm is endogenous to the corpus, as linguistic units do not actually have frequencies in language (Lafon 1980). *Textometry* represents a particularly effective approach to the analysis of corpora, whether it is *corpus-driven* (i.e. led by the corpus) or *corpus-based* (i.e. validated by the corpus) (Tognini-Bonelli 2001).[8] Its statistical tools and methods could thus be used in both phases of our analyses, since they include functionalities that allow for both the quantitative and the qualitative observation of the corpus.

The whole of the corpus was submitted to two tools that perform exploratory analyses of textual data, and which were chosen for their complementarity. The most inductive analyses of the corpus were performed with DTM-VIC software, developed by Ludovic Lebart, and the qualitative investigations with Serge Fleury's *Le Trameur*.[9]

---

8.   The most recently developed methods actually take textual linearity and reticularity into account.

9.   These programs can be downloaded free of charge:
DTM: http://www.dtmvic.com/05_SoftwareF.html
Le Trameur: http://www.tal.univ-paris3.fr/trameur/

4.   Structure of the corpus – enunciative device and institutional constraints

In order to investigate the structure of the corpus, we first performed a correspondence analysis (CA) on all 82 reports making up the corpus. The first factorial plane we obtained is presented in Figure 3.
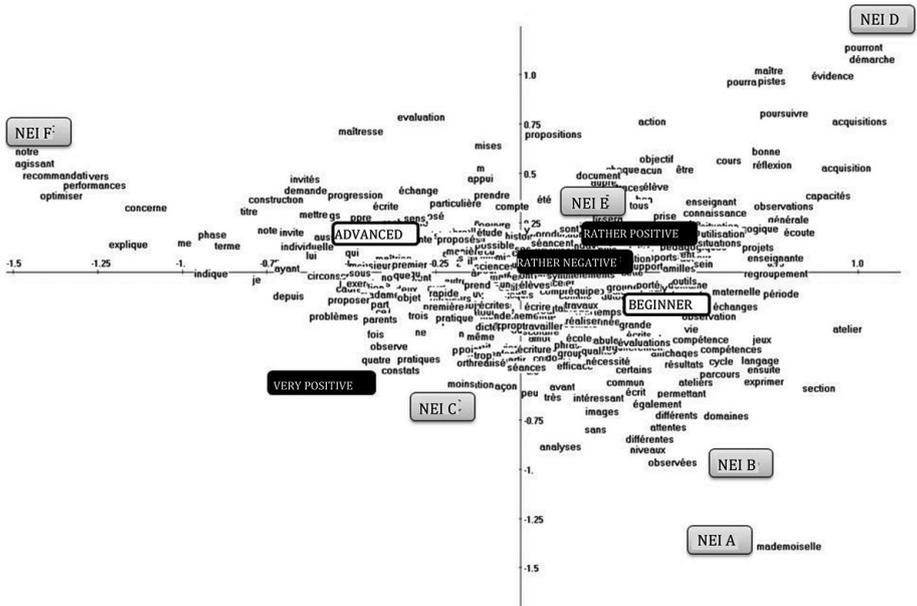

Figure 3.   First factorial plane from the CA with projection of the NEIs' categorical variables (assessment and length of service) – 82 full reports

As well as allowing us to observe the most significant contrasts and similarities in a large set of data, and identify the lexical and stylistic characteristics that determined the structure of the reports, the CA also provided information on the evaluative nature of the genre. As inspection reports are part of an evaluative practice, we might assume that the texts are primarily organized around an evaluative vocabulary that differentiates between positive and negative reports, as is the case for web 2.0 genres and Internet user ratings. However, this proved not to be the case, as the CA did not plot the reports according to the positive or negative nature of the assessment, instead contrasting the strongest assessments (i.e. either very positive or very negative) with more moderate ones (i.e. fairly positive or fairly negative). In other words, the sharply positive and negative reports were grouped together and contrasted with the more moderate reports, whether these

were positive or negative. The polarity of the overall assessments therefore had no significant character, as confirmed by our use of the bootstrap test.[10] As we can see in Figure 4, the confidence ellipses overlap considerably, showing that the moderate reports had many characteristics in common, unlike the more strongly worded reports.
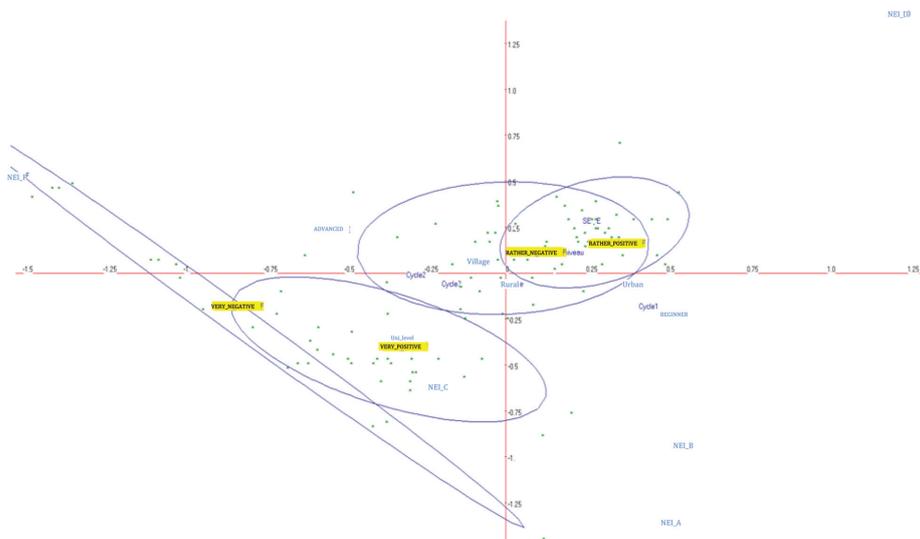

Figure 4. Confidence ellipses around the four modalities of the overall assessments

These initial observations reflect what we said in Section 1.3. about the changing nature of the inspection system and the increasingly evaluative role of inspection reports. Moreover, very negative reports were rare, such that the moderate or mixed reports primarily contrasted with very positive ones. We explore this distinction further in the following sections describing the analysis of the most significant thematic and stylistic poles.

Clear thematic poles reflect distinct conceptions of teaching and of the practice of inspection (4.1.). We can also discern what are sometimes marked personal styles (4.2.), for despite significant compositional and generic constraints, some

---

10. An extremely strict test of the stability of a structure that involves making statistical inferences about *new* samples taken from an initial sample. Thus, we disturbed the sampling of the 120 reports by placing them in a box and pulling out a report at random, *and returning it to the box each time* 120 times (on average, 70% of the reports). A sample treated in this way is a *replication* of the initial dataset. Other replications are constructed (30 or so are sufficient for a context like ours), and the analysis of the set of replications defines the halos around each point, marked by the confidence ellipses.

writers stood out very clearly from the others, allowing us to talk about report writing *styles*.

4.1   Thematic poles

In Figure 3, we can find the most significant lexical variations by observing the lexical items situated at the ends of the two axes.

4.1.1   *Teacher performance versus student acquisition (Axis 1, horizontal)*

This axis contrasts words such as *agissant* (*effective*), *notre* (*our*), *recommandations* (*recommendations*), *optimiser* (*optimize*), *performances* (*performance*), *explique* (*explains*), *vers* (*toward*), *concerne* (*concerns*), and *me* (*me*) on its negative side with words such as *démarche* (*approach*), *pourront* (*could*), *atelier* (*workshop*), *acquisitions* (*acquisitions*), *acquisition* (*acquisition*), *évidence* (*apparent*), *section* (*class*), *capacités* (*abilities*), *période* (*period*), and *écoute* (*attention*) on its positive side.

In other words, one pole is centered on the *teacher* and associated with a vocabulary of *performance* (e.g. *performance*, *optimize*, *phase*, *construction*, *progression*, *grade*, etc.), and the opposite pole is centered on the *student* and associated with a vocabulary of acquisition (e.g. *acquisition*, *abilities*, etc.).

This axis clearly reveals a contrast between NEIs D and F. NEI F is characterized by the use of terms such as *performance* and *optimize*, which focuses on the result of the teaching act presented to him. The term *optimize* pops up in all the reports written by NEI F, except for those containing very positive assessments. Each teacher whose practice is rated as sufficiently satisfactory or not very satisfactory is thus advised to *optimize* an aspect of his or her work. The focus on the students' performance likewise concerns the result rather than the process that may have led to it. By contrast, NEI D favors the *acquisition* of knowledge or knowhow, observes the *approach* taken by the teacher, and looks not at the students' *performance* but at their *abilities*.

We can assume that NEI F has a teleological conception of the teacher, whereas NEI D adopts a more open perspective.[11] This contrast clarifies the position of the *very positive* reports as far as the assessment of the teacher's performance is concerned (cf. Figure 3). The focus on the teacher triggered more enthusiastic and

---

11.   Nevertheless, this contrast is not entirely clear at first sight: the term *construction*, which refers to the Vygotskian perspective of the construction of knowledge in the pedagogical discourse, situated near NEI F, should logically accompany the term *approach* that *characterizes* NEI D. However, *Le Trameur's* concordancer shows that NEI D likewise used *construction*, but in the configuration "construction of learning". By contrast, NEI F used *construction* with a variety of complements, such as *construction of meaning*, *construction of sounds*, *construction of graphics* and *construction of language*, which explains why *construction* occurs so frequently in his reports.

more direct comments than the focus on the students – even though, at the end of the day, the effect that teaching has on the students should be the motivation behind the assessment. In NEI F's discourse, we find an almost systematic association between observation of the teacher and the *very positive* character of the report, for example:

> *A professional of considerable caliber. Mrs F3 takes the recommendations she receives into account. She regularly updates her teaching practices.*

We can also find enunciative differences between the discourses of NEIs D and F referring to their self-representations of their involvement in the exchanges they report and the unfolding of the dialogue. NEI F set himself apart from the rest of the NEIs through his use of the pronoun *me*, as in the following utterance:

> *Even though Mrs F7 tells* me *that the performance of her students is sufficiently homogenous on the whole, she nonetheless indicates to* me *that the particular situation of one student in the* GS [*class for 5-year-olds*] *is giving her cause for concern.*

This very characteristic form of NEI F's discourse includes the strong presence of *I*. Although it is closer to the center of Axis 1, it still appears in the list of words specific to F's discourse that was automatically generated by the software (cf. Figure 4 below).[12] In contrast to NEI F, NEI D favored an indirect manner, using the future tense of the modal verb *can*. This contrast is exemplified in two types of utterances that seek to give advice: "I have pointed out to Mrs X" versus "Mrs X could", where one type of self-enunciation serves to highlight NEI F and one to obscure D.[13]

4.1.2   *A traditional vocabulary that recommends versus a contemporary vocabulary that describes (Axis 2, vertical)*

Regarding the second factor, the terms on the negative side, namely *mademoiselle* (*young lady*), *observées* (*observed*), *niveaux* (*levels*), *analyses* (*analyses*), *sans* (*without*), *différentes* (*different* [*feminine plural*]), *attentes* (*expectations*), *domaines* (*domains*), *différents* (*different* [*masculine plural*]), *images* (*images*) and *écrit* (*writ-*

---

12.   The word "I", located at –0.7 on Axis 1, does not lie right at the end of the axis because it is used throughout the corpus and is less distinctive than terms like *performance* or *optimize*, which figure almost exclusively in NEI F's discourse. However, the location of *I* shows that it was used preferentially by F.

13.   A comprehensive study of the reported exchanges would reveal the abundant use of *me* and *I* by NEI F, who continually presents himself as being in the midst of speaking. At the end of the day, he is very much the hero of the inspection visit.

ing) are contrasted with the terms on the positive side, namely *pourront* (*could* [3rd *person future plural*]), *démarche* (*approach*), *maître* (*schoolmaster*), *maîtresse* (*schoolmistress*), *pourra* (*could* [3rd *person future singular*]), *pistes* (*strategies*), *poursuivre* (*pursue*), *évaluation* (*evaluation*), *actions* (*act*), *mises* (most frequently in the collocation *mises en oeuvre* [*implementation*]) and evidence (*apparent* [in *to make apparent, to show*]).

On this vertical axis, a key contrast concerns the designators of the elements observed, with people (*schoolmaster, schoolmistress* and, less significantly, *students*) on the positive side, but primarily objects (*image*, *writing*) and abstractions (*levels*, *domains*) on the negative side. Consequently, *practices* are *observed* on the negative side, whereas on the positive side, the future modals indicate recommendations, and the focus is on *approaches*. NEIs D, E, and F used *schoolmaster/schoolmistress, implementation,* and *approach*, whereas NEIs A, B, and C preferred to use *practices* and *skills*.

The contrast between *implementation* and *approach* on the one hand, and *analysis* and *practices* on the other, reflects the opposition between a top-down approach to instruction and a horizontal one. The term *implementation*, intentionally used twice, corresponds to a vision of instruction as the application of a protocol that has been thought out beforehand. Similarly, the term *approach* refers to the elaboration of a protocol. By contrast, when the NEIs mention *practices*, they refer to a set of professional acts, whether these are tools, the teacher's discourse, or the unfolding of the sessions, which is certainly supposed to find its origin in the planning of the session but in principle cannot be separated from it (as the terms *reflective practices* show, for example). Thus, whereas the term *implementation* connotes a kind of exteriority on the part of the teacher with regard to his or her action, *practices* seems more integrated.

The use of *observed*, almost always in the noun phrases *sessions observed* and *disciplines observed*, indicates a desire to objectify the inspection. By contrast, on the positive side of the axis, we find the terms *approaches* and *strategies*, which refer to (1) the teachers' pedagogical approaches, and (2) the pedagogical strategies recommended by the NEIs. Highly characteristic of NEI D, *approaches* and *strategies* are classifying designations, whereas the adjective *observed*, associated with A and B's discourse, highlights these NEIs' neutral stance.

The word *skill*, present in both the singular and the plural on the negative side, weighs less heavily on this axis than the terms commented on above. However, it is worth a mention because of its particular value in the institutional discourse: the skill-based approach, "an attempt to modernize the curriculum, reorient it, and take into account not only knowledge, but the ability to transfer it and put it to use" (Perrenoud 2000), is coherent with the "common core of knowledge and skills", the frame of reference for compulsory education in France

since 2005.[14] In the NEIs' discourse, the use of *skill*, be it in the singular or in the plural, referred in the vast majority of cases to these common-core skills. Thus, from the point of view of the interdiscourse, *skill* is part of a technical and contemporary vocabulary. It is also part of the essentially mandatory register of official discourse. Thus, it is surprising to find this term on only one side of Axis 2, as it belongs to a professional vocabulary that should be shared and therefore not characterize either axis. Its underuse by NEIs D, E, and F should be considered as characteristic, and not its presence in A, B, and C's discourse, where it has no real significance.

Consequently, it is the presence of the dated terms *schoolmaster* and *schoolmistress* (the contemporary, institutional term is *teacher*) that most strongly characterizes the discourse of NEIs D, E, and F.[15] D and F are contrasted on the horizontal axis, but come together on fundamentals like the status of the individuals being observed. Students still call their teachers by the appellatives *maître* and *maîtresse*. *Teacher*, the term most often used in curriculums and official instructions, designates people by their function and places the emphasis on the core activity of their profession, namely teaching. It is thus associated with the content to be communicated and with know-how and professional activity. In statutory documents, *maître* is generally used today to designate a grade (*maître de conférences* [*Associate Professor*]*, maître auxiliaire* [*non-certified teacher*]), and *schoolmistress* is virtually absent from the Ministry of National Education website (3 occurrences). These NEIs therefore adopt an intermediate position between the daily discourse of the students, parents, and teachers who favor *maître* or *maîtresse,* and the noospheric

---

14.   The common core document is currently in the process of being simplified. The initial version of this document, dating from 2006, is available online: http://cache.media.education.gouv.fr/file/51/3/3513.pdf

15.   The contemporary preference for *teacher* and the decreasing use of (*school*)*master* is illustrated in quantitative terms by the inverse variation between the elementary school curriculums of 1985 and those of 2008: 37 occurrences of *schoolmaster* and 0 occurrences of *teacher* in 1985 versus 22 occurrences of *schoolmaster* and 48 occurrences of teacher in 2008. Analysis of all the curriculums published between 1985 and 2008 reveals a steady increase in the use of *teacher* to the detriment of *schoolmaster*, with a particularly strong shift in 2002:

– 1985: no occurrences of the noun *enseignant* (*teacher*) (one occurrence of the gerund *enseignant*).
– 1991: 0.2 *teacher* for 1 *schoolmaster*
– 1995: 0.25 *teacher* for 1 *schoolmaster*
– 2002: 1.8 *teacher* for 1 *schoolmaster*
– 1991: 2.2 *teacher* for 1 *schoolmaster*

discourse that favors *teacher*. By contrast, the other four NEIs are resolutely on the side of the noosphere.[16]

We can apply a similar analysis to the term *practices* (especially in the phrase *professional practices*, as the terms *professional* and *practices* tend to form a collocation in the NEIs' discourse). *Practices* also features in one of the subtitles of the report template (*Role of evaluation in the teaching practices*). As with *skills*, the absence or rarity of this term in some reports is almost more significant than its strong presence in others. It was scarcely employed at all by NEIs D and F, which means that they were out of step not only with the other four NEIs we observed, but also with the institutional discourse. We only indirectly took account of the words *analyses* and *expectations* on the negative side of the axis, because they, too, feature in one of the report template's subtitles, and in the vast majority of cases, it was in this context that they emerged in the concordancer. Their presence therefore reflected the NEIs' decision to use the template, rather than an idiomatic usage. However, these two terms are significant precisely because they featured in the template: it is important to note that the vertical axis contrasts the two or three NEIs who used the template (on the negative side) with others, on the positive side, who considerably distorted it, showing that they took full responsibility for all the attributes, both hierarchical and professional, of their function.

4.2   Personal styles

As well as highlighting different conceptions of the inspection and what should be observed and assessed, the lexical axes also showed up individual differences. For instance, NEIs D and F had very particular personal styles indeed, contrasting not only with each other on the first axis of the factorial chart (Figure 3), but also with NEIs A, B, and C, who were grouped together on the negative side of Axis 2. We were able to identify the words that characterized their discourse, on both the positive and the negative sides, using data and text mining (DTM) software (value test; see Figure 5).

Generally speaking, from the first factorial analysis onwards, we noted that the reports were significantly grouped by author. Figure 5 shows the confidence ellipses around the different NEIs: none of the ellipses overlap, and their small size clearly confirms the (statistical) significance of each author.

---

16.   Likewise, NEI D borrowed the very frequent usage of the axiological adjective *good* (54 occurrences of *good* [feminine], 21 occurrences of *good* [masculine]) from daily discourse. She was without doubt the NEI whose discourse contained the least professional terminology.

```
1 qui            11.374
2 agissant       10.765
3 je              9.649
4 concerne        9.091
5 construction    9.046
6 madame          8.407
7 vers            8.095
8 que             6.865
8 également       -5.065
7 section         -5.303
6 évaluations     -5.424
5 classe          -5.616
4 compétences     -6.317
3 enseignante     -8.826
2 1               -9.586
1 et             -16.040
```

```
Class  4          testv  p
1 être            10.645
2 maître          10.324
3 pourront         8.607
4 démarche         8.273
5 pourra           8.211
6 bonne            8.022
7 enseignante      7.717
8 pistes           7.489
8 cadre           -5.481
7 monsieur        -5.520
6 plus            -5.683
5 j               -6.226
4 analyses        -6.233
3 qui             -6.413
2 madame          -8.378
1 je              -9.209
```

```
Class  1          testv  p
1 mademoiselle    10.045
2 analyses         8.993
3 langage          7.788
4 ateliers         7.785
5 écrit            7.416
6 section          6.569
7 observation      6.472
8 observées        5.548
8 cours           -4.002
7 poursuivre      -4.043
6 agissant        -4.077
5 ses             -4.228
4 réflexion       -4.483
3 concerne        -4.898
2 été             -5.108
1 qui             -6.542
```

Figure 5. Confidence ellipses around the NEIs–unique vocabulary of NEIs A, D, and F

As we have just indicated, the NEIs who were the furthest away from the center (those who distanced themselves most significantly and who were therefore the most singular) were A, D, and F.

F contrasted with D (and with A, albeit to a lesser extent) on the first factor, whereas D and F contrasted with A on the second factor. We have already identified and described the contrasts between F and D on the first factor (see Section 3.1 above).

We based our analysis of personal styles (still only exploratory) on these three NEIs. For this reason, we only show the words that were unique to them in Figure 5.

Whereas the preceding sections deal with the characterization of the DEIs' socio- and technolect, it is their idiolects that were the focus of our research here. We also examined the NEIs' descriptions of their practices in the questionnaire we distributed.

4.2.1 *Lexical contrasts*
NEI F was characterized above all by significant usage of the relative pronoun *qui* (*who*) (v-test +11.3) and the conjunction *que* (*that*) (+6.8), revealing a tendency toward a subordinating writing style. A and D, however, who contrasted with F on Axis 1 (teacher's performance vs. students' acquisition) were not positively

characterized by grammatical words, and A even underused the relative pronoun *who* (−6.5).

An indication of the writers' enunciative commitment to their writing, the personal pronoun *I* likewise significantly characterized the styles of F (+9.6) and D (−9.2). It is worth recalling here that D was characterized by the usage of the verb *pouvoir* (*can*), nearly always modal and in the future tense. NEI D also gave advice in an indirect manner, and her manner of expression distanced her from her remarks, even though she passed judgement on her observations.

From a lexematic perspective, NEI D was characterized by the use of the adjective *good* in the feminine (v-test +8, present 57 times in the 14 reports produced by this NEI, equivalent to four occurrences on average per report). In addition to its inherent axiological seme, it contained a moralizing element here. Rather surprisingly, given that reports are supposed to be objective observations, *bonne* (*good*) was used to qualify nouns such as *situation* (*situation*), *attitude* (*attitude*), *facture* (*technique*), *compréhension* (*comprehension*), *mise en place* (*setup*) and *réflexion* (*reflection*). In constructions like *good reflection* or *good setup*, for example, we can see that the adjective's meaning is clearly subjective, and the absence of any other qualification heightens this impression. This use of *good* to describe situations for which axiological assessing scales exist is in some ways similar to the use of the old-fashioned character of the appellative *maître* (*schoolmaster*) commented on above. It reveals a strong personal involvement in the judgment of the actions being observed, and indeed of the protagonists, as they are preferentially designated by the co-presence of the terms *schoolmaster* and *teacher* on the positive axis. NEIs D and F are grouped together on Axis 2 as the authors of axiological discourses that reveal a rather traditional view of teaching practices. They personally involve themselves, either directly or indirectly, in their function of giving advice to teachers, all the while clearly indicating a hierarchical distance (predominance of *I*/future modal).

Contrasting with NEI F on both axes, NEI A is characterized by the overuse of objectifying vocabulary: *analyse* (*analysis*), *observations* (*observations*), and *observées* (*observed* – in the collocation *practices observed*). The contrast with F is lexically materialized by three words: *qui* (*who*), *agissant* (*effective*), and *concerne* (*concerns*). These words are unique to F's discourse and are noticeably absent from A's. The contrasting use of the term *section* (*class*) (very frequent in A's reports, but strangely absent from F's) does not have the same meaning: used in the collocations *grande section*, *moyenne section*, and *petite section*, corresponding to classes for 5-, 4-, and 3-year-olds, it specifically designates nursery school classes, and its frequency simply indicates that A's reports frequently concerned nursery schools and those of F elementary schools. This is corroborated by the presence in A's reports of the terms *langage* (*language*) and *ateliers* (*workshops*), which are both associated with nursery school.

A also contrasts with D on Axis 2 (traditional and axiological vocabulary vs. contemporary and objectifying vocabulary) by his use of *analyses*, which was underused by D.

4.2.2   Contrasts in background

NEIs A, D, and F had distinct professional backgrounds: A (male) and D (female) had only recently embarked on their careers as NEIs, whereas F (male) had been an NEI for 10 years. A and F completed the full training program <u>10 years apart from one another</u>. D, on the other hand, was recruited without having to take the competitive entrance exam, and did not undergo this training. These characteristics may explain the differences and similarities between these NEIs:

– A, having only recently been trained, used a contemporary vocabulary and utilized techniques he had learned very recently. The more traditional vocabulary employed by F can be explained by the fact that he trained much earlier, while the vocabulary used by D may be due precisely to not having undergone any training, thus obliging her to refer to her initial teacher training, which by then was already dated;
– It may also be the training and its legitimizing effect that explains the enunciative marks used by the three NEIs. Without doubt, F's length of service and the legitimacy he derived from it made him feel self-assured and influenced his propensity to use *je* (*I*) and involve himself personally in the narration of the observations and interactions with the teacher being assessed. *I*, not one of the words that characterized A's discourse (i.e. neither over- nor underused), was strangely absent from D's discourse. Correlated with her lack of training, this may reflect the difficulty she had feeling like a legitimate inspector.

In addition to the processes of linguistic legitimization, the NEIs' responses to the questionnaire about their professional stance revealed differences that likewise reflected their feelings of legitimacy or otherwise. Regarding the primary role of the inspection report, A and F responded that it was to *assess the teacher's professional skills,* whereas for D, it was to *provide the LEA's inspectorate with a picture of the teacher's work.* We can assume that she felt the need to refer back to her superiors in order to validate her observations and her own professional practice.

4.2.3   Procedural contrasts

The three NEIs also composed their reports in different ways. First of all, the writing medium varied. A and F used word processors straightaway, and in fact began composing their reports right from the start of the process, while they were observing the teacher. By contrast, D did not bring her computer into the classroom,

taking notes by hand during the observation and reviewing them in their entirety once she had returned to her office. The responses of these three NEIs led us to identify three separate working periods: during the classroom observation, during the interview, and outside the classroom after the visit. The tools (word processor vs. pen and paper) were used in the following manner:

NEI F	NEI D	NEI A	
Composing the report directly on the computer	Taking handwritten notes	Composing the report directly on the computer + jotting notes down in a notebook to prepare for the interview	*Classroom observation*
Word processor	Taking handwritten notes	Taking handwritten notes	*Interview*
Resuming use of the word processor	Composing the report on the computer	Resuming use of the word processor	*Back at the office*

We can hypothesize that the variations in the composition medium during the observation period–pen and paper for D, word processor for A and F – affected the NEIs' cognitive activity. In the first case, the composition primarily corresponded to note-taking, which Pétillon (2004) calls "writing with urgency". The goal is to record as quickly as possible the information that will later be organized and reformulated to constitute the text of the report: "taking notes within a work situation consists in understanding a flood of information expressed in words, picking out the essential or useful parts, and composing a written product that cannot be compared to an explicit and linearly organized text" (Piolat 2010, 55). Numerous psychological studies have already demonstrated the significant cognitive effort required for note taking (cf. Piolat 2006 for a review). Despite all this effort, notes do not constitute a text that is readable for others, and the composition's communicative dimension pertains solely to the communication that can be said to take place within the author's mind *at a later time*. In other words, the notes must be sufficiently clear for them to be used later on. The task that NEIs A and F took upon themselves, namely directly composing the report on the computer in the classroom during the actual observation, implies a supplementary dimension: taking the reader into account, or rather the final readers of the report, namely the ADNES and the teacher being assessed. The task was more complex still for NEI A, because of his sporadic recourse to pen and paper whenever he noticed things he felt warranted questions in the interview phase. A precise study of the NEIs' handwritten notes and *online* writing – an investigation that has been planned but not yet performed – would allow us to address the difference between the two

procedures in detail. A recording of NEI A has already shown that the direct composition of the report on a computer cannot be divided into steps corresponding to note-taking followed by formatting. On the contrary, NEI A simultaneously formulated and formatted his text, frequently stopping to put segments of text in italics or capital letters. At least as far as the writing process is concerned, he did not go through phases of content elaboration, as the syntactic-lexical characteristics already complied with the report norms. When he returned to his text following the visit, NEI A carried out a general tidying-up, during which he inserted a few new elements, particularly for the *interview* section. NEI D, on the other hand, had to start composing the report from the very beginning, based on the notes she had taken during the session. During report composition, therefore, the activity differs depending on which instrument was used, screen or paper.

The three NEIs we observed here also differed from one another in terms of procedure, that is, the order in which they filled out the various sections of the report. There are six of these sections, which are listed in the left-hand column of the table below, in the order in which they appear in the template report. The other three columns contain numbers indicating the order in which each NEI filled out these sections.

NEI F	NEI D	NEI A	
Before the visit	1	1	Identification
Before the visit	6	6	Monitoring of the students
	2	3	Conformity of the institutional tools
	3	5	Pertinence of professional writing
1	4	2	Sessions observed
	5	4	Work and product of the students
		7	Role of evaluation in the teaching practices

Whereas D filled the sections out one after another, following the order provided by the template and justifying this procedure by her desire "to take the template into account", A stated that he proceeded according to the recommendations of the LEA's working group: "begin by noting what you see first (*natural* order of things)." Rather than a choice or a habit, this "facilitated" the procedure for him, "knowing that we have to take in a lot of information in a limited space of time". As for NEI F, he did not really seem to take the sections into account at all, adopting a procedure all his own. He indicated that the organization of the writing depended on the situation he had before him and the elements he viewed

as important in that situation, thus reflecting a relative independence from the institution, doubtless to do with his length of service. These differences in procedure appear to reflect the NEIs' greater or lesser willingness to adapt themselves to the collective constraints. This is certainly not the least of the paradoxes of this profession, whose members represent the state and speak with official authority (i.e. collectively, as state employees), while at the same time acting and speaking as individual inspectors in their own districts.

5.   Conclusion

This initial exploration of inspection reports allowed us to identify several different dimensions of this professional genre, reflecting interactions between institutional constraints and the liberties taken by individual NEIs. Intended both to assess and to foster teachers' professional development, these reports play an important role, in that they reflect professional practices at a given moment in time, but one that is ultimately limited, owing to the rarity of the visits and the ineffectual nature of the suggestions that are given. Nonetheless, the form and content of these reports clearly illustrate the NEIs' investment in this mission, and above all show the variety of stances the reports' authors are able to adopt within what is a restricted form with well-defined institutional concerns. Results revealed disparate representations, with as many elements to evaluate (teacher's performance vs. students' learning) as ways to assess them and to convey those assessments.

Additionally, these reports can reveal different professional (or personal?) styles. For instance, NEIs A, D, and F stood out particularly clearly in the factorial chart, the lexical contrasts without doubt reflecting major divergences in their vision of the profession. Lying at the interface between professional stances and elements that are more about each inspector's personality, these examples of professional writing are objects that symbolize the relationship between an NEI and the teachers in his or her LEA, and establish writing as the locus of contradictory interactions between roles and individuals.

References

Cauterman, Marie-Michèle, and Bertrand Daunay. 2007. "Le métier d'enseignant de français au crible du rapport d'inspection." *Recherche* 47: 85–94.

Ferrier, Jean. 1999. "L'évaluation des maîtres des écoles primaires." *Recherche et formation* 30: 101–120.

Fraenkel, Béatrice. 2008 "Comment tenir un registre?" *Langage et société* 124: 59–71.
    DOI: 10.3917/ls.124.0059

Gasse-Granjean, Marie-José, and Benoît-Michel Tock. 2003. "Peut-on mettre en relation la qualité de la mise en page des actes avec le pouvoir de leur auteur?" In *Les actes comme expression du pouvoir*, ed. by Marie-José Gasse-Grandjean, and Benoît-Michel Tock, 99–123. Brepols: Turnhout.

Lafon, Pierre. 1980. "Sur la variabilité de la fréquence des formes dans un corpus." *Mots* 1: 127–165. DOI: 10.3406/mots.1980.1008

Paquay, Léopold. (Ed.). 2004. *L'évaluation des enseignants: tensions et enjeux.* Paris: L'Harmattan.

Perrenoud, Philippe. 2000. "L'approche par compétences: une réponse à l'échec scolaire?" In *Actes du XXe colloque de l'AQPC.* http://francais-au-lycee-dz.e-monsite.com/medias/files/approche-par-competence.pdf

Pétillon, Sabine. 2004. "Abréviations et genèse du texte parlementaire: l'écriture en urgence des secrétaires analytiques." In *Écriture abrégées (notes, notules, messages, codes…). L'abréviation entre pratiques spontanées, codifications, modernité et histoire*, ed. by Nelly Andrieux-Reix, Sonia Branca, and Christian Puech, 163–174. Gap: Ophrys.

Piolat, Annie. 2006. *La Prise de notes* (2nd ed. revised). Paris: Presse universitaires de France.

Piolat, Annie. 2010. "Approche cognitive de la prise de notes comme écriture de l'urgence et de la mémoire externe." *Le français aujourd'hui* 170: 51–62. http://www.cairn.info/zen.php?ID_ARTICLE=LFA_170_0051 DOI: 10.3917/lfa.170.0051

Si Moussa, Azzedine. 2000. "Les rapports d'inspection du 1er degré, contenus et finalités: un exemple à La Réunion." *Revue française de pédagogie* 133: 75–86.
DOI: 10.3406/rfp.2000.1022

Tognigi-Bonelli, Elena. 2001. *Corpus Linguistics at Work.* Amsterdam: John Benjamins Publishing Company.

Cognitive insights through writing studies

Conflict in writing

Actions and objects

David Galbraith & Veerle M. Baaijen

Southampton Education School, University of Southampton /
Centre for Language and Cognition, University of Groningen

This chapter argues that writing involves an interaction between conflicting cognitive systems, one designed for the construction of mental objects and the other for the taking of actions. It characterises the construction of mental objects as a problem-solving process involving the retrieval of content from episodic memory and the manipulation of content in working memory. The system for action involves the synthesis of content guided by implicit constraints within semantic memory. The chapter then reviews research investigating the effects of different types of planning and individual differences in goals and beliefs on the development of understanding during writing and on the quality of text. It concludes by discussing the effectiveness of different drafting strategies and implications for theories of writing.

Keywords: writing processes; dual-process model; knowledge constituting; knowledge transforming; writing strategies

1. Introduction

Writing should be easy. Given that we have mastered the art of transcribing language into visual signs, it should simply be a matter of transcribing whatever thoughts we have into a visual form. This process of transcription is of course a uniquely human accomplishment, taking years of education and practice before it becomes fluent, but once transcribing has been mastered writing is surely simply a matter of transcribing thought into an external form. Aren't our minds filled with thoughts every waking moment of the day? Don't we chatter away, more or less happily, to one another all the time in the course of our everyday interactions? Why, then, is writing so difficult?

The answer, of course, is that writing is rarely just a matter of transcribing thought. It may approximate to this sometimes: when, for example, we write an informal letter, or chat with one another on Facebook. But "serious" writing

involves creating a "knowledge object" capable of standing on its own independently of the author (Galbraith, van Waes & Torrance 2007). This knowledge object, once created, becomes part of what Karl Popper called 'world 3' – the world of objective contents of thought (Popper 1972; see also Bereiter 2002). This world includes cultural artefacts, and is created by the mental processes and behavioural dispositions that make up Popper's 'world 2'. In our view, what makes writing different from other forms of language production, and more than simply a matter of transcribing thought into a visual form, is that it involves creating an independent knowledge object and that this is designed to make a contribution to world 3.

We argue in this chapter that there is a conflict at the heart of the writing process between the systematic and well-organised nature of the final product and the processes involved in text production. This is a conflict between a system designed for constructing objects and a system designed for action. In the next section, we will describe Bereiter & Scardamalia's (1987) knowledge transforming model of writing and argue that it describes a system for constructing objects. We will then summarise a dual-process model of writing (Galbraith 2009) and argue that it incorporates a system designed for action and explains how this conflicts with the object system. We will then describe the results of two recent experiments designed to elucidate the processes involved in planning and text production. We will conclude by assessing the implication of these experiments for the two models of writing.

2.   Writing as the construction of a knowledge object

Bereiter & Scardamalia (1987) characterise reflection in writing as an interaction between two problem spaces – a content space and a rhetorical space. The content space is, in Bereiter & Scardamalia's (1987, 302) words:

> made up of knowledge states that may be broadly characterized as *beliefs.* It is the kind of space in which one works out opinions, makes moral decisions, generates inferences about matters of fact, formulates causal explanations, and so on.

By contrast, the rhetorical space is:

> specifically tied to text production. The knowledge states to be found in this kind of space are *mental representations of actual or intended text* – representations that may be at various levels of abstraction from verbatim representation to representations of main ideas and global intentions. Whereas the goal states in the content space are knowledge (in the sense of warranted beliefs), the goal states in the rhetorical space are plans for achieving various purposes in composition.

We think that these spaces can be seen as roughly equivalent to Popper's world 2 and world 3. The content space is orientated towards an individual's beliefs

and involves processes designed to manipulate these; the rhetorical space is an individual's mental representation of the desired knowledge object to be created in world 3. Writing is a matter of retrieving content from the content space in response to the requirements of the rhetorical space. The beliefs that the writer selects from the content space are shaped by the properties of the knowledge object that is to be created. Writing, therefore, shapes an individual's thinking in terms of the shared, and culturally created, properties of the knowledge objects in world 3. Klein (1999) characterises this as the *genre* hypothesis in his review of theories of writing-to-learn.

Development in writing, in this view, consists of a movement from a *knowledge-telling* approach to writing, in which text production is controlled primarily by the content space, to a *knowledge-transforming* approach to writing, in which text production is controlled by an interaction between content and rhetorical spaces. The result is that knowledge telling is a "think-say process of composition" (Bereiter & Scardamalia 1987, 304) in which the writer translates their existing beliefs into words. By contrast, knowledge transforming involves a "two-way process of information transfer, which results in the joint evolution of the composition and the writer's understanding of what he or she is trying to say". (*ibidem*) Thus, development of writing can be seen as internalising the cultural practices embodied in the knowledge objects of world 3. Once these practices are internalised, writing is transformed into an interaction between content and rhetorical space, leading to the development of the writer's understanding as they write, and enabling the writer to make more effective contributions to world 3. Learning how to write effectively enables the process of writing to also become a process of writing-to-learn, in which the production of effective text and the development of understanding go hand in hand (Klein & Kirkpatrick 2010).

In addition, Bereiter and Scardamalia also emphasized that development in writing could be influenced by increases in working memory capacity during development, and shared the common assumption in the field that cognitive overload is a fundamental problem in writing. However, they had relatively little to say about how this might be alleviated by different drafting strategies.

Research on drafting strategies (see Kellogg 1994 for a review) assumes that their function is to reduce the number of different activities that have to be carried out at the same time. In particular, it is assumed that, within a limited capacity cognitive system, resources demanded by text production can reduce the resources available for higher level planning and evaluation. This competition for resources can be reduced by separating out the different components of the writing process so that each can be carried out more effectively. Thus, in an outlining strategy, the writer concentrates on generating and organising their ideas before then focussing on translating their outline into text. In a revising strategy, text is

produced spontaneously in an initial draft, and then revised into a rhetorically appropriate form.

Kellogg's (1994) research has suggested that outlining leads to the production of higher quality text than when the writer tries to combine planning and text production in a single draft. One possibility is that this is because this enables the writer to carry out knowledge-transforming processes more effectively. Thus, when creating the outline free from the demands of full text production, the writer may be better able to construct a model of the knowledge object to be created. And, having defined their goals in rhetorical space more explicitly, the writer may be better able to engage in the interaction between content and rhetorical space required for knowledge transforming during the production of the text itself.

In conclusion, we have argued that knowledge transforming involves goal-directed problem solving designed to create a knowledge object in world 3. This involves three main assumptions about the writing process. First, producing effective text is associated also with the development of the writer's understanding. Second, text production is a relatively passive process, equivalent to knowledge telling, and is controlled by higher level problem solving operations. Third, strategies like outlining, which enable writers to separate these higher level processes from the lower level processes, should enable writers both to produce higher quality text and to develop their understanding more effectively.

3.   Writing as action

In this section, we argue that the knowledge-transforming model underestimates the active role of text production in the generation of content. We describe a dual-process model of writing (see Galbraith 2009, for an overview), in which text production is treated as a form of action, out of which the knowledge object ultimately emerges. The dual-process model makes three main claims.

First, it claims that spontaneous text production is not simply a matter of knowledge telling, in which ideas are retrieved from memory and translated into text, but is an active knowledge-constituting process, guided by implicit constraints within semantic memory, which leads to the development of the writer's understanding.

Second, explicit planning is assumed to operate on existing ideas in episodic memory, as well as ideas created by the knowledge-constituting process, to construct a knowledge object in working memory. This is analogous to the knowledge-transforming process in that it involves creating a coherent mental model of the text designed to satisfy rhetorical goals. However, by itself, this is not associated with the development of understanding.

Third, both processes are assumed to be required for the production of effective text. The knowledge-constituting process is required to articulate the writer's implicit understanding of the topic, and the explicit planning process is required to organise the text and adapt it to rhetorical goals. Writing is a combination of actions to create content and planning processes to construct an explicit rhetorical object. However, the fact that the two processes are organised in different ways leads to a fundamental conflict in writing. This conflict is different to the cognitive overload assumed in problem-solving models of writing. It is not about the number of different processes that can be carried out at the same time but about which of two different paths to pursue at any given moment. Individual differences in writing are assumed to reflect the extent to which writers prioritise the two processes and the different ways in which they resolve the conflict between the processes.

In the next two sections, we first explain the processes involved in the knowledge-constituting process, characterising them as a form of action. We then describe the episodic and semantic memory systems that the knowledge-constituting and knowledge-transforming processes operate on.

3.1   The knowledge-constituting process

To give you an idea of the principles involved in the knowledge-constituting process, consider the very simple, 'toy' network shown in Figure 1. This consists of a
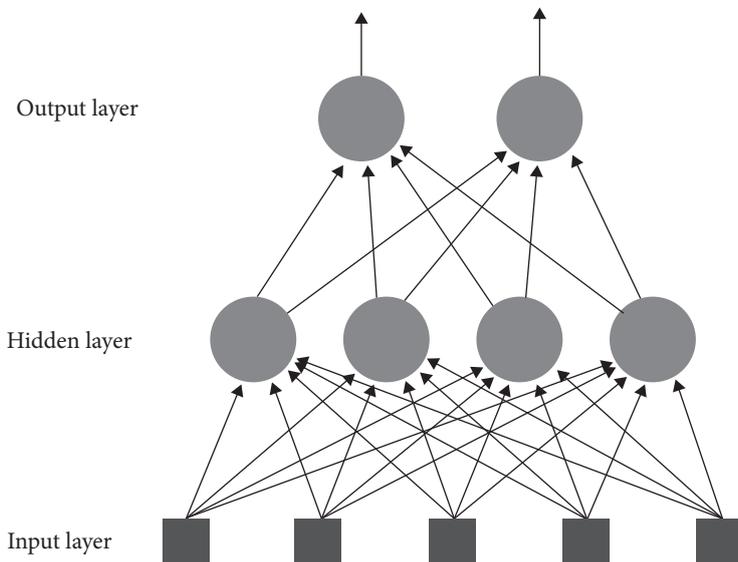
Figure 1. A simple feedforward network

set of simple units, roughly analogous to neurons, organized in three layers. Each unit has the function of summing up the activation passed to it via the connections from other units (shown as arrows in the diagram), and then passing on activation to other units in the network. The work of processing information is done by the connections between units, which vary in strength depending on the network's learning history. Thus, when a pattern of activation is presented at the input layer some of the units in the input layer will be strongly activated whereas others will be much less so. Each unit will then pass its activation on to the units in the next, hidden, layer, with the amount of activation it passes on depending on the strength of the connection it has with each of the units it is connected to. The overall effect on the units in the hidden layer will depend on the pattern of activation presented at the input layer and the varying strengths of the connections between this layer and the units in the hidden layer. The hidden layer will in turn pass activation forward to the output layer, which will then produce a response to the input the network has received.

The key feature of such a distributed network is that the same set of units and connections is used to produce responses to all its different inputs. Learning in such networks consists in the gradual adjustment of the strength of the connections between units to enable the same set of connections to reproduce existing responses while at the same time learning to produce new material. This is a slow process, with connections being gradually adjusted over time to assimilate new knowledge. The result is a fixed set of connections, abstracted from individual experiences to create the network's knowledge.

Scaled-up, this kind of distributed network is assumed to constitute semantic memory, with the equivalent of the hidden layer corresponding to a set of semantic features, and responses to input being synthesized by the passage of activation through the fixed set of connections between the units and layers (see Churchland (2012) for a recent overview of these operating principles and their philosophical implications, and Rogers & McClelland (2004) for specific applications to semantic memory phenomena).

The dual-process model claims that this constitutes one source of content in writing. This has two broad characteristics. First, in order to act we have to be able to bring together everything that we have learned and apply it to our present circumstances. The function of a distributed representation is not just to represent the regularities that we have induced from our past, it is also to provide the resources for acting in the present. This system is used precisely when we are required to take action in words. Indeed, the ultimate reason why a semantic memory system has evolved in the first place is precisely in order to enable us to be able to rapidly bring together all our past experiences and apply them to our current circumstances.

Second, although the synthetic nature of the process means that content is produced spontaneously without conscious deliberation, this does not mean that this is simply associative knowledge telling. Synthesis is controlled by the fixed connections between units in the network. It is these connections which constitute the implicit organisation of the writer's knowledge. The dual-process model characterises this as the writer's disposition towards the topic. Furthermore, because this is built up through exposure to examples within different rhetorical situations and genre contexts, this is not simply a repository of de-contextualised content in a separate content space. Instead, the implicit relationships between content reflect the contexts within which that content has been learned. The key feature of the writer's disposition is that it is implicitly organised: it is abstracted from the totality of the writer's experience and constitutes the writer's implicit self.

By itself, this is an account of how the content of individual bursts of language is produced. However, each of these bursts is only a partial representation of the content within the writer's disposition. In order to fully articulate this implicit understanding the network has to continue to create further content. The second key feature of the knowledge-constituting process is that this is achieved by inhibitory feedback from the output to the hidden units constituting the writer's disposition. Such inhibitory connections are a common feature of recurrent networks designed to produce sequential behaviour (see Glasspool, Shallice & Cipolotti 2006, for an example of a network designed for spelling). In the present context, inhibitory feedback has the effect of suppressing units associated with the initial output, with the result that even though the external input to the network remains the same the network produces new content on the next cycle of synthesis. Notice, here, that this content is not available initially: it is only synthesized once the preceding content has been created. The sequence of content consists of bursts synthesised at the point of utterance, and the writer's understanding is discursively constituted in the text. Interrupting this process before it is complete – by, for example, evaluating individual bursts in terms of rhetorical goals – will prevent the writer from constituting their implicit understanding in the text.

In summary, the knowledge-constituting process consists of the synthesis of content, controlled by the fixed connections within the writer's disposition, followed by inhibitory feedback to create a sequence of bursts.

3.2   Episodic and semantic memory

A major problem with a distributed form of representation is that new examples (or experiences) tend to overwrite earlier learning: in modifying the

connections required to represent a new example, previous examples are no longer recoverable. McClelland, MacNaughton, and O'Reilly (1995) suggested that, in order to overcome this problem, two complementary learning systems are required. Both use the same general principles of processing, however, the first of these – located in the *hippocampus* – creates a sparsely coded representation of individual items of content, whereas the second – located in the *neo-cortex* – creates a distributed representation in which content is represented as different patterns of activation across the same units. In effect, then, events or experiences are stored in two different ways: they are stored as individual memories or assimilated within existing knowledge and become part of one's general knowledge. This core distinction between two distinct forms of representation has proved capable of accounting for a wide range of biological, neuropsychological, and behavioural data (see Norman 2010; O'Reilly, Battacharyya, Howard & Ketz 2011 for recent reviews.) Particularly relevant in the present context is the suggestion by Winocur, Moscovitch, and Bontempi (2010, 2340) that: "There is a dynamic interplay between the two types of memory such that one or the other may be dominant depending on their relative strength and the circumstances that elicit them at retrieval. As a result, retention and retrieval are continually evolving processes in which the memories can interact and influence each other."

The dual-process model claims that these two systems provide different sources of content in writing. The *episodic memory* system consists of a memory of ideas as individual, already created objects; the *semantic memory* system provides the conceptual resources required to take action in the present. The writing process varies depending on which of these systems is used to generate content. When the semantic memory system is dominant, writing becomes a knowledge-constituting process; when the episodic memory system is dominant, writing becomes a knowledge-telling process; when ideas retrieved from episodic memory, or synthesized by semantic memory, are evaluated and manipulated in working memory to satisfy rhetorical goals, writing becomes a knowledge-transforming process.

Writing is at its most effective when writers are able to constitute their implicit understanding in the text, and select and organise the ideas in the text so that they satisfy rhetorical goals. However, these two processes conflict with one another. The knowledge-constituting process needs to be allowed to unfold, burst by burst, in the text in order for the writer to develop their understanding, and is inhibited when it is interrupted by external goals, or when a predetermined organisation is imposed on the text. The knowledge-transforming process operates on fixed objects within working memory, and is reduced in effectiveness when ideas are not fixed but are still in the course of being constituted.

The two different processes vary in how active they are depending on a range of factors. These include:

1. *The state of the writer's knowledge*. When writers have relatively little experience in a domain, they will not have been able to build up a rich distributed representation in semantic memory, and in consequence the knowledge-constituting process will be less active, and they will be more reliant on individual ideas stored in episodic memory. This provides an alternative to Bereiter and Scardamalia's (1987) explanation of knowledge telling in younger writers.
2. *How thought is represented*. When thought is represented economically, in note-form, episodic memory is assumed to dominate, and the explicit organising process is facilitated, while the knowledge-constituting process is reduced. When thought is articulated as explicit connected propositions in full text, the synthetic process is assumed to dominate, while the explicit organising process is reduced.
3. *The goals towards which writing is directed*. When writing is directed towards rhetorical goals, the explicit knowledge-transforming process will be prioritised; when writing is directed towards dispostional goals, the knowledge-constituting process will be prioritised.
4. *Beliefs about how the processes should be combined*. Writers may be taught, or develop their own, strategies for combining different writing processes. For example, the traditional outlining strategy taught in schools would be expected to facilitate the knowledge-transforming process but reduce the knowledge-constituting process.

4. Empirical evidence

The main evidence for the dual-process model is research showing that different types of writers develop their ideas under different writing conditions (see Galbraith 1992; 1999, and 2009; Galbraith, Torrance & Hallam 2006).

These experiments have typically used Snyder's (1979) self-monitoring scale to distinguish between writers who prioritise either rhetorical or dispositional goals. High self-monitors are "particularly sensitive to the expression and self-presentation of relevant others in social situations and use these cues as guidelines for monitoring (that is regulating and controlling) their own verbal and non-verbal self-presentation" (Snyder 1979, 89), and are therefore assumed to be more likely to direct their writing towards rhetorical goals. By contrast, low self-monitors' "self-presentation and expressive behaviour […] seems, in a functional sense, to be controlled from within by their affective states (they express it as they feel it)

rather than moulded and tailored to fit the situation" (Snyder 1979, 89), and are therefore assumed to direct their writing towards dispositional goals.

In brief summary (see Galbraith 2009, for a review), these experiments have consistently shown that low self-monitors produce more new ideas during text production than high self-monitors, and that only the new content produced by low self-monitors is associated with the development of the writer's understanding. None of these studies found any relationship between the number of new ideas produced during text production and the development of understanding for the high self-monitors. These results are consistent with the dual-process model's claim that dispositionally guided text production is not simply a matter of knowledge telling but is an active process leading to the development of the writer's understanding.

In the earliest study, Galbraith (1992) also found that when writers were asked to make notes in preparation for an essay, rather than write full text, the effect was reversed, and high self-monitors produced more new ideas than low self-monitors. However, these were not associated with increases in understanding and Galbraith (1992) concluded that they reflected a reorganisation of existing ideas rather than the development of new content. This is consistent with the dual-process model's second claim that, although explicit planning directed towards rhetorical goals does lead to the reorganisation of content, it does not lead to the development of understanding.

This latter finding has not been followed up until relatively recently. In addition, none of the research we have described so far has assessed the quality of the texts, or how this is related to the development of understanding. In the next section, we describe two more recent experiments designed to assess the effects of planning and text production on the development of understanding and text quality. The first assesses the extent to which making an outline involves the transformation of knowledge, and how this relates to the quality of the text that is then produced. The second assesses the effects of writing beliefs and different types of planning on text production.

4.1   Constructing an outline

Although research has consistently suggested that outlining has beneficial effects on text quality (Kellogg 1994), there has been hardly any research investigating whether this is because it enables writers to carry out knowledge transforming more effectively. On the assumption that an outline enables writers to represent ideas in an abbreviated form, and to concentrate on constructing a model of the text to be written, one would expect that writers would be better able to transform their knowledge to satisfy rhetorical goals, and that this would be related to the quality of the final text.

In a recent study, Galbraith, Hallam, Olive, and Le Bigot (2009) compared low and high self-monitors and investigated the role of different components of working memory in knowledge transformation during the construction of an outline (see also Galbraith, Ford, Walker & Ford (2005), for a previous study addressing similar issues).

Two groups of low and high self-monitors were asked to write an argumentative text in three phases: listing ideas; creating an outline for the text; and then writing the text itself. They did this under one of four conditions, varying in the nature of the secondary task carried out during the creation of the outline, but otherwise identical. In the *control* condition, all constructed their outlines as normal. In the *spatial* and *visual* conditions, participants were asked to carry out a secondary task designed to load on the spatial and visual components of the visuo-spatial sketchpad (VSSP) in working memory (see Baddeley 1986). In the fourth – *interference* – condition, the participants were briefly interrupted by an unrelated stimulus towards which they had to give a brief response.

Four measures were taken to assess knowledge transforming during the construction of the outline: (i) the number of ideas in the list produced before outlining; (ii) the number of new ideas introduced during outlining; (iii) the number of rhetorical headings included in the outline; and (iv) latent semantic analysis (LSA) (Landauer, McNamara, Dennis & Kintsch 2007) was used to assess the semantic similarity between the content produced in the initial list of ideas and the content subsequently produced in the outlines.

There were two important findings. First, the high self-monitors changed the content of their outlines more compared to the initial list of ideas than the low self-monitors did (as indicated by lower LSA scores). Second, the number of new ideas introduced during outlining was reduced in the spatial condition compared to the other conditions.

These results suggest that two factors affect the way ideas are transformed during outlining: (i) rhetorical goals increase the extent to which content is adjusted during outlining; (ii) the spatial component of working memory affects the extent to which this content is differentiated into distinct ideas. Galbraith et al. (2009) argued that outlining involves the construction of a mental model of the text designed to satisfy rhetorical goals. This provides further support for the dual-process model's claim about the role of knowledge transforming in planning. Furthermore, it suggests that this literally involves the construction of a spatially extended knowledge object in working memory.

The key question is how knowledge transformation during outlining is related to text quality. However this was not explicitly addressed by Galbraith et al. (2009). We have therefore carried out a regression analysis on their data, regressing text quality on the idea change measures in the outlines, while controlling for the

Table 1. Regression of text quality on experimental conditions (step 1) and idea change measures (step 2)

	Step 1			Step 2		
	B	SE	β	B	SE	β
(Constant)	5.92	0.47		7.90	1.11	
Visual condition[a]	−1.49	0.62	−0.30[*]	−1.63	0.58	−0.33[*]
Spatial condition[a]	−1.32	0.63	−0.27[*]	−0.73	0.63	−0.15
Interference condition[a]	−1.80	0.61	−0.37[**]	−1.82	0.56	−0.38[**]
Self-monitoring[b]	0.25	0.44	0.06	0.36	0.42	0.09
Number of ideas in initial list				−0.08	0.07	−0.13
Rhetorical headings in outline				0.22	0.10	0.23[*]
New ideas in outline				0.83	0.42	0.26[*]
LSA difference				−6.70	1.77	−0.46[***]

Note. $R = .53$, adjusted $R^2 = .21$, $F(8, 85) = 3.79$, $p < .001$.
[a] dummy coded (control = 0, condition = 1)
[b] dummy coded (low self-monitors = 0, high self-monitors = 1)
[*]$p < .05$, [**]$p < .005$ [***]$p < .001$

number of ideas produced in the initial list and the experimental conditions. The final regression model is shown in Table 1.

Overall, the model accounted for 21% of the variance in quality scores. As can be seen in Table 1, at step 1, all three of the experimental conditions were associated with significantly lower text quality than the control condition. When the idea change measures are added in at step 2 (after the conditions and the number of initial ideas have been controlled for), three factors show significant relationships with text quality: (i) the more rhetorical headings included the outline the higher the text quality; (ii) the more new ideas added to the outline the higher the text quality; (iii) but the more overall change in content between the initial list and the outline the *lower* the text quality.

These results clearly contradict Bereiter & Scardamalia's (1987) knowledge-transforming model. They suggest that, although high self-monitors do change the content of thought more during the construction of an outline, as would be predicted if they adapt their thought to rhetorical goals, this is negatively related to the quality of the resulting text. Instead, the quality of the final text is related to the extent to which relatively *unchanged* content is differentiated into separate ideas and organised in terms of rhetorical goals. This is consistent with the dual-process model's claim that knowledge transformation during planning involves the creation of a coherent mental model of existing ideas which facilitates effective writing but which does not by itself lead to the development of the writer's understanding.

4.2   Effects of outlining on text production

The preceding analysis suggests that, although outlining may improve the quality of writing, this is *not* because it enables writers to transform their knowledge more effectively. Perhaps, however, making one's goals explicit before writing enables the writer to transform their knowledge more effectively during text production itself. Alternatively, the dual-process model would expect outlining to reduce knowledge constituting during text production, and hence to prevent the development of understanding.

In a recent study, Baaijen, Galbraith, and de Glopper (2014) examined the effects of outlining and writing beliefs on the development of understanding and text quality. University students were asked to write an article for a university newspaper and their keystrokes were logged using Inputlog (Leijten & van Waes 2006) to assess the extent to which text was modified in the course of text production.

Writing beliefs were measured using the Writing Beliefs Inventory (WBI) designed by White and Bruning (2005). The WBI measures two sets of writing beliefs: *transmissional beliefs*, which represent a belief that writing involves the transmission of information from authoritative sources to the reader, and *transactional beliefs* scale, which represent the belief that writing is an emotional experience involving the development of understanding as the text is constructed. White and Bruning treat these as equivalent to a contrast between a knowledge-telling model of writing (a combination of high transmissional and low transactional beliefs) and a knowledge-transforming model of writing (a combination of low transmissional and high transactional beliefs). Consistent with the knowledge-transforming model, White and Bruning found that the two sets of beliefs had additive effects on text quality, with low transmissional beliefs and high transactional beliefs being associated with higher writing quality.

Baaijen et al. (2014) questioned this interpretation of the two sets of beliefs, pointing out that they are uncorrelated with one another (as both they and White and Bruning found), and hence do not, on the face of it, correspond to a single dimension. They suggested, instead, that the two sets of beliefs are about different aspects of writing: *transmissional* beliefs are about the content to be written about; *transactional* beliefs are about the process by which text is produced. They suggested that low and high transactional writers take an equally active approach to writing but differ in how they combine planning and text production. In particular, they suggested that transactional beliefs are about how explicit planning and text production processes are combined. Low transactional writers believe that writing should be a controlled process in which predetermined ideas are translated into text; high transactional writers believe that writing involves an interaction with the text in which ideas are developed during the course of text production. Therefore, they hypothesised that outlining will benefit low-transactional writers

because it will enable them to develop a coherent global structure for their text which they can then use to control text production. However, since this operates by facilitating the explicit organising process this will not be associated with a development of understanding. By contrast, it predicts that outlining will be less beneficial for high transactional writers because, although it might facilitate the explicit organising process, in doing so it will reduce the development of understanding during writing.

The findings supported the dual-process model. First of all, they showed that, as can be seen in Figure 2, outlining benefits low transactional writers but not high transactional writers.

This is clearly consistent with the dual-process model. However, a possible explanation in terms of knowledge transforming could be that this is because high transactional writers have a higher working memory capacity and don't need the external support of the outline in order to engage in knowledge transforming. If this is correct, then one would expect that low transactional writers should experience less development of understanding than high transactional writers when writing non-planned texts, but similar levels of development in understanding when they write outline planned texts.

This hypothesis was directly contradicted by the results for the development of understanding, which are plotted in Figure 3.

There are a number of important findings here. First, the low transactional writers did not experience increases in understanding in any of the conditions ($p > .55$ in all conditions), with their scores close to zero in all conditions, and no significant difference between the outline and synthetic conditions ($p > .15$).
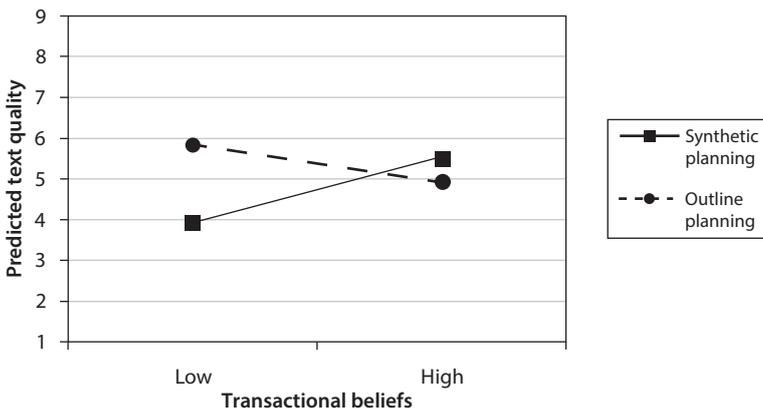
Figure 2. Relationship between transactional beliefs and text quality as a function of type of planning
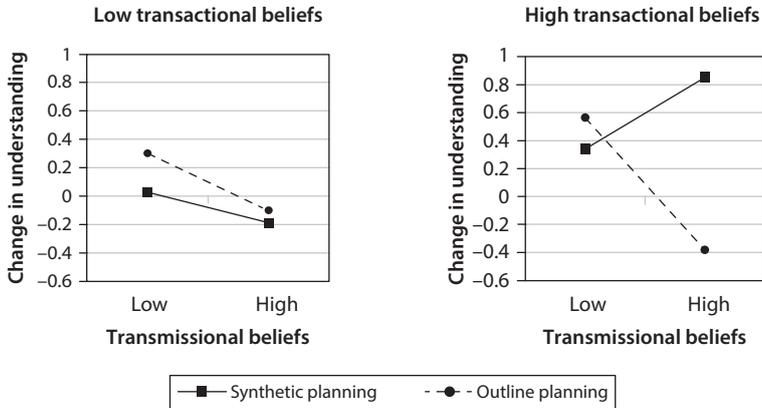
Figure 3. The relationship between transactional beliefs and changes in understanding as a function of transmissional beliefs and type of planning

By contrast, the high transactional writers experienced increases in understanding in all conditions (p < .05) except for when they also had high transmissional beliefs and made an outline before writing. This is compatible with Baaijen et al.'s (2014) claim that low and high transactional writers differ in how they control text production, with high transactional writers allowing their implicit understanding to drive text production whereas low transactional writers use their pre-existing explicit ideas to control text production. The findings also suggest, however, that the high transactional writers' general approach can be overridden when they also try to convey authoritative ideas and reinforce this by making an outline before writing.

For present purposes, the key finding is that, although outlining does improve quality, it only does so for a sub-group of writers, and does not do so by increasing the development of understanding. Quality and the development of understanding do not go hand in hand, contrary to the predictions of the knowledge-transforming model.

The hypothesis about the difference in process between low and high transactional writers received further support from Baaijen et al.'s (2014) analysis of the extent to which writers revised their text during text production. Although low and high transactional writers revised their texts to a similar extent, they differed in how revision was associated with the development of understanding and text quality. For low transactional writers, greater amounts of revision were associated with poorer quality text and unrelated to the development of understanding. This supports the claim that low transactional writers try to get their ideas straight first and then "translate" these ideas into text; revision is essentially *reactive* (see

Galbraith & Torrance 2004), occurring when the text does not adequately express pre-determined explicit ideas, and involves trying to modify the text to fit these ideas, rather than changing the ideas themselves. Outlining enables low transactional writers to create a more coherent set of ideas to guide text production, hence reducing the extent to which text needs to be revised during writing, and improving the quality of the final text. However, since this is based on pre-existing ideas in episodic memory, this does not lead to the development of understanding during text production.

By contrast, for high transactional writers, greater amounts of revision were associated with greater development of understanding and unrelated to text quality. This supports the claim that high transactional writers prioritise the text production process and focus on articulating their implicit understanding in the text, adjusting and developing the global structure of the text in response to the development of their understanding in the text. Hence, their understanding develops in the course of text production, and is associated with a greater amount of revision during writing. Outlining reduces the extent of text modification and hence the extent to which the writer's understanding develops during text production, particularly when the writer is trying to transmit pre-determined content derived from authorities (high transmissional beliefs).

5. Conclusion

In this chapter we have focussed on the contribution of global planning processes – as embodied in outlining – to the creation of written knowledge objects. To conclude, we want to summarise our conclusions about the nature of outlining in writing, and consider its place as a drafting strategy to be taught to novice writers, before considering more general implications for theories of writing.

5.1   The nature of outlining

Our first conclusion is that the main function of outlining is to create a coherent mental model of the text. We argue that this involves trying to create a spatially extended representation of the text to be written. This is based on ideas stored in episodic memory and involves reorganising these into a coherent mental model of the text that satisfies rhetorical goals. This is most effective when content is differentiated into separate ideas and these are organised in terms of the writer's rhetorical goals. To that extent this is a knowledge-transforming process, in which existing content is transformed into a rhetorical object, and the form of this object is the product of an interaction between content and rhetorical spaces. Indeed, we

argue that this literally involves constructing a spatial representation of the object of thought.

However, this does not involve a development of the writer's understanding or perhaps, to be more precise, only involves a partial development of understanding – a better understanding of the global structure of the knowledge object rather than the formulation of novel propositions about the topic. Its effects, therefore, are limited to a specific aspect of the writer's understanding.

The second question here is whether, by making their ideas and goals more structured and explicit before producing the text itself, writers are better able to develop their understanding during writing. Does having a more structured "rhetorical space" enable writers to evaluate and modify content proposed by "content space" more effectively?

Baaijen et al.'s (2014) finding that, when low transactional writers are allowed to make an outline before writing, they improve the quality of their text but do not develop their understanding clearly contradicts this. Furthermore, the fact that improvements in quality and developments in understanding are unrelated to one another contradicts the idea that these are necessarily linked. Instead, these appear to be independent aspects of writing.

The findings for high transactional writers suggest further that outlining is not necessary for effective writing or for the development of understanding during the production of text itself. The high transactional writers in the synthetic planning condition were able to do both. In addition, outlining may, for some writers, actively prevent the development of understanding. Thus, for high transactional writers who also have high transmissional beliefs, synthetically planned writing was strongly associated with the development of understanding whereas, for the same group of writers, outline planned writing was associated with a complete absence of developments in understanding. Given that high transmissional beliefs are that content should be based on authoritative sources rather than personal opinion, this suggests that the negative effect of outlining is conditional on the kind of content that is written about. Our interpretation of this is that, when writing about sources, writers have to formulate the sources in their own terms if their understanding is to develop. Synthetic planning enables them to constitute their understanding in the text, whereas outline planning constrains them to trying to translate pre-determined ideas stored in episodic memory.

Our general conclusion is that although outlining does have beneficial effects on quality for writers who view writing as a translation process (low transactional beliefs), it does not have any beneficial effects on quality for writers who are concerned with developing their understanding during text production (high transactional beliefs). In itself, it does not directly affect understanding, but can influence this negatively through its effects on the text production process.

5.2   Implications for drafting strategies

Outlining tends to be treated as a general strategy that should be useful for all writers – "you must plan your work" is common advice in educational contexts. This is reinforced by a view of writing as primarily a reflective process, in which producing rhetorically effective text and developing one's understanding are treated as two sides of the same coin. If, however, these are not two sides of the same coin then it becomes important to distinguish between the goals of writing. Baaijen et al.'s findings suggest that, although outlining might be a useful strategy for writing effective text, it might be a risky strategy for developing one's understanding. If our purpose is the development of understanding through writing then Baaijen et al.'s results suggest that teaching should target low transactional writer's beliefs, making them aware that a less controlled form of text production can promote the development of understanding. It is an important question for future research as to whether (and how) writers' beliefs about basic components of the writing process can be changed. Second, synthetically planned text production would be a better strategy for developing understanding. This may be particularly relevant when – as in many academic contexts – it is important for students to develop their understanding of authoritative sources. In such circumstances, all writers, rather than just those with high transmissional beliefs, are trying to demonstrate their understanding of these sources.

For writing-to-learn to be effective, then, it is important that writers understand the form of text production that it involves, and that writing is synthetically planned. Future research is needed to establish whether other forms of non-outline planned writing – including completely spontaneous text production – are more or less effective than synthetic planning.

The second set of questions here is about the effectiveness of an outlining strategy compared to other forms of drafting strategy. Kellogg's (1994) research has suggested that outlining is more effective than other drafting strategies, particularly that it is better than a revision strategy, in which an initial unplanned draft is produced and then revised. However, Galbraith and Torrance (2004) have argued that Kellogg's negative findings about the revision strategy are a consequence of the form of revision strategy that he implemented in his experiments. This involved writing an initial draft without pre-planning and without evaluating how well it was expressed, postponing evaluation and revision of expression until after the draft had been produced. Galbraith and Torrance (2004) argued that the key ingredient of a revision strategy should not just be that evaluation of expression is postponed but that the initial draft should not be explicitly organised, and should instead follow the path of thought as it unfolds. The second revision stage

then is concerned not just with revision of expression but with identifying the organisation implicit within the initial draft and rewriting this in a more rhetorically appropriate form.

In Baaijen et al.'s (2014) study, participants were asked to produce a single draft in all conditions, with only the nature of the planning before writing and the writer's beliefs varying. An important question for future research is how these conditions would compare with conditions where a revision strategy of the form specified by Galbraith and Torrance was implemented (see Kieft, Rijlaarsdam, Galbraith & van den Bergh 2006, for some suggestive findings). As things stand, Baaijen et al.'s (2014) findings suggest that outline planning writing is more effective for text quality than synthetically planned writing for low transactional writers, and makes little difference for high transactional writers. One interpretation, therefore, might be that outlining is more effective as a general strategy when the writer's goal is to produce high quality text, particularly in a practical teaching context, where it might be difficult to tailor teaching to different writing beliefs. However, this interpretation neglects the fact that outlining does not develop understanding for low transactional writers and inhibits the development of understanding for some high transactional writers. Furthermore, it assumes that the development of understanding does not contribute to text quality.

In fact, as Baaijen et al.'s (2014) study showed, high transactional writers produced higher quality text than low transactional writers in the synthetic planning condition, and this was precisely the condition in which high transactional writers produced the greatest developments in their understanding, and low transactional writers did not develop their understanding. According to the dual-process model, text quality depends on both the articulation of the writer's understanding and the organisation of the text into a coherent and rhetorically appropriate form. But these are also assumed to be mutually conflicting processes, with explicit organisation inhibiting the development of understanding, and the development of understanding disrupting explicit organisation. In Baaijen et al.'s study, the low transactional writers' approach was supported by an outlining strategy. However, the high transactional writers were not provided with a strategy consistent with their approach. Outlining may have enabled them to better organise their texts – just as it did the low transactional writers – but this was at the expense of the articulation of their understanding in the text. And to the extent that they were still able to develop their understanding, this would be expected to be at the expense of the explicit organisation of the text. Hence, there would be no overall beneficial effect on text quality for these writers. However, if they were allowed to use the revision strategy described by Galbraith and Torrance (2004), they should be able constitute their understanding in the initial draft, and then organise this into a

rhetorically appropriate form in the subsequent revision stage. The prediction of the dual-process model is that this would be of higher quality than writing produced by an outline planning strategy.

5.3   Actions and objects

We have argued in this paper that, in treating writing as a process of constructing knowledge objects, the knowledge-transforming model neglects the extent to which text production is an active knowledge-constituting process, and hence does not capture the way in which understanding develops during writing. Thus, although there is evidence that writers who prioritise rhetorical goals do change the content of their thought more than other writers, and that this occurs particularly when they are able to focus on higher level thinking free from the demands of text production, this is not associated with the development of their understanding. Instead, this appears to involve the reorganisation of existing ideas to satisfy rhetorical goals, as would be expected if it operates on an episodic store of fixed mental objects. Insofar as this has beneficial effects on quality, it appears to be because of its structuring effects rather than because of its effect on the development of understanding. In general, there is no evidence that it has a beneficial effect on the development of understanding, or that the development of understanding is directly linked to the production of better quality text.

We have suggested instead that the explicit structuring effects involved in the construction of knowledge objects need to be combined with the knowledge-constituting process. This provides a mechanism by which new content can be created, and involves treating thought as a form of action. In this view, knowledge objects are the residue of the movement of thought, and are shaped from this residue once thought has been externalised. The development of understanding and the production of effective text are not intrinsically linked, and need to be coordinated in a particular way if they are to be combined effectively.

References

Baaijen, Veerle M., David Galbraith, and Kees de Glopper. 2014. "Effects of writing beliefs and planning on writing performance." *Learning and Instruction* 33: 81–91.
DOI: 10.1016/j.learninstruc.2014.04.001
Baddeley, Alan. 1986. *Working Memory.* New York: Oxford University Press.
DOI: 10.1017/s0033291700025228
Bereiter, Carl. 2002. *Education and Mind in the Knowledge Age.* Hillsdale, NJ: Lawrence Erlbaum Associates.
Bereiter, Carl, and Marlene Scardamalia. 1987. *The Psychology of Written Composition*. Hillsdale, NJ: Erlbaum. DOI: 10.2307/327270

Churchland, Paul M. 2012. *Plato's Camera: How the Brain Captures a Landscape of Abstract Universals.* Cambridge, Massachusetts: The MIT Press. DOI: 10.1007/s11023-012-9295-x

Galbraith, David. 1992. "Conditions for discovery through writing." *Instructional Science* 21: 45–72. DOI: 10.1007/bf00119655

Galbraith, David. 1999. "Writing as a knowledge-constituting process." In *Knowing What to Write,* ed. by Mark Torrance, and David Galbraith, 139–160. Amsterdam, NL: Amsterdam University Press. DOI: 10.5117/9789053563076

Galbraith, David. 2009. "Writing as discovery." *British Journal of Educational Psychology Monograph Series II 6 – Teaching and Learning Writing*: 5–26. DOI: 10.1348/978185409x421129

Galbraith, David, and Mark Torrance. 2004. "Revision in the context of different drafting strategies." In *Revision: Cognitive and instructional processes,* ed. by Linda Allal, Lucile Chanquoy, and Pierre Largy, 63–86. Dordrecht, NL: Kluwer Academic Publishers. DOI: 10.1007/978-94-007-1048-1

Galbraith, David, Sheila Ford, Gillian Walker, and Jessica Ford. 2005. "The contribution of different components of working memory to knowledge transformation during writing." *L1 – Educational Studies in Language and Literature* 5: 113–145. DOI: 10.1007/s10674-005-0119-2

Galbraith, David, Mark Torrance, and Jenny Hallam. 2006. "Effects of writing on conceptual coherence." *Proceedings of the 28th Annual Conference of the Cognitive Science Society*: 1340–1345.

Galbraith, David, Luuk van Waes, and Mark Torrance. 2007. "Introduction." In *Writing and Cognition: Research and Applications*, ed. by Mark Torrance, Luuk van Waes, and David Galbraith, 1–10. Amsterdam, NL: Elsevier. DOI: 10.1075/wll.13.2.07man

Galbraith, David, Jenny Hallam, Thierry Olive, and Nathalie Le Bigot. 2009. "The role of different components of working memory in writing." *Proceedings of the 31st Annual Conference of the Cognitive Science Society*: 3028–3033.

Glasspool, David, W., Tim Shallice, and Lisa Cipolotti. 2006. "Towards a unified process model for graphemic buffer disorder and deep dysgraphia". *Cognitive Neuropsychology* 23(3): 479–512. DOI: 10.1080/02643290500265109

Kellogg, Ronald, T. 1994. *The Psychology of Writing.* New York: Oxford University Press.

Kieft, Marleen, Gert Rijlaarsdam, David Galbraith, and Huub van den Bergh. 2006. "The effects of adapting a writing course to students' writing strategies." *British Journal of Educational Psychology* 77(3): 565–578. DOI: 10.1348/096317906x120231

Klein, Perry D. 1999. "Reopening inquiry into cognitive processes in writing-to-learn." *Educational Psychology Review* 11(3): 203–270.

Klein, Perry D. and Lori C Kirkpatrick. 2010. "A framework for content area writing: Mediators and moderators." *Journal of Writing Research* 2(1): 1–46.

Landauer, Thomas K., Danielle S. McNamara, Simon Dennis, and Walter Kintsch. 2007. *Handbook of Latent Semantic Analysis.* Mahwah, NJ: Lawrence Erlbaum Associates. DOI: 10.1177/0261927x08317982

Leijten, Marielle, and Luuk van Waes. 2006. "Inputlog: New perspectives on the logging of online writing." In *Computer keystroke logging and writing: Methods and applications*, ed. by Kirk P.H. Sullivan, and Eva Lindgren, 73–93. Amsterdam: Elsevier.

McClelland, James. L., Bruce, L. McNaughton, and Randall C. O'Reilly. 1995. "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory." *Psychological Review* 102: 419–457. DOI: 10.1037//0033-295x.102.3.419

Norman, Kenneth A. 2010. "How hippocampus and cortex contribute to recognition memory: Revisiting the Complementary Learning Systems model." *Hippocampus* 20(11): 1217–1227. DOI: 10.1002/hipo.20855

O'Reilly, Randall C., Rajan Bhattacharyya, Michael D. Howard, and Nicholas Ketza. 2011. "Complementary Learning Systems." *Cognitive Science* April 2011: 1–20. DOI: 10.1111/j.1551–6709.2011.01214.x

Popper, Karl R. 1972. *Objective Knowledge*. Oxford: Oxford University Press. DOI: 10.1007/bf02381520

Rogers, Timothy, T., and James, L. McClelland. 2004. *Semantic Cognition*. Cambridge, MA: MIT Press.

Snyder, Mark 1979. "Self-monitoring processes." *Advances in Experimental Social Psychology* 12: 86–131. DOI: 10.1016/s0065-2601(08)60260–9

White, Mary J., and Roger Bruning. 2005. "Implicit writing beliefs and their relation to writing quality." *Contemporary Educational Psychology* 30: 166–189. DOI: 10.1016/j.cedpsych.2004.07.002

Winocur, Gordon, Morris Moscovitch, and Bruno Bontempi. 2010. "Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions." *Neuropsychologia* 48: 2339–2356. DOI: 10.1016/j.neuropsychologia.2010.04.016

Analyzing writing process data

A linguistic perspective

Mariëlle Leijten[*], Luuk Van Waes[**] & Eric Van Horenbeeck[**]
[*]University of Antwerp & Research Foundation – Flanders (FWO)
[**]University of Antwerp

In this paper we briefly introduce keystroke logging as a research method in writing research, focusing more explicitly on the recently developed linguistic analysis technique. In a case study of two elderly people (healthy versus demented), we illustrate some aspects of this linguistic approach. This analysis aggregates event-based data from the character level to the word level, while taking into account all the revisions that occurred during the composing process. The linguistic process analysis complements the logged process information with results from a part-of-speech tagger, a lemmatizer, a chunker, a syllabifier, and also adds word frequencies. The enriched word level information – together with action time and pause time at the word level – opens up new perspectives in the analysis of process dynamics, once more establishing a closer link between process and product analysis. We thus test the complementary diagnostic accuracy for Alzheimer's disease, mainly focusing on cognitive and linguistic aspects that characterize the process of written language production.

Keywords: keystroke logging; linguistic process analysis; Alzheimer's disease; Inputlog

1. Introduction

Writing research has a rich tradition of research on writing products as well as on writing processes (cf. recent overviews in Berninger 2012; MacArthur, Graham & Fitzgerald 2008; Bazerman 2008; Bazerman et al. 2010). Since the 1980's, there has been a growing focus on process research due to the increase in interest in cognitive psychology. The introduction of the writing model (Flower & Hayes 1981) by Hayes and Flower opened up new areas of research. The first model was based on protocol analysis and mainly provided insights into the internal mechanisms necessary for writing (e.g. memory, planning,

problem solving) (Hayes 2012b, a). In the 1996 model the writing medium was the subject of renewed attention (Hayes 1996), mainly due to the fact that the computer gradually became the standard for text production. However, not only have the methods of text production changed considerably, the technical possibilities for studying writing have also evolved. For instance, keystroke logging and eye tracking have been implemented as observation and research tools enabling us to gain a better understanding of the cognitive processes involved in writing. Although there is an increasing interest in and focus on real-time processes, we think it remains very important trying to establish a link between a writer's observed mental processes and the textuality of writing from a product perspective. In this section, we focus on a writing process study in which we use keystroke logging data to specifically examine the crossroads at which the linguistic characteristics of the written product and the writing process itself meet.

2.   Keystroke logging

Keystroke logging is a widely used and popular method in writing research. One reason is undoubtedly the fact that it is an unobtrusive method for studying underlying cognitive processes and scarcely interferes with the natural writing process (Sullivan & Lindgren 2006; Leijten & Van Waes 2012; Van Waes et al. 2012). In addition, it is also possible to combine it with complementary observation techniques, like thinking aloud or eye-tracking. Moreover, keystroke logging enables researchers to collect fine-grained pause and revision data and may therefore make it possible to analyze writing processes from a wide range of perspectives. Keystroke logging has been widely used in cognitive writing process research in the broadest sense, for instance in domains like writing development, second language learning, developmental language disorders such as dyslexia, translation, professional writing, on-line writing, etc. An increasing number of studies now report keystroke logging research experiments (e.g. Gunawardhane et al. 2013; Van Waes, Leijten & Remael 2013; Baaijen, Galbraith & de Glopper 2014; Robert & Van Waes 2014; Wininger 2014; Doherty & O'Brien 2014) or describe specific aspects of the research method itself (Ehrensberger-Dow & Perrin 2009; Jakobsen 2011; Baaijen, Galbraith & de Glopper 2012; Galbraith and Baaijen this volume). In addition, there are a number of recent articles focusing on theory development (Leblay & Caporossi 2014; Caporossi & Leblay 2011; Leblay & Caporossi this volume; Macgilchrist & Van Hout 2011; Miller, Lindgren & Sullivan 2008; Van Waes & Leijten 2013; Risku, Windhager & Apfelthaler 2013; Leijten et al. 2014).

In Europe, three free keystroke logging programs are available, each focusing on specific niches: ScriptLog, Translog, and Inputlog.

2.1 Experimental research into writing processes: ScriptLog (www.scriptlog.net)

ScriptLog (Wengelin et al. 2009) was developed by researchers at the universities of Gothenburg, Lund (Sweden) and Stavanger (Norway) for the study of writing processes. It was originally a Macintosh program, then a Windows program, and at the time of writing, a new platform-independent (Windows, MacOS, Linux) version is being tested (ScriptLog 2013: Johansson et al. 2014).

ScriptLog creates a writing environment with a build-in text editor and makes it possible to incorporate frames for different types of elicitation material, such as pictures, texts, movie clips or sounds (for example for dictation experiments). The new version includes extra experimental facilities that enable researchers to set up different writing experiments, for example using dual/triple-task paradigms. The set-up of the environment is controlled in a design module. When activated, ScriptLog keeps a record of all keyboard events, the exact screen position corresponding to these events, and their temporal distribution.

Like other keystroke logging programs, ScriptLog allows the researcher to play back a recorded session – or a selected extract from it – in real time on the basis of the log file. In addition, the analysis module enables the researcher to analyze time distributions across the writing process both for predefined patterns and for user-defined patterns, for example for a particular word string or for a regular expression. Finally, ScriptLog allows researchers with access to an eye tracker to enhance the study of the interplay between writing, monitoring (reading) and revision by integrating eye tracking data. (Currently only SMI eye trackers, more models will be added.) Data on the distribution of visual attention during writing help, for instance, to determine the extent to which pauses are used for monitoring. Data gathered via ScriptLog can now be converted to the Inputlog XML format, thus enabling researchers to conduct Inputlog analyses on ScriptLog data.

2.2 Writing research in translation settings: Translog (www.translog.dk)

Translog was developed at the Copenhagen Business School (Denmark) with the primary aim of studying the writing processes of translators translating a source text from one language into another (Jakobsen 2006). Since 2012, a newly programmed version has been available under the name Translog-II (Carl 2012). It has two interdependent components: a *Translog-Supervisor* component to prepare

a logging project and replay it afterwards, and a *Translog-User* component. The latter displays a predefined source text in the top half of the window together with a translation frame in which the target text can be edited. Translog can be used in combination with various eye tracking programs (Tobii 1750 and EyeLink 1000). Since translation research has its own specificities, Translog-II focuses mainly on providing visualization options to show the relation between the source and the target text.

2.3 Writing research in educational and professional settings: Inputlog (www.inputlog.net)

Inputlog was developed at the University of Antwerp (Belgium) to log writing processes in both ecological and experimental settings (Leijten et al. 2014; Leijten & Van Waes 2013). The program logs all keyboard and mouse events in every Windows environment. In the case of texts written in MS Word, extra characteristics relating to the input events are logged to permit fine-grained writing analyses (see below). The program also logs text production with speech recognition systems (Dragon Naturally Speaking, Nuance) and tracks copy-and-paste actions that relate to the use of external digital sources (e.g. the internet).

Inputlog 6.0 features five modules:

1. *Record*: This module logs (keyboard, mouse, and speech) data in Microsoft Word and other Windows-based programs and assigns a unique time stamp (ms) to the data.
2. *Pre-process*: As it is often necessary to prepare and clean up logged data prior to analysis, this module makes it possible to process data from various perspectives: event-based (keyboard, mouse, and speech), time-based or based on changes between Windows (sources: MS Word, Internet etc.). The filter provides an easy way to delete 'noise' at logging session start-up or shut-down. For example, if additional questions are asked at the beginning of the period of observation when the logging session has already started, this pause time (noise) can be excluded from the data analysis.
3. *Analyze*: This module is the heart of the program. It features three process representations (the general and linear logging file and the S-notation of the text) and four aggregated levels of analysis (summary, pause, revision, and source analyses). Additionally, a process graph can be produced. The current version also offers a linguistic process analysis which returns the results from

a part-of-speech tagger, a lemmatizer and a chunker, as well as the syllable boundaries and word frequencies (cf. below).

4. *Post-process*: This module integrates single or multiple log files from Inputlog or other observation tools (Dragon Naturally Speaking and eye tracking data). It is also possible to merge multiple output files for further analysis in, for instance, SPSS, R or MLW in.

5. *Play*: This module allows researchers to play back the recorded session at various levels (time or revision-based). The playback is data-based (not video-based) and the playback speed is adjustable. A logged session can also be reconstructed revision by revision.

The described keystroke logging programs are distributed for free for non-commercial use to researchers and teachers (for a general overview of keystroke logging tools and their characteristics, please see www.writingpro.eu).

3.   Linguistic analysis with Inputlog 6.0

Recently Inputlog has implemented a linguistic analysis in which linguistic information is added to the process data (Leijten et al. 2012). As described, the data output from keystroke loggers is mainly based on capturing each character and mouse movement as an isolated event. However, we are now able to aggregate the logged process data from the letter level (keystroke) to the word level. This has allowed us to merge the word-level output with existing lexical and Natural Language Processing (NLP) tools. The logged process data is annotated with different kinds of linguistic information: part-of-speech tags, lemmata, chunking, named entity information, syllabification and frequency information. We will describe these in more detail below. We start with a short overview of the programming challenges we have had to deal with and provide a description of the linguistic analysis flow: what steps are needed to transform character level process analyses into word-level process analyses?

3.1   Aggregating log data from character level to word and sentence level

A number of challenges have to be addressed before the log data of Inputlog can be aggregated to the word level (or higher):

1. First, the concept of a 'word' or a 'sentence' does not exist in the log file; these items have to be reconstructed because the atomic unit is a key press, a mouse movement, a button click.

2. Second, text production is characterized by revisions to the previously pro-
duced text and these may occur either immediately or be initiated at some
distance from the word to be revised.

3. Third, additions and deletions can be nested, occurring not in situ but decided
on away from the target.

To cope with the non-linearity of writing processes, it is necessary to map the shift-
ing and changing events to the locations where the effects are generated. This can
be done using S-notation. S-notation (Severinson Eklundh & Kollberg 2002; Van
Horenbeeck et al. 2012) contains information about the types of revision (inser-
tion or deletion), the order of these revisions, and the breaks in the text where the
writing process was interrupted.

Consider the following French sentence at the end of a writing process:

(1)   "*Des questions sur la science, sur la science et sur l'évolution. Fin.*"

Figure 1 shows the test sentence (1) that we are studying together with all the
changes rendered in the S-notation.

Des·questions·sur·la·s[x]^1|$_1$cience,·sur·la·science·et·sur{·l'évolution}^{4}·[le·progrès]^2|$_3${.}^{3}|$_4$·Fin.|$_2$·

Figure 1. Example of S-Notation

Square brackets indicate a deletion, curly braces an insertion and the verti-
cal pipe symbol, called a 'break', is used to mark the position at which the pro-
cess was interrupted. The subscript numbers next to the pipe symbol have a
corresponding superscript number at either an insertion or at a deletion. In this
example: the word 'l'évolution' is surrounded by curly braces indicating that
it has been inserted. The insertion is indicated by superscript number 4. This
means that it was the 4th revision out of a total of 4 interventions. The vertical
pipe symbol with subscript 4 appears before the last word of the sentence and
marks the position where the author decided to insert 'l'évolution' instead of 'le
progrès', a word that has been deleted as indicated by the square brackets sur-
rounding it.[1]

1. The French sentence is a translation of an English example taken from the Inputlog
manual (Leijten & Van Waes 2014).

3.2   Finding words in a stream of events

Inputlog includes a regular expressions-based tokenizer that divides sequences of characters into words and sentences while aggregating the pause time from the individual events. The program uses two data sets: the event log (a file with the 'idfx' extension) used to identify the word boundaries and then, subsequently, a computed list of revisions and edits calculating the pause and the action time once the words have been reconstructed. The main processing method creates a token object that contains the word in its final appearance, its revisions, its position in the final text string, all pause time information (before, within and after words), and the type of processing used for the word (normal production, revised word, deleted word).

For every revision, the process first defines whether it is 'immediate' or 'delayed' (Leijten, De Maeyer & Van Waes 2011). An immediate revision is defined as a revision that takes place before a new word is produced; a delayed revision, in contrast, is characterized by the fact that one or more words are produced between the decision to change the text (the 'break') and the actual deletion or insertion. Initially, every token is assigned the 'normal production' marker which is used for linear text production. However, if the software detects at a later stage that the word contains revisions, the process type is changed to 'revised'. Line feeds, tabs, and other non-characters are replaced with a placeholder to make them visible.

Insertions or deletions may extend over many events, e.g. [le progrès]from the example is a deletion of nine characters and one space. Because the characters are presented as the separate words 'le' and 'progrès', opening or closing symbols inserted by the S-notation are missing. In this case the first word has the opening bracket: ['le' and the second word the closing bracket: 'progrès']. The program adds the missing symbol and returns [le] and [progrès]. Another issue is that the linguistic analysis expects all punctuation symbols to be detached from the front or the back of a word and to be saved as separate tokens. Hence, the word 'Fin.' with the end-of-sentence dot at the end of the example sentence is rendered as 'Fin .' with a space between the word and the endpoint.

Finally, each word is accompanied by timing information such as start time and end time, the word action time and pause time (Figure 2). When the final word is the result of one or more revisions (deletions, insertions) then the pause and action times of the editing events are added to the word production time. Revisions are presented at the word level in combination with the before-word and after-word pauses and the word production time.

Revisions	S-notation	#Chars	Token	Part of Speech	Chunk	Lemma	Syllable	BfrWord-2	BfrWord-1	Word Prod	Within WordPause	AftWord+1
	Des	3	Des	DT	B-NP	de	de	0	0	608	530	141
	questions·	9	questions	NNS	I-NP	question	question	141	312	1731	1653	172
	sur·	3	sur	IN	B-PP	sur	sur	172	187	421	343	156
	la·	2	la	DT	B-NP	la	la	156	250	250	172	234
1-I	s[x]cience··	9	science	NNP	I-NP	science	science	234	218	5304	5242	187

Figure 2. Example of the timing information on word level for a part of the example French sentence considered here

3.3   Flow of the linguistic analysis

S-notation makes a vital contribution by moving the logging process from character to word level. Once words and sentences are recognized, all the tools developed for the NLP framework become available to conduct linguistic analyses on the process data, including word-level revisions and deleted fragments. To support the linguistic analysis of the writing process, a client-server version of Inputlog has been developed (Leijten et al. 2012). Although the logging process and most analyses can be performed using a desktop version of Inputlog, the linguistic modules need to interact with specialized programs, sometimes backed up by very large data files. By decoupling data capture from the analytics, it is possible to add in any NLP module that receives Inputlog data via a communication layer. A workflow procedure presents the data in sequence to the different NLP packages and collects the final output. Because all data traffic is performed using simple text files, cooperation between different software packages is conceivable. The extension has been developed for English and Dutch in order to provide a proof of concept, but it is largely language-independent.

The LT3 Linguistic Preprocessing toolkit developed by the Language and Translation Technology Team of the University of Ghent (LT3serv.ugent.be) is currently used as the main toolkit (Figure 3).[2] The different linguistic tools are installed on an LT3 web server. The output is a tab-delimited string in UTF-8 containing the following fields: token, part-of-speech, probability of the part-of-speech tag, lemma, probability of the lemma, chunk information, named-entity, probability of the named entity, absolute frequency, relative frequency, and

2.   These toolkits are also available for French and German.

syllabification (The manual belonging to Inputlog 6.0 has more details on the different components and the tags used for the part of speech tags and the chunks: Leijten & Van Waes 2014).
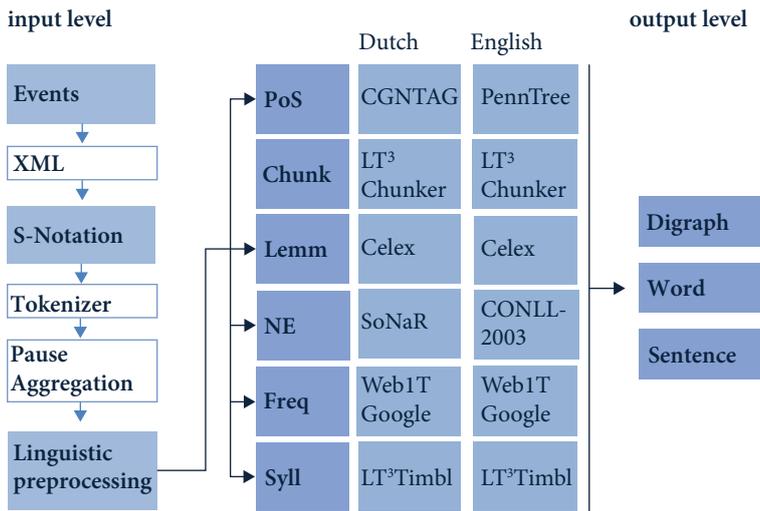
Figure 3. Schematic representation of the flow used in the linguistic analysis performed by Inputlog 6

3.4 Part-of-speech (PoS) tagger

Part-of-speech tagging (PoS tagging), also called grammatical tagging or syntactic word-category disambiguation, is a piece of software that reads text and assigns an appropriate part of speech to a word such as noun, verb, adjective, etc. based on both its definition and its relationship with adjacent words (e.g. Part of Speech column in Figure 2). Because many words have more than one syntactic category, the tagger tries to determine which of the syntactic categories is the most likely for a particular use of a word in a sentence (Manning & Schütze 1999). The English PoS tagger uses the Penn Treebank tag set which contains 45 distinct tags. The Dutch part-of-speech tagger uses the CGN tag set codes which are characterized by a high level of granularity (Van Eynde, Zavrel & Daelemans 2000).

3.5 Chunker

Text chunking combines syntactically related consecutive words into non-overlapping, non-recursive chunks on the basis of a fairly superficial analysis. The LT3 chunkers are rule-based and contain a small set of constituency and distituency rules. Constituency rules define the part-of-speech tag sequences that can

occur within a constituent (such as preposition + noun) while distituency rules define the part-of-speech tag sequences that cannot be adjacent within a constituent (such as noun + preposition). The chunks are represented by means of IOB tags (See Figure 2). In the IOB-tagging scheme, each token belongs to one of the following three types: I (inside), O (outside) and B (begin); the B and I tags are followed by the chunk type, e.g. B-VP, I-VP (Jurafsky & Martin 2009).

3.6   Lemmatizer

The base form (lemma) for each orthographic token is generated during lemmatization. For verbs, the base form is the infinitive. For most other words, the base is the stem, i.e. the word form without inflectional affixes. The lemmatizers make use of the predicted PoS codes to disambiguate ambiguous word forms. For instance 'Paris' can be a city or a person. It is classified as a city, for instance, when it is preceded by a preposition of place (bought in) and not by a preposition of possession (bought from). The lemmatizers were trained on the English and Dutch parts of the Celex lexical database, respectively.

3.7   Named entity recognizer

Named entity recognition (NER) sequences of words in a text that belong to predefined NER categories such as names of persons, products, or locations are identified in a text. The Dutch NER system is trained on the 1-million-word subset of SoNaR and identifies the following six NER categories: person, organization, location, product, event, and miscellaneous. The English NER system is trained on the CONLL-2003 shared task data and discerns four NER categories: person, organization, location, and miscellaneous.

3.8   Frequency

Word-frequency information for English and Dutch is retrieved from frequency lists derived from the Web1T Google corpus which is available from LDC.[3] The frequency lists contain the 2 million most frequent words in Dutch and English. The word frequencies are presented both as absolute frequencies and relative frequencies (expressed as percentages).

3.9   Syllabification

Syllabification was approached as a classification task: a large instance base of syllabified data was presented to a classification algorithm which automatically

---

3.   Details about the data set can be found at www.ldc.upenn.edu/Catalog/docs/LDC2006T13/readme.txt.

learned the patterns needed to syllabify unseen data. The syllabification tools were trained on Celex using Timbl as classification algorithm.

We will illustrate the concept of linguistic analysis on the basis of a case study taken from a writing research project investigating the cognitive characteristics of people with Alzheimer's disease.

4. Cognitive writing process characteristics in Alzheimer's disease

A large variety of neuropsychological tests are available for the diagnosis of Alzheimer's disease (AD). In some of these tests, linguistic processing – both oral and written – is an important factor. Language disturbances might serve as a strong indicator of an underlying neurodegenerative disorder like AD. However, the current diagnostic instruments for language assessment mainly focus on product measures, ignoring the importance of the process that leads to written or spoken language production. A more process-oriented approach should allow researchers to describe and analyze writing data from a temporal perspective, focusing on motor, cognitive, and linguistic aspects. Keystroke logging data (writing), potentially complemented by eye tracking data (reading while writing), provides an excellent basis for an adequate description of these processes.

To our knowledge, this is the very first project to test whether cognitive and linguistic aspects that characterize the process of written language production could provide *a complementary and accurate diagnosis* of AD.

4.1 Participants

Three groups of participants were involved in the study:

1. Patients with mild dementia due to AD (n = 5),
2. Patients with mild cognitive impairment (MCI) due to AD (n = 8),
3. A group of age-matched cognitively healthy elderly persons (n = 20).

The patients were recruited from the Memory Clinic of the Antwerp, Middelheim and Hoge Beuken Hospital Network (ZNA), Belgium. All the patients were diagnosed by Prof. Dr. Engelborghs and underwent an extensive neuropsychological examination (Van der Mussele et al. 2012).[4]

---

4. The AD patients met the NINCDS-ADRDA criteria of probable AD (McKhann et al. 1984). Their mini-mental state examination (MMSE) score (Folstein et al. 1975) was above 20 (mild dementia). MCI patients met the criteria of Petersen et al. (2004), and also complied with the new diagnostic criteria of 'MCI due to AD' (Sperling et al. 2011).

4.2   Task

The three groups of participants were instructed to write two short descriptive texts on a computer. We opted to use two figurative elicitation tasks (see Figure 4a and b) which are part of standardized aphasia test batteries (Goodglass, Kaplan & Barresi 1983; Mesulam et al. 2003; Visch-Brink et al. 2014). On the basis of this picture, the participants produced a brief text in which they described the scene presented to them. To evaluate consistency of task execution, we used two comparable scene pictures, while picture elicitation was counterbalanced to avoid order effects.

Figure 4. (a-left) 'Kitchen' task by Goodglass and Kaplan (1983); (b) 'Living room' by Visch-Brink et al. (2014)

4.3   Case study

In this study we describe the cognitive processes that characterize the text production of the participants in a controlled task environment. We will first describe the results of some more general process analyses using standard measures employed in keystroke logging research. These include, for example, time taken to perform the task, active writing time, number of pauses and mean length of pauses at different levels, as well as the product/process ratio (i.e. proportion between product and process measures. We will then introduce certain linguistic and product measures in order to control more precisely for word and phrase characteristics that might influence pausing behavior. Since this kind of automated linguistic data analysis has not yet been fully tested for difficulties in aggregating and filtering, we will present a case study in which we selected two participants. We selected one healthy elderly woman (Elise*, 81 years old) and one woman with dementia (Mary*,[5] 79 years old). The participants were matched on three levels: age, education, and career. Both women were about 80 years old, had attended school until

---

5.   *The names of the participants were changed for privacy reasons.

they were 19 years old and they had worked in jobs requiring them to type texts. Readers should note that the main aim of this paper is not to identify differences between the two participants. Instead, the main reason for presenting this case study is because we want to explore the potential value of adding a linguistic perspective to writing process research, and pause analyses in particular, and investigate whether the two approaches can complement one another.

In the same way as in spoken language, we expected that cognitively impaired elderly persons would take longer to produce a (shorter) picture description. Consequently, we expected the proportion of active writing time relative to pausing time to decrease between the healthy elderly and the cognitively impaired elderly (Schilperoord 1996; Van Waes & Schellens 2003). Table 1 gives an overview of some process indicators characterizing the writing processes of the two participants.

Table 1. Mean product, process, and pause characteristics of both picture-depicting tasks

	Elise (healthy)	Mary (demented)
Product information		
Number of words in final text	56	41
Number of words in final text (per minute)	11.76	6.73
Process information (pause threshold: 2000 ms)		
Process time	0:04:46	0:05:58
Total pause time	0:02:03	0:03:45
Percentage active writing time (%)	56.65	36.73
Mean number of pauses	24.50	38.50
Mean pause duration (in seconds)	5.08	5.86
Median pause duration (in seconds)	3.23	3.82
Number of characters produced (incl. spaces)	328.5	236
Number of characters produced per minute (incl. spaces)	68.86	38.77
Product/process ratio	0.95	0.99
Mean words produced per sentence	24.17	42.00
Mean word length per sentence	4.59	4.68

The results indicate that Mary (demented – d) took about a minute longer to write the descriptive texts and that her final texts were on average 15 words shorter than Elise's (healthy – h). Thus, compared to Elise, she produced about half the number of words per minute (Elise: 11.76 vs Mary: 6.73). This was due mainly to the amount of pausing time: if we consider the pause analysis based on a threshold of 2 seconds, then Elise(h) paused 25 times on average in both writing tasks, while

Mary(d) paused about 39 times. Consequently, Elise(h) exhibited 20% more active writing time than Mary(d). The average length of their pauses was about 5–6 seconds. The fact that the product/process ratio was close to 1 shows that both writers performed almost no revision. The data also show that the number of words produced per sentence is in itself not a very reliable measure. The number of words produced was about the same as the total text length, indicating that Mary(d) did not use sentence markers. Therefore, pauses within and between words will be a more reliable metric.

In addition to the general pausing behavior, we expected that the mean pause length *within words* and *between words* would help us to further discriminate the healthy elderly from the cognitively impaired elderly (Wengelin 2006; Kellogg 2008; Lindgren et al. 2011). Table 2 (top – Threshold of 2 seconds) shows that Mary(d) made almost twice as many pauses within words as Elise(h) and that the pauses were on average 3 seconds longer. If we aggregate the pauses between words (pause after a word + pause before a word; Leijten & Van Waes 2014) then Elise(h) paused about 43 times and Mary(d) about 29 times at the between-word level. Individual pauses might be below the chosen threshold, but taken together they might exceed the threshold and become relevant (See Figure 5: AW: after words; BW: before words; ww: within words).

However, if we focus only on pauses before words, then Mary(d) made twice as many individual pauses of longer than 2 seconds than Elise(h). The length of individual pauses was about 4 seconds.

translation		the			cakebox									
inputlog events	–	d	e	–	k	o	e	k	e	n	d	o	o	s
pause time (in ms)	374	297	312	343	1810	218	343	203	203	249	500	312	296	8877
pause location	AW	BW	ww	AW	BW	ww	ww	ww	ww	ww	ww	ww	ww	ww
summed pauses	sum (671)			sum (2153)										

Figure 5. Example of aggregated between-word pauses for Elise(h) in boxes (AW = after-word pause; BW = before – word pause)

Although in writing research a pause threshold of 2 seconds is quite common, we should be aware of the fact that this causes a lot of data loss, especially at the between and within-word level. When focusing on higher level processes, this is not problematic. However, in the current study we are also interested in more low level processes. Therefore, we lowered the threshold to 200 ms and recalculated the same parameters. This added about 20% more data for the healthy elderly and 25% for the cognitively impaired elderly participant. Following this manipulation,

Table 2. Description of mean pause characteristics in both picture-depicting tasks

	Elise (healthy)	Mary (demented)
Pause information (threshold 2s)	0:02:03	0:03:45
Number of pauses within words	8.00	14.50
Mean pause duration within words (s)	4.67	7.66
Number of pauses between words	43.00	29.00
Mean pause duration between words (s)	6.08	4.97
Number of pauses before words	8.50	20.00
Mean pause duration before words (s)	3.98	4.61
Number of pauses after words	5.50	3.00
Mean pause duration after words (s)	3.10	3.96
Pause information (threshold 0.2s)	0:04:40	0:05:57
Mean number of pauses	349	246
Mean pause duration (in seconds)	0.80	1.59
Median pause duration (s)	0.38	0.66
Number of pauses within words	209.5	153
Mean pause duration within words (s)	0.63	1.47
Median pause duration within words (s)	0.37	0.61
Number of pauses between words	53.5	39
Mean pause duration between words (s)	1.74	2.99
Number of pauses before words	66	52.5
Mean pause duration before words (s)	0.99	2.45

Elise(h) had about 100 more pauses than Mary(d), but her pauses within words were of a mean duration of 600 ms while the pauses made by Mary(d) lasted about twice as long (1470 ms). About 15% of these pauses above the threshold of 200 ms were between words. Again the mean pause duration for Mary(d) was more than 1 second longer.

The above-mentioned measures are common in writing process research (mean pause length within and between words, burst length, process/product ratios). However, using the data from the linguistic analyses we can further refine the concept of 'pause location', especially at the between-word level. The general pause data revealed a difference in the way the two participants dealt with pauses before and after words. We expect that focusing on the pause behavior associated with specific word categories will reveal useful additional features enabling us to further differentiate our observations relating to pre- and post-word pauses. The related literature tells us, for instance, that the elderly in general find it more

difficult to choose the correct verb than the correct noun (Yi, Moore, & Grossman 2007).

In the linguistic analysis, pauses are represented in three different ways: Before-WordPause2 (i.e. the pause immediately following the previous word: technical term 'after word pause'), BeforeWordPause1 (i.e. the pause immediately preceding the word), and AfterWordPause (i.e. the pause immediately after the last character of the word). The 'between word pauses' are therefore calculated as the sum of the BeforeWordPause2 and BeforeWordPause. To a certain extent, this resembles the definition of between-word pauses in handwriting, which are defined as the time it takes to lift the pen when ending a word and starting a new one.

translation	I			see				a			
inputlog events	i	k	–	z	i	e	–	e	e	n	–
pause time (in ms)	0	358	1124	6364	546	1061	1155	1310	312	437	1341
pause location	BW	ww	AW	BW	ww	ww	AW	BW	ww	ww	AW
summed pauses			sum (7488)				sum (2465)				

Figure 6. Example of general analysis including pause information (Mary(d)) in boxes (AW = after-word pause; BW = before-word pause)

In Figure 6 we see a coded fragment in which Mary paused for 7488 ms between the words 'ik' [I] and 'zie' [See]. Mary paused for 1124 ms after the production of 'ik', and she paused for 6364 ms before she started to produce the word 'zie'. One of the main challenges in keystroke logging research is to better understand these kinds of pause patterns. Are both pauses related to the production of the next word? Is there a gradual shift of attention, e.g. from the evaluation of the previous word to the planning (and translation) of the next word (Maggio et al. 2012)? The distribution of between-word pauses is also partly determined by personal preference. Some people systematically seem to pause for longer immediately after a word than before a word. By contrast, Mary exhibited a pattern of longer pauses immediately preceding the word compared to the length of her after-word pauses (see Figure 6 and Figure 7).

S-Notation (Dutch)	S-notation (translation)	BeforeWord 2	BeforeWord1	AfterWordPause
ik-	I	0	0	1124
zie-	see	1124	6364	1155
een-	a	1155	1310	1341
kind-	child	1341	2824	499

Figure 7. Example of linguistic analysis including pause information (Mary(d))

Table 3 presents the basic pausing information from the linguistic analysis. This analysis complements the pause analysis data previously presented in Table 2. In Table 2 we reported an average of 53.5 pauses between words for Elise(h) and 39 for Mary(d) for the 0.2 ms pause threshold. However, if we fine-tune the pause analysis for the conduct of our linguistic analysis, we can look in greater detail to the 99 pauses for Elise(h) and 60 for Mary(d) in both writing tasks. Since we decided to focus on those pausing times that clearly indicate cognitive effort related to *producing* a word, we excluded revisions from the current evaluation since they disrupt the data by introducing cognitive effort of a different kind. We also removed extremely long pauses of more than 10 seconds (2 in the case of BFW-1 and 5 for variable BFW-2). Finally, we had to manually correct the automated word reconstruction of Inputlog in a few instances. Examples of such corrections are incorrectly connected words (*halende ~ halen de*) and grossly misspelled words (*kantwkanteken ~ kantelen*). As a result of this intervention, the number of pauses in Table 3 differs slightly from the numbers and means mentioned in Table 2.

Table 3. Mean pause duration before words (–1 and –2)

	Elise (healthy)	Mary (demented)
Pause information (threshold 0.2s)		
Total number of pauses	96	43
Mean duration Before Word Pauses (sum of –1 and –2)	1718	2661
Mean duration Before Word Pause (–1)	817	1958
Mean duration Before Word Pause (–2)	901	704
Mean duration After Word Pause (–2)	837	749

The pauses between words (before-word pauses –1- and –2) were about 1 second shorter for Elise(h) than for Mary(d). The summed pauses for Elise(h) consisted of two pauses of comparable length, whereas the pauses for Mary(d) were more than twice as long as the preceding pause (–2) just before a new word was produced (–1).

Figure 8 shows the number and mean of the most frequently used word categories (The information on pausing times is presented in Table 5 in the Appendix). By selecting word categories that were used at least 5 times, we provide an overview of more than 90% of the data for each participant (Elise(h): 93.75%, and Mary(d): 90.70%). The difference between the two participants is due to the fact that Elise(h) regularly used connectives (4) and adjectives (7) in her text, whereas only one adjective occurred in Mary's text. The remainder of the infrequently used

word categories were adverbs and unspecified tokens (spec). (An overview of the word categories identified by the linguistic analysis is provided in the Inputlog manual (Leijten & Van Waes 2014)).
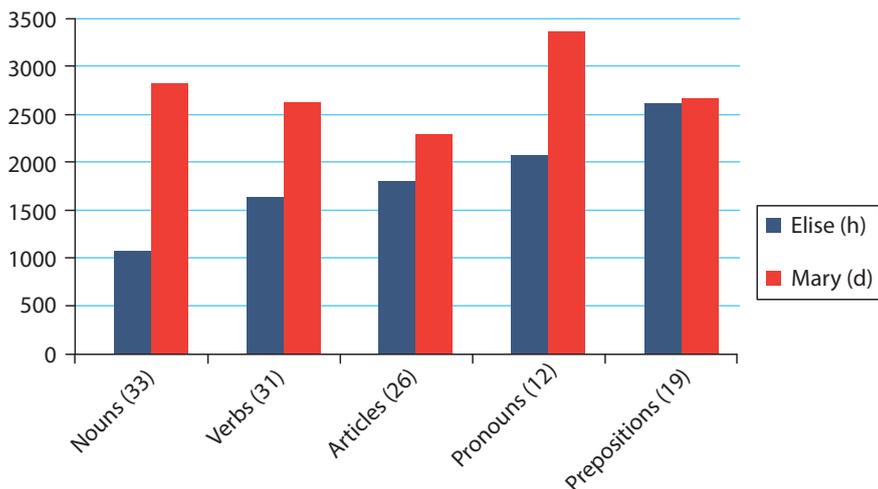
Figure 8. Number of between word pauses and mean pause duration before words per word category

The least demanding word category for Elise(h) seems to have been nouns (1077 ms), with the pause length lengthening gradually from verbs to articles and then on to pronouns. On average, Elise reflected for longest (2630 ms) in the case of prepositions, which often introduced more extensive prepositional phrases including articles. This same hierarchy is not reflected in Mary's data. The differences between the word category-related before-word pauses fluctuated less but were still in all cases longer than those produced by Elise. In particular, nouns, verbs, and pronouns seem to be more cognitively demanding for the participant with dementia, since the mean pause durations on these items were about 1 second longer than for the healthy elderly participant, Elise. The data shows that producing a pronoun required the most effort for the demented participant.

Importantly, the pattern of mean pause lengths before articles and nouns differed between Elise(h) and Mary(d). Mary(d) required a lengthy pause before articles and an even longer pause before nouns (as shown in Figure 5), while Elise(h) required a longer pause before articles than before nouns.

Figure 9 shows that to write the noun phrase *'the kitten'*, Elise paused for 3229 ms before the article *the*, and 1030 ms before the noun *kitten*. Pauses after the production of an article were in general relatively short (437 ms). A similar pattern can be found before the production of the more complex noun phrase *'the*

goldfish (*in the bowl*)'. In this case, the initial pause was longer than 4 seconds. These examples clearly demonstrate the importance of, and the added value conferred by, linguistic diversification in between-word pausing patterns. The extra layer to the pause analysis refines the interpretation of cognitive pauses to a large extent. However, they also show that further fine-tuning of the data is undoubtedly needed in order to better explain the complexity of these pausing patterns, both relative to one another and as a function of the syntactic structure.

S-Notation (Dutch)	S-Notation (translation)	BeforeWord1+2	AfterWordPause
dat-	that	1482	2683
de-	the	3229	437
poes-	kitten	1030	3073
de-	the	4071	359
goudvis-	goldfish	827	327
in-	in	811	281
de-	the	562	562
kom-	bowl	1030	2980
wil-	will	4150	608
vangen-	catch	889	3135

Figure 9. Partial sentence showing pausing times before articles and nouns (Elise(h)). [translation at word level]

To a certain extent, this information is already provided by the chunker since this returns information allowing us to discriminate between pauses at the beginning of a chunk and those inside a chunk. Table 4 summarizes the data for both persons from this perspective.

Table 4. Mean pause duration before words (−1 and −2) at the beginning of a chunk and inside a chunk

	Elise (healthy)	Mary (demented)
	Mean	Mean
Beginning	2061	2600
Inside	1049	2821

The mean pause length of the healthy elderly participant Elise was twice as long at the beginning of a chunk as inside a chunk. By contrast, Mary(d) exhibited a pause length of 2600 ms at the beginning of and about 2800 ms inside a chunk. In combination with the pausing data from Table 5 (Appendix), this suggests that

Mary's efforts were more fragmented and occurred at a lower level. It seems that her text production evolved as a staccato word-by-word sequence. Every word required an almost equal amount of effort: at the beginning of a phrase, within a phrase, at the beginning of a chunk, or inside a chunk. Elise's pattern, on the other hand, seems to reflect more diversification, probably due to the fact that she was able to plan larger text sections.

5.   Conclusion and discussion

Keystroke logging has become instrumental to observe and analyze writing processes. This chapter summarizes the use of keystroke logging as a research technique in general. It also reviews three freely available research tools: ScriptLog, Translog and Inputlog.

To date, (automated) keystroke logging analyses have been mainly based on data obtained at the character level. Although it is clear that this fine-grained, low-level approach leads to very interesting insights, a long tradition of product analysis has taught us that more high-level analyses could also open up new avenues of research. Therefore, Inputlog has been extended by a so-called linguistic analysis in which data is aggregated through to the word level. This module facilitates linguistic process analysis by taking account of the dynamics of writing as the text unfolds. The linguistic module has been developed in English and Dutch, but can potentially also be used for other (Western) languages thanks to the generic approach adopted during its development.

This chapter explains the operation of the module and provides a case study by way of example. In this case study, we show that it is very important to connect the general mental processes observed in writers, on the one hand, with the linguistic features of the text, on the other hand. The case study clearly shows that 'a pause' is too broad a concept, even when we subdivide pauses into different levels (character – word – sentence etc.). We contend that in order to better understand the underlying cognitive processes, the concept of 'pause' needs to be further defined.

In the case study, we described the cognitive processes characteristic of the text production of two elderly people in a controlled task environment. We selected a healthy elderly woman (Elise) and a demented woman (Mary) whose profiles matched in terms of age, education and working career. The product data showed that the healthy elderly participant was able to produce a longer text (about 10 more words) to describe the picture presented to her. When production time is taken into account, it took the demented participant about 2 minutes longer to produce the texts. Moreover, her texts were shorter and she composed about 7 words per minute. In contrast, the healthy elderly participant produced almost

twice as many words per minute (about 12 words). Mary(d) paused about 39 times, whereas Elise(h) paused 25 times. However, as stated above, comparing pausing behavior based on a 2 s pause threshold is perhaps not the best approach if we also wish to address lower-level differences (cf. average pause length of 5.08 s for the healthy elderly participant compared to 5.86 s for the demented elderly woman). Our further results, involving an analysis of within – and between – word pauses using a lower threshold of 200 ms, showed that the pauses were twice as long for the demented participant than for the healthy participant (i.c. within words: 0.80 versus 1.59 s; between words: 1.74 versus 2.99 s).

Furthermore, the new automated linguistic analysis showed that the demented participant took about three times as long to produce nouns (difference of 1750 ms) and twice as long to produce verbs (difference of 1000 ms). By contrast, the pause time before articles differed by about 400 ms. The combined results of the various levels of pause analysis as a function of linguistic feature showed that Mary(d) struggled throughout the writing process as she moved from word to word and that this occurred both at the beginning of a phrase and during a phrase. Elise(h) seemed to produce phrases more fluently and in longer bursts. These production units reveal a pausing behavior with a quite considerable within-participant variance and seem to be defined, to a large extent, by linguistic and syntactic characteristics.

We hope to have demonstrated that automated linguistic analysis provides a large volume of rich data that opens up new avenues for writing process analyses based on keystroke logging. The added value brought about by the further differentiation between different types of between-word pauses undoubtedly merits further exploration and will hopefully lead to a better understanding of the underlying cognitive processes that characterize pause behavior. It is important to remember, however, that – despite the use of sophisticated NLP tools – this type of analysis is more sensitive than, e.g. a general pause analysis. Process data are much more complex than product data, and therefore a certain degree of 'noise' occurs. A typical example is the case in which an unfinished word is deleted during the process, and is presented as such to the linguistic analysis. For instance, when analyzing Mary's data, we had to deal with data loss of about 25% due to complexities in the data, mainly in the form of unrecognized (non-existent or misspelled) words. Adding linguistic features to pauses at the word level has proved to be a first step and is certainly worth further exploration. Moreover, although we believe that adding linguistic features to the pause analysis is an important first step in further diversifying the analysis of cognitive processes, it should be remembered that 'a pause' is still a complex construct that needs to be defined in greater detail and from other theoretical perspectives. For instance, pauses between words are made up of *before* and *after*-word pauses and individuals deal with these in differ-

ent ways, as they do in the case of pauses before and after a full stop (Van Waes & Leijten 2011; 2014). Consideration of this type of interpersonal difference – perhaps in combination with the study of individual motor and typing skills – constitutes an avenue that is clearly worthy of further exploration.

As stated in the introduction, the present research project combines process information with linguistic characteristics. Future analyses will focus on the richness of the written output relative to the cognitive effort invested by writers in order to produce these texts. The process measures can be matched to product measures (final text), including *word diversity* and *expressivity*.

During the remainder of this research project, it is our goal to describe, on a larger scale, the changes that occur during the different stages of AD development, on the one hand, and to test the diagnostic potential for discriminating AD sufferers from controls, on the other. Furthermore, by linking writing process data to lexica and by using NLP tools, we will be able to analyze the data on a higher, more complex level, while also using more advanced statistical techniques that take into account the hierarchical character of the data and the underlying patterns. In this way, we hope to stimulate interdisciplinary research at the crossroads of product and process analysis.

Acknowledgements

References

Baaijen, Veerle M., David Galbraith, and Kees de Glopper. 2012. "Keystroke Analysis Reflections on Procedures and Measures." *Written Communication* 29(3): 246–277.
DOI: 10.1177/0741088312451108

Baaijen, Veerle M., David Galbraith, and Kees de Glopper. 2014. "Effects of writing beliefs and planning on writing performance." *Learning and Instruction* 33(0): 81–91.
DOI: 10.1016/j.learninstruc.2014.04.001

Bazerman, C., ed. 2008. *Handbook of Research on Writing: History, Society, School, Individual, Text*. New York and London: Routledge, Taylor & Francis Group.

Bazerman, Charles, Robert Krut, Karen Lunsford, Susan McLeod, Suzie Null, Paul Rogers, and Amanda Stansell. 2010. *Traditions of Writing Research*. New York and London: Routledge, Taylor & Francis Group.

Berninger, Virginia 2012. *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*. New York and London: Routledge, Taylor & Francis Group.

Caporossi, Gilles, and Christophe Leblay. 2011. "Online writing data representation: a graph theory approach." *Advances in Intelligent Data Analysis* X:80–89.
DOI: 10.1007/978-3-642-24800-9_10

Carl, Michael. 2012. "Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research." Paper read at *LREC*.

Doherty, Stephen, and Sharon O'Brien. 2014. "Assessing the Usability of Raw Machine Translated Output: A User-Centered Study Using Eye Tracking." *International Journal of Human-Computer Interaction* 30(1): 40–51. DOI: 10.1080/10447318.2013.802199

Ehrensberger-Dow, Maureen, and Daniel Perrin. 2009. "Capturing translation processes: a multi-method approach." *Across Languages and Cultures* 20(2): 275–288.
DOI: 10.1556/acr.10.2009.2.6

Folstein, Marshall, Susan E. Folstein, and PR McHugh. 1975. "Mini-mental state. A practical method for grading the cognitive state of patients for the clinician." *Journal of Psychiatric Research* 12: 189–198.

Flower, Linda, and John R. Hayes. 1981. "A cognitive process theory of writing." *College Composition and Communication* 32: 365–387. DOI: 10.2307/356600

Goodglass, Harold, Edith Kaplan, and Barbara Barresi. 1983. *Boston Diagnostic Aphasia Examination* (*BDAE*). Philadelphia: Lea and Febiger. DOI: 10.1002/ana.410160524

Gunawardhane, Suranga DW, Pasan M De Silva, Dayan SB Kulathunga, and Shiromi MKD Arunatileka. 2013. "Non invasive human stress detection using key stroke dynamics and pattern variations." Paper read at *Advances in ICT for Emerging Regions* (ICTer), 2013 International Conference on. DOI: 10.1109/icter.2013.6761185

Hayes, John R. 1996. "A new framework for understanding cognition and affect in writing." In *The science of Writing: Theories, Methods, Individual Differences, and Applications*, ed. by C.Michael Levy, and Sarah E. Ransdell, 1–27. Mahwah: New Jersey: Lawrence Erlbaum Associates.

Hayes, John R. 2012a. "Modeling and remodeling writing." *Written Communication* 29(3): 369–388. DOI: 10.1177/0741088312451260

Hayes, John R. 2012b. "My Past and Present as Writing Researcher and Thoughts About the Future of Writing Research." In *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*, ed. by Virginia Berninger, 3–26. New York: Taylor and Francis Group, Psychology Press. DOI: 10.4324/9780203805312

Jakobsen, Arnt. L. 2006. "Translog: Research methods in translation." In *Computer Keystroke Logging and Writing: Methods and Applications*, ed. by Kirk P. H. Sullivan, and Eva Lindgren, 95–105. Oxford: Elsevier.

Jakobsen, Arnt L. 2011. "Tracking translators' keystrokes and eye movements with Translog." *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies* 94: 37. DOI: 10.1075/btl.94.06jak

Johansson, Victoria, Åsa Wengelin, Johan Frid, and Roger Johansson. 2014. "ScriptLog 2013 state of the art." In *Training school on keystroke logging.* University of Antwerp, Belgium.

Jurafsky, Daniel S., and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Vol. 3. New Jersey: Pearson Education Inc. DOI: 10.1162/089120100750105975

Kellogg, Ronald T. 2008. "Training writing skills: A cognitive developmental perspective." *Journal of Writing Research* 1(1): 1–26.

Leblay, Christophe, and Gilles Caporossi. 2014. *Temps de l'écriture: Enregistrements et représentations.* Vol. 12: Academia/L'Harmattan.

Leijten, Mariëlle, Sven De Maeyer, and Luuk Van Waes. 2011. "Coordinating sentence composition with error correction: A multilevel analysis." *Journal of Writing Research* 2(3): 331–363. DOI: 10.1007/s11145-009-9190-x

Leijten, Mariëlle, Lieve Macken, Veronique Hoste, Eric Van Horenbeeck, and Luuk Van Waes. 2012. "From character to word level: Enabling the linguistic analyses of Inputlog process data." In *European Association for Computational Linguistics, EACL – Computational Linguistics and Writing (CL&W 2012): Linguistic and Cognitive Aspects of Document Creation and Document Engineering,* ed. by Michael Piotrowski, Cerstin Mahlow, and Robert Dale. Avignon. http://aclweb.org/anthology/W/W12/W12-03.pdf.

Leijten, Mariëlle, and Luuk Van Waes. 2012. "Inputlog 4.0: Keystroke Logging in Writing Research." In *Learning to Write Effectively: Current Trends in European Research,* ed. by Mark Torrance, Denis Alamargot, Montserrat Castelló, Franck Ganier, Otto Kruse, Anne Mangen, Liliana Tolchinsky, and Luuk Van Waes, 363–366. Emerald Group Publishing Limited.

Leijten, Mariëlle, and Luuk Van Waes. 2013. "Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes." *Written Communication* 30(3): 358–392. DOI: 10.1177/0741088313491692

Leijten, Mariëlle, and Luuk Van Waes. 2014. *Manual Inputlog 6.0.* Antwerp: University of Antwerp.

Leijten, Mariëlle, Luuk Van Waes, Karen Schriver, and John R. Hayes. 2014. "Writing in the workplace: Constructing documents using multiple digital sources." *Journal of Writing Research* 5(3): 285–336.

Lindgren, Eva, Mariëlle Leijten, and Luuk Van Waes. 2011. "Adapting to the reader during writing." *Written Language & Literacy* 14: 188–223.

MacArthur, Charles A., Steve Graham, and Jill Fitzgerald. (Eds.). 2008. *Handbook of Writing Research.* New York, NY: The Guilford Press. DOI: 10.1111/j.1467-873x.2008.00423.x

Macgilchrist, Felicitas, and Tom Van Hout. 2011. "Ethnographic discourse analysis and social science." Paper read at *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research.*

Maggio, Severine, Bernard Lété, Florence Chenu, Harriet Jisa, and Michel Fayol. 2012. "Tracking the mind during writing: Immediacy, delayed, and anticipatory effects on pauses and writing rate." *Reading and Writing* no. 25 (9): 2131–2151. DOI: 10.1007/s11145-011-9348-1

Manning, Christoper, D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing.* Cambridge, MA: The MIT Press. DOI: 10.1162/coli.2000.26.2.277

McKhann Guy, David Drachman, Marshall Folstein, et al. 1984. "Clinical diagnosis of Alzheimer's disease." Report of the NINCDSADRDA work group under the auspices of the Department of Human Services Task Force on Alzheimer's disease. *Neurology* 34: 939–944.

Mesulam, M-Marsel, Murray Grossman, Argye Hillis, Andrew Kertesz, and Sandra Weintraub. 2003. "The core and halo of primary progressive aphasia and semantic dementia." *Annals of Neurology* 54(5): 11–14. DOI: 10.1002/ana.10569

Petersen, Ronald C. 2004. "Mild cognitive impairment as a diagnostic entity." *Journal of Internal Medicine* 256(3): 183–194.

Risku, Hanna, Florian Windhager, and Matthias Apfelthaler. 2013. "A dynamic network model of translatorial cognition and action." *Translation Spaces* 2(1): 151–182. DOI: 10.1075/ts.2.08ris

Robert, Isabelle S., and Luuk Van Waes. 2014. "Selecting a translation revision procedure: do common sense and statistics agree?" *Perspectives*: 1–18. DOI: 10.1080/0907676x.2013.871047

Schilperoord, Joost. 1996. *It's about time. Temporal aspects of cognitive processes in text production*. Amsterdam/ Atlanta: Rodopi.

Severinson Eklundh, Kerstin, and Py Kollberg. 2002. "Studying writers' revision patterns with S-notation analysis." In *Contemporary tools and techniques for studying writing*, ed. by Thierry Olive, and C. Michael Levy, 89–104. Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-010-0468-8_5

Spelman Miller, Kristyan, Eva Lindgren, and Kirk P. H. Sullivan. 2008. "The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging." *TESOL Quarterly* 42(3): 433–454.

Sperling, Reisa A., Paul S. Aisen, Laurel A. Beckett, David A. Bennett, Susanne Craft, et al. 2011. "Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." *Alzheimers Dement* 7: 280–292.

Sullivan, Kirk P.H., and Eva Lindgren. 2006. *Computer Key-Stroke Logging and Writing*. Edited by G. Rijlaarsdam, *Studies in Writing*. Oxford: Elsevier Science.

Van Eynde, Frank, Jakub Zavrel, and Walter Daelemans. 2000. "Part of speech tagging and lemmatisation for the Spoken Dutch Corpus." In *Proceeding of the Second International Conference on Language Resources and Evaluation*, ed. by M. Gavrilidou et al., 1427–1433. Athens.

Van Horenbeeck, Eric, Tom Pauwaert, L. Van Waes, and M. Leijten. 2012. *S-notation: S-notation markup rules* (Technical Description). Antwerp: University of Antwerp.

Van Waes, Luuk, and Mariëlle Leijten. 2011. "Observing and analysing digital writing processes with Inputlog." In *Antwerp Summer School on Writing Process Research: Keystroke logging and Eyetracking*. Antwerp. DOI: 10.1177/0741088313491692

Van Waes, Luuk, and Mariëlle Leijten. 2013. "Vlot schrijven-Een multidimensioneel perspectief op 'writing fluency'." *Tijdschrift voor taalbeheersing* 35(2): 160–182. DOI: 10.5117/tvt2013.2.waes

Van Waes, Luuk, and Mariëlle Leijten. 2014. Inputlog 6.0: Pause and fluency analysis." In *Keystroke logging training school*. Antwerp.

Van Waes, Luuk, Mariëlle Leijten, and Aline Remael. 2013. "Live subtitling with speech recognition. Causes and consequences of text reduction." *Across Languages and Cultures* 14(1): 15–46. DOI: 10.1556/acr.14.2013.1.2

Van Waes, Luuk, Mariëlle Leijten, Åsa Wengelin, and Eva Lindgren. 2012. "Logging tools to study digital writing processes." In *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*, ed. by Virginia Wise Berninger, 507–533. New York/Sussex: Taylor & Francis.

Van Waes, Luuk, and Peter Jan Schellens. 2003. "Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers." *Journal of Pragmatics* 35(6): 829–853.

Visch-Brink, Evy, Dorien Vandenborre, Hyo Jung De Smet, and Peter Mariën. 2014. *The Comprehensive Aphasia Test-NL*, *Pearson*. Amsterdam.

Wengelin, Åsa. 2006. "Examining pauses in writing: Theories, methods and empirical data". In *Computer Keystroke Logging and Writing: Methods and Applications*, ed. by Kirk P. H. Sullivan, and Eva Lindgren, 107–130. Oxford, UK: Elsevier.

Wengelin, Åsa, Mark Torrance, Kenneth Holmqvist, Sol Simpson, David Galbraith, Victoria Johansson, and Roger Johansson. 2009. "Combined eye-tracking and keystroke-logging methods for studying cognitive processes in text production." *Behavior Research Methods* 41(2): 337–351. DOI: 10.3758/brm.41.2.337

Wininger, Michael. 2014. "Measuring the evolution of a revised document." *Journal of Writing. Research* 6(1): 1–28.

Yi, Hyon-Ah, Peachie Moore, and Murray Grossman. 2007. "Reversal of the Concreteness Effect for Verbs in Patients with Semantic Dementia." *Neuropsychology* 21(9): 9–19. DOI: 10.1037/0894–4105.21.1.9

Appendix

Dutch example of final text produced by healthy elderly woman: Elise(h) (81)

"De ene ramp na de andere: de afwasbak van de mama loopt over (is de kraan geblokkeerd ?) zoonlief wil heimelijk koekjes uit de koekendoos halen, zijn stoel kantelt en hij zal waarschijnlijk op de grond vallen. Wil kleine zus ook een koekje of lacht zij hem uit ?Antwoord op het volgende plaatje."

Dutch example of final text produced by elderly woman with dementia Mary(d) (79):

"ik zie een kind dat een bord iot de kast wenst te halende moeder is eeen bord aan jet afdrogen. het stoeltje waarop de jongen staat is aanhet kantwkanteken; ik denk fat er verscheidene bit borden zullensneuvelenmm moeder is aan het afdeogen er valt warze p op de gron, grond xus zie ik nog andere ongelukkengebeuren."

Table 5. Number of pauses and mean pause duration before words (–1 and –2) per word category

	Elise (healthy)		Mary (demented)	
	Number	**Mean**	**Number**	**Mean**
Articles (26)	19	1812	6	2288
Nouns (33)	23	1077	10	2827
Verbs (31)	20	1639	11	2625
Prepositions (19)	13	2630	6	2681
Pronouns (12)	8	2084	5	3385

Index

This volume aims at contributing to an interpretive approach to writing and its dynamics. It offers a general scope on the process-product interface by multiplying the points of view on both the process and the product and their links. The book presents new findings and perspectives in the study of language and writing, both theoretical and methodological (e.g. dual process models of writing, pragmatics of writing, linguistic analysis of psycholinguistic units such as bursts of production). It also presents new tools for a longitudinal approach to the writing steps, key-stroke logging with integrated linguistic modules, and textometric analysis of written texts. The volume is composed of five sections that highlight different approaches to writing from the viewpoint of multiple disciplines: Anthropology, Cognitive Psycholinguistics, Communication Studies, Didactics (Applied Linguistics), Discourse Analysis, Literacy, Sociolinguistics and Text Genetics. This book will be relevant for scholars and students interested in writing, text analysis, literacy, learning and teaching.

John Benjamins Publishing Company