



Științele limbajului

**Roxana Patraș
Loredana Cuzmici**
(editors)

**Tools, Methods, and Solutions
for the Exploration of Romanian Corpora**

INSTITUTUL EUROPEAN

Roxana Patraș, Loredana Cuzmici (editors), *Tools, Methods, and Solutions for the Exploration of Romanian Corpora*

© 2024 Institutul European Iași, pentru prezenta ediție

INSTITUTUL EUROPEAN

Iași, Str. Bălușescu nr. 2, et. I, O P 6, CP 1309

euroedit@hotmail.com.; www.euroinst.ro

Descrierea CIP a Bibliotecii Naționale a României

Tools, methods, and solutions for the exploration of Romanian corpora /

ed.: Roxana Patraș, Loredana Cuzmici. - Iași : Institutul European, 2024

Conține bibliografie

ISBN 978-606-24-0395-9

I. Patraș, Roxana (ed.)

II. Cuzmici, Loredana (ed.)

811.135.1

All rights reserved. No part of this work covered by the copyrights hereon may be reproduced or copied in any form or by any means – graphic, electronic, or mechanical, including photocopying, recording, taping – without written permission of publisher.

PRINTED IN ROMANIA

Roxana Patraş, Loredana Cuzmici
(editors)

TOOLS, METHODS, AND SOLUTIONS FOR THE EXPLORATION OF ROMANIAN CORPORA

INSTITUTUL EUROPEAN
2024

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS–UEFISCDI, project number PN-III-P1-1.1-TE-2019-0127, within PNCDI III.

The revision and editing of this collection was performed by BolderWords translation & revision and was supported by the Altissia Chair in Digital Cultures and Ethics held by Chris Tanasescu at Université catholique de Louvain.

Table of Contents

INTRODUCTORY NOTES	7
Tactics: your DH is not my DH Roxana PATRAȘ	9
CREATING TOOLS FOR ROMANIAN CORPORA	23
RELATE: a modern processing platform for Romanian language Vasile PĂIȘ, Radu ION, Andrei-Marius AVRAM, Maria MITROFAN, Dan TUFIȘ	25
Tools and Resources for Digitizing Historical Romanian Documents Victoria BOBICEV, Tudor BUMBU, Ludmila MALAHOV, Alexandru COLESNICOV, Svetlana COJOCARU, Liudmila BURTSEVA, Cătălina MĂRÂNDUC	53
Digitalization of Romanian Lexicography Elena Isabelle TAMBA	85
ADAPTING TOOLS FOR ROMANIAN CORPORA	97
Postdigital Creativity: modeling poetry translation with multiplexes, neural networks, and large language models Raluca TANASESCU, Chris TANASESCU	99
Challenges of Digitization and Digitalization for the Study of Writing: from archival survey to process analysis Georgeta CÎȘLARU	125
Towards a Digital Corpus-Based Method for Assessing Language Level in EFL Student Writing: a case study on Romanian undergraduate literary analyses Alexandru ORAVIȚAN, Mădălina CHITEZ	139
DIGITIZING ROMANIAN HERITAGE	151
A Computer-Assisted Analysis of Eminescu's <i>Doină</i> Ioana GALLERON	153
Mapping Eastern European Political Discourse and Inequalities through Public Monuments: a digital cartography crowdsourcing project Voica PUȘCAȘIU	175
The Title of the Feuilleton Novel: a model of semantic annotation Lucreția PASCARIU	193
Exercises in Literary Geography: where hajduks roam in the second half of the nineteenth century Alexandra OLTEANU	227

<i>DIGITUS DEI EST HIC!</i>	251
Proxy Structures in Computational Analyses of Text Corpora	253
Laura PRICOP	
The Many Faces of Virtual: charting the family resemblance network for conceptual understanding	267
Matei Alexandru STOENESCU	
CARTOGRAPHIES OF THE ROMANIAN NOVEL: THEN AND NOW	297
The Recontextualization of a Genre	299
Loredana CUZMICI	
Notes on the Contributors	321

Challenges of Digitization and Digitalization for the Study of Writing: from archival survey to process analysis

Georgeta CÎȘLARU

This article looks at the impact of digitization and digitalization on the study of writing from an interdisciplinary perspective, with a specific focus on *keylog* data. According to Brennen and Kreiss (2016), *digitization* can be defined as the material process of converting analog streams of information into digital bits, and *digitalization* as the way domains of social life are restructured around digital communication and media infrastructures. The former is necessarily determined by the latter, with tools and practices mutually conditioning each other.

The impact of digitization on the very modalities of research in the human sciences is widely recognized (Dalbello 2011). Digital Humanities are changing the epistemologies and ontologies underlying research (Berry 2012). One of the most obvious impacts on humanities research, of which a large part involves analyzing and interpreting written texts (McGillivray *et al.* 2020), is the volume of archives that can be processed thanks to digitization on the one hand, and the automation it enables on the other (Moretti 2007).

Since its emergence as a technology (Auroux 1994), writing has undergone many changes, evolving from physical to digital media. Recent publications have examined this evolution and highlighted the many changes it has entailed (Van Waes *et al.* 2006; Kruse *et al.* 2023). These changes range from the use of word processing software (Heilmann 2023) to the development of digital tools for writing (argumentation, note-taking, etc., e.g., Benetos 2023), learning and regulation (plagiarism, self-reflexivity, e.g., Devitt *et al.* 2023). Digital technologies have not only impacted the ways of writing and favored new discourse genres (Rowley-Jolivet and Carter-Thomas *in print*), they have also helped to develop new methodologies for investigating writing and the writing process (Van Waes *et al.* 2011).

Impacts on the nature of the data

Archives in writing studies

For some disciplines in the humanities, access to archives is a prerequisite for research. In the field of writing studies, the literature has focused on the evolution of writing systems, from a phylogenetic perspective, or on texts in the making, from an ontogenetic perspective. The latter is at the heart of critical genetics (Hay 2002).

The anteriority of the written text has also interested linguists in its variational, dynamic, and enunciative compositional dimension, shifting the research focus from the text entity to the act or process of production, writing, observed from drafts. Since the 1970^s, critical genetics has provided a vast body of work on the archaeology of texts, studying the traces of the writing process (*cf.* Hay 2002, Grésillon 2016 [1994]). Genetic criticism, which was first interested in literary creativity, happy accidents and the protean milestones of writing (*cf.* Grésillon 2008, 11), places at the center of its analysis erasure, reformulation and other operations that ultimately constitute the four operations of rewriting: addition, deletion, replacement and displacement (Grésillon 1989). These operations are all traces that reconstitute entire sections of the process: for instance, “erasure is one of the elements capable of bearing witness to the temporal dimension specific to any writing process” (Grésillon 2016 [1994], 83). The identification by textual genetics of these four operations of text revision introduces a distributional perspective that opens the way to multiple analyses. Linguistic issues were rapidly identified and have been the subject of several development proposals (*cf.* Grésillon and Lebrave 2008).

Attempts have been made to describe types of writers according to the way they use (or do not use) drafts. Some authors consider the draft itself and the “drafting” activity as part of the process of idea generation and organization, that is, planning (Sharples and Pemberton 1990). The aim is to understand the impact of different types of drafts (lists of words, plans, smoothed drafts, etc.) on the writing itself and on the quality of the text produced (Piolat and Roussey 1996). Other authors identify types of writers according to the more or less planned or programmatic nature of their drafts. Galbraith (1992), for example, distinguishes between “executor” writers (who base their drafting on highly organized drafts) and “discoverer” writers (who write from very

sketchy, poorly organized drafts, generating ideas and their plan essentially as they write). Fenoglio (2007) points out that this distinction between programmatic and free writing seems rather random, apart from the typical cases of Zola and Proust, respectively executor and discoverer.

Whether the focus is on revision operations, or on the complex dynamics of cognitive processing underlying written production, everything concurs in considering the textualization process in a non-linear way. In genetic criticism studies, the observation of rewriting operations, various graphic marks on drafts and the analysis of conceptual elaborations also show the non-linearity of the writing process (for recent work, *see* Fenoglio 2015). The text is conceived in terms of deployment, building on pre-existing elements (the pre-text, made up of memorized information, standards, formats) and with elementary units such as keywords, proto-words, etc., which are then used to create the text, as shown by *Figure 1*.

Born-digital archives and the writing process

While there is no consistent evidence of the benefits of replacing handwriting by typing in early writing instruction (Wollscheid *et al.* 2016), the use of digital tools to compose a text has impacted writing practice. Keyboard writing in itself is the exact opposite of handwriting, as it operates with discrete script signs, which may influence the apprehension of linguistic units: letters, words, and phrases (Doquet 2006). Physical space management is also different, with text displayed in small windows on the screen, and revision traces are evanescent (Anis 1995).

The effects on the process and the product of writing are diverse, but in all cases, social practice in itself and the object of study are affected, as reported by various studies. Whilst reading on paper shows more positive affordances for Finnish college students, digital writing is valued because it enables the text to be quickly and efficiently edited (Taipale 2014). Williams and Beam (2019) point out in their review that digital technologies and writing tools increase motivation and social collaboration in students. Dahlström and Boström (2017) showed that university students produce longer and linguistically more correct narrative texts on tablets compared to handwritten texts. They also noticed differences in representing

processes, with material processes dominating in digital writing, and material and relational processes dominating in handwriting.

Compared to handwritten or typewritten data, texts produced by means of digital tools are formally and qualitatively different, and offer different research perspectives. Drafts become evanescent and not archival, and revisions leave no visible physical trace. The use of digital tools also offers new resources for scientific investigation, however, and new solutions for accessing the virtual “traces” of writing, for instance by recording the progressive construction of the text through its different versions. It is then possible to run linguistic processing (automatic or otherwise) on the raw texts collected directly. First and foremost, this allows longitudinal approaches in short diachrony, by comparing the versions of texts from each writing session and identifying variants that interfere with the linearity of the texts. We call this macro-segmentation. In this way, textual genetics can calculate additions, replacements and deletions to obtain an overall view of the modifications a text has undergone. Software such as EDITE MEDITE6 (Fenoglio and Ganascia 2007) or ALLONGOS (Lardilleux *et al.* 2013) compare each new text version with the previous one. For example, ALLONGOS compares versions and indicates which revision operations have been carried out.

The various writing sessions reveal stabilized versions of the text in the making and constitute traces of the stages of discursive configuration. Ferrer (2011, 182) insists on the role and cognitive interest of drafts for textual configuration, considering variants as “protocols for making of a text”.

Process analysis

The “protocols for making a text” have always intrigued researchers seeking to understand the cognitive mechanisms of writing. Previously, researchers exploited verbal protocols in order to make mental mechanisms “observable” by asking individuals to describe aloud the activities they performed, either concomitantly or retrospectively (Levy and Ransdell 1995, 1996). These protocols led to models of written communication distinguishing planning (cognitive activities: activation-selection, organization), writing (transcription, translation: verbalization, from thought

to language), and revision (reading-evaluation-correction) (Flower and Hayes 1980; Hayes and Flower 1980).

These reflective approaches were intended to compensate for the lack of access to the writing process. In the 1990^s, the evolution of digital tools made it possible to access process information through screen capture, then by means of tablets, screen-capture or key-stroke logging tools. These developments led to the study of eye movements and grapho-motor activity during writing on Eye-and-Pen tablets (Chesnet and Alamargot 2005), the study of L2 writing in university students via the screen capture tool Screencast-O-Matic (Seror 2013), process recording and analysis via keylogs such as Inputlog (Leijten and van Waes 2006) or Scriptlog (Strömquist and Malmsten 1998). Wengelin and Johansson (2023) offer a detailed description of the tools and their functionalities.

A chrono-topological longitudinal approach to writing and textualization

Key-stroke logging software provides a chronologically structured record of all the operations involved in writing on a keyboard: typing characters and spaces, revisions (deletions, insertions, moves, etc.), pauses sequencing the process, or consultation of sources outside the document. All these data can be updated in their longitudinal dynamics thanks to their rereading function, extracted by type (pause, typing, revision), classified by series, etc. Recorded writing has thus become a means and an object of analysis, in cognitive psychology, through experimental studies, but also in communicative approaches, such as the pragmatics of writing interested in progress analysis (*see*, for example, Perrin 2013). By recording and analyzing chronometric data associated with production, such as pauses or writing speed, it is possible to attempt to infer the cognitive mechanisms at play. *Pausology*—the study of pauses in language production—is an essential part of any approach to the processual dimension of language production. Pauses provide a first level of segmentation of language flow, whether oral or written, and enable us to associate the production of textual units with the dynamics of the writer's or speaker's mental functioning (*cf.* Fayol 1997).

By offering longitudinal access to the chronological unfolding of the writing process, keystroke recording tools stage a representational paradox. While the studies

based on material archives and drafts demonstrate the non-linearity of the writing process, the temporality of recorded writing is always linear. Only by relating this chronological linearity to the topology of the text is it possible to articulate temporality to the non-linearity of the process. For example, graphic representations of the writing process proposed by Leblay and Caporossi (2015) clearly illustrate the recursivity of revision operations, and *Inputlog* graphically represents the progress of writing in terms of characters typed or deleted, pauses, and sources consulted (*Figure 3*).

In addition to the writer's actions on the text, *Inputlog* software records all the writer's actions on the computer, outside the word processor. This means that, while writing, it is possible to consult various files stored on the computer, go on the Internet, copy text, write and send an e-mail, etc. All this information is recorded, making it possible to analyze the sources consulted by the writer. With all this information recorded, it becomes possible to analyze the sources consulted by the writer and see whether they have been integrated into the text. The software can also encode metadata to link a writing process to a specific writer, or compare data from different writers. It can also use the generic file created to perform a number of analyses of the writing process. *Inputlog* provides information on texts written (*i.e.*, number of words, sentences, words per sentence in the final text or during the session), on revision operations (by creating files summarizing the various operations carried out), on pauses (*i.e.*, number, average duration, pauses between words, sentences, paragraphs) and execution periods (see below). It also calculates keystroke metrics to analyze the typing skills of writers.

The identification of textual operations performed by writers when composing their text is based on S-notation, a language developed specifically to support the analysis of revision episodes (Severinson Eklundh and Kollberg 1996). The software also enables a wide range of analyses of the writing process to be carried out automatically. From this point of view, digital development affects not only the nature of the archive (born-digital, the digital edge) and the point of view on data (processual), but also the place of the analyst, by offering him or her immediate avenues of observation.

Linguistic insights

In the field of linguistics, two methodological questions emerge as the object itself mutates: the first concerns the flow of language data recorded in real time, and raises the question of what can be considered the object of linguistic analysis, *i.e.*, how to discriminate between “analyzable” linguistic units? The second concerns the transition from the language sequence level to the text level, *i.e.*, how analysts recover the text as a product when moving away from it both upstream in time and with regard to its unitary structure.

Pauses in the writing process can be considered as segmentation markers, according to which it is possible to extract and categorize segments of linguistic production. Pauses segment textual sequences in the process of elaboration. These sequences can be considered to have the status of linguistically attested empirical data. We can define pausing as spontaneous segmentation, or auto-segmentation, insofar as pauses are the work of the writers and not of the analyst or the tool, which merely records and measures them. From this perspective, the primary object of analysis we focus on is linguistic sequences written uninterruptedly between two pauses in production. We refer to these sequences as *bursts of writing*, a concept borrowed from the psycholinguistics of writing (*cf.* Chenoweth and Hayes 2001) and reworked in the frame of linguistics by Cislaru and Olive (2018). From a linguistic point of view, we can thus consider textual bursts as units of written production, or units of performance.

The duration of pauses, the length of the bursts, and the way in which pauses and bursts are structurally arranged constitute ecological data that can be studied in their own right. This new form of segmentation does not compete with all the other forms of segmentation used to analyze texts and discourse (repeated segments, sequences, periods, motifs, formulas, etc.), nor does it represent an exclusive alternative, but rather a temporal and processual approach to text and writing (Cislaru and Olive 2018).

This processual view of writing brings it closer to oral production data, implying similar questions about flow segmentation and the status of the pause (Grosman *et al.* 2018). In this respect, we can consider that we are dealing with a new epistemological mutation engendered by the evolution of digitalization.

Last but not least, on-line revisions reveal metadiscursive approaches whose traces are undetectable in the final product and which are sometimes identifiable only

through the movement that consists in filling a digital space with two different forms successively. Yet the relationship (which it might be questionable to call “equivalence”, as in reformulations) between these forms and between the chosen form, the text in itself and the writing process remains to be clarified: should these forms be considered revisions, or on the contrary production, if they impact the evolution of the text? These issues led to new classification suggestions (Conijn *et al.* 2022) and show that revisions may be considered as process-oriented or product-oriented.

Conclusion

This article takes advantage of a brief historical overview of studies on the writing of texts to show the contributions, implications and mutations of digital tools in the study of writing, moving from the physical archive to the born-digital archive, from the sequential study of the process to on-line study. This evolution shows the transformation of the object of study itself, with the new issues it implies and the fields of research that are being redrawn.

Annex 1: Figures

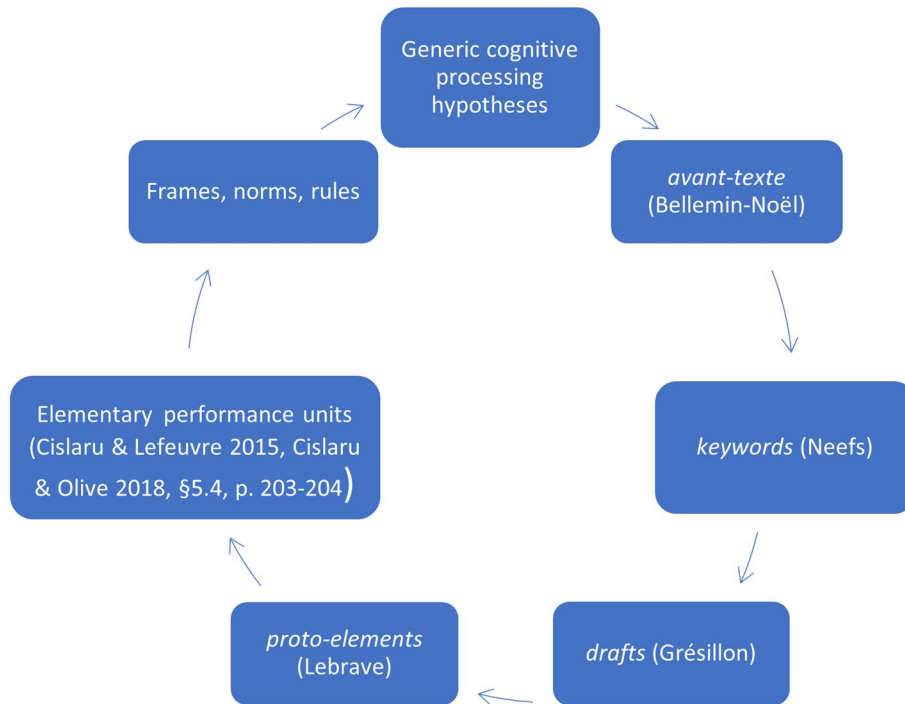


Figure 1: Cognitive-linguistic text deployment.

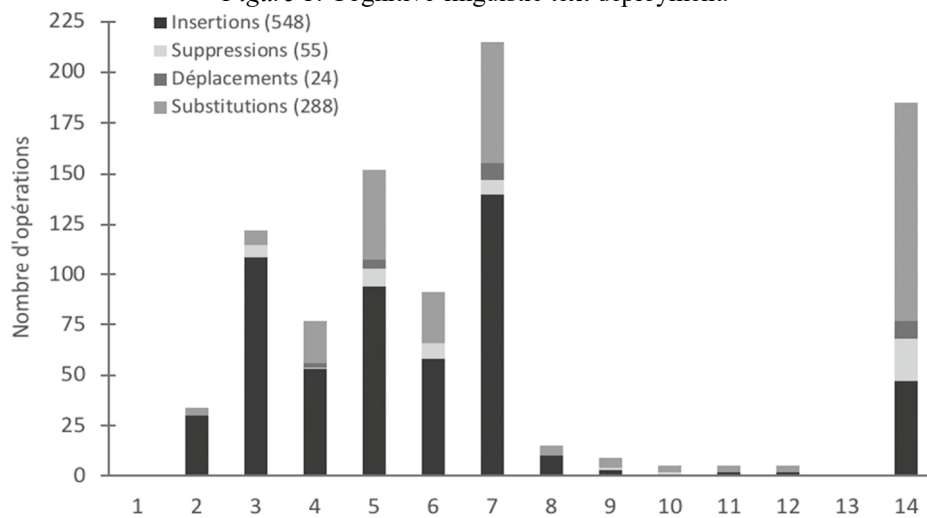


Figure 2: Number of rewriting operations for a 14-session writing process.

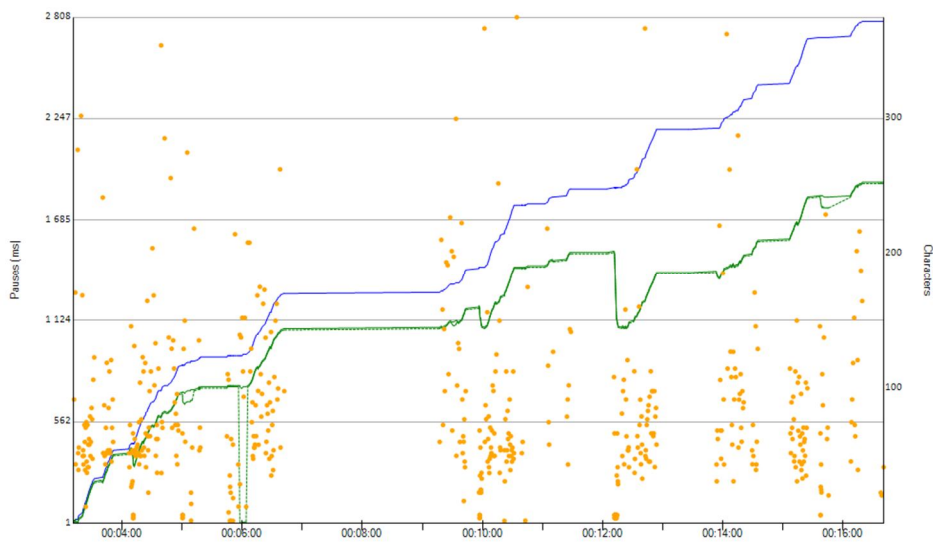


Figure 3: Inputlog representation of a writer's productivity during a social report writing session.

The yellow dots indicate pauses, each of whose duration (in ms) is shown on the left vertical axis, with the moment of occurrence indicated on the horizontal axis. The green curve represents the sum of characters present in the text. The blue curve indicates the cumulative sum of characters typed since the start of the writing session. A descending green curve indicates the deletion of characters. The dotted green curve indicates the cursor position (Cislaru and Olive 2018).

References

- Anis, Jacques. 1995. "Le traitement de texte: écriture ou méta-écriture?" *Repères* 11: 15–27.
- Auroux, Sylvain. 1994. *La révolution technologique de la grammatisation*. Liège: Mardaga.
- Bellemin-Noël, Jean. 1972. *Le texte et l'avant-texte*. Paris: Larousse.
- Benetos, Kalliopi. 2023. "Digital Tools for Written Argumentation." In *Digital Writing Technologies in Higher Education. Theory, Research, and Practice*, edited by Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani, 81–99. Springer.
- Berry, D. M. 2012. "Introduction: Understanding the Digital Humanities." In *Understanding Digital Humanities*, edited by D. M. Berry, London: Palgrave Macmillan. https://doi.org/10.1057/9780230371934_1.
- Brennen, J. Scott, and Daniel Kreiss. 2016. "Digitalization." In *The International Encyclopedia of Communication Theory and Philosophy*, 1–11. John Wiley & Sons.
- Chenoweth, Norton Ann, and John R. Hayes. 2001. "Fluency in Writing: Generating Text in L1 and L2." *Written Communication* 18 (1): 80–98.
- Chesnet, David, and Denis Alamargot. 2005. "Analyse en temps réel des activités oculaires et grapho-motrices du scripteur: intérêt du dispositif 'Eye and Pen'." *L'année psychologique* 105 (3): 477–520.
- Cislaru, Georgeta, and Florence Lefevre. 2015. "The Instrumental Use of Verbless Sentences in Writing and Rewriting: A Longitudinal and Genre-Contrasted Point of View." In *Writing(s) at the Crossroads: The Process/Product Interface*, edited by Georgeta Cislaru, 57–80. Amsterdam–Philadelphia: John Benjamins.
- Cislaru, Georgeta, and Thierry Olive. 2018. *Le processus de textualisation*. Bruxelles: De Boeck.
- Conijn, Rianne, Emily D. Speltz, Menno van Zaanen, Luuk van Waes, and Evgeny Chukharev-Hudilainen. 2022. "A Product- and Process-Oriented Tagset for Revisions in Writing." *Written Communication* 39 (1): 97–128. <https://doi.org/10.1177/074108832-11052104>.
- Dahlström, Dahlström, and Boström, Boström. 2017. "Pros and Cons: Handwriting Versus Digital Writing." *Nordic Journal of Digital Literacy* 12 (4): 143–161.
- Dalbello, Marija. 2011. "A Genealogy of Digital Humanities." *Journal of Documentation* 67 (3): 480–506.
- Devitt, Ann, Kalliopi Benetos, and Otto Kruse. 2023. "Writing and Learning: What Changed with Digitalization?" In *Digital Writing Technologies in Higher Education. Theory, Research, and Practice*, edited by Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani, 423–433. Springer.
- Doquet-Lacoste, Claire. 2006. "L'objet insaisissable: l'écriture sur traitement de texte." *Genesis* 27: 35–44.

- Fayol, Michel. 1997. *Des idées au texte. Psychologie cognitive de la production verbale, orale et écrite*. Paris: Presses Universitaires de France.
- Fenoglio, Irène. 2007. "Du texte avant le texte. Formes génétiques et marques énonciatives de prévisions textualisantes." *Langue française* 155: 8–34.
- Fenoglio, Irène. 2015. "From Writing Under Production to the Finished Product: A Processual Threshold." In *Writing at the Crossroads: The Process-Product Interface*, edited by Georgeta Cislaru, 25–50. Amsterdam – Philadelphia: John Benjamins.
- Fenoglio, Irène, and Jean-Gabriel Ganascia. 2007. "MEDITE: un logiciel pour l'approche comparative de documents de genèse." *Genesis* 27: 166–168.
- Ferrer, Daniel. 2011. *Logiques du brouillon. Modèles pour une critique génétique*. Paris: Éditions du Seuil.
- Flower, Linda S., and John R. Hayes. 1980. "The Dynamics of Composing: Making Plans and Juggling Constraints." In *Cognitive Processes in Writing*, edited by Lee W. Gregg and Erwin R. Steinberg, 31–50. Hillsdale, NJ: Lawrence Erlbaum.
- Galbraith, David. 1992. "Conditions for Discovery Through Writing." In *Computers and Writing. Issues and Implementations*, edited by Mike Sharples, 45–72. Dordrecht: Kluwer Academic Publishers.
- Grésillon, Almuth. 1989. "Fonctions du langage et genèse du texte." In *La naissance du texte*, edited by Louis Hay, 177–192. Paris: José Corti.
- Grésillon, Almuth. 2008. *La mise en œuvre. Itinéraires génétiques*. Paris: CNRS Éditions.
- Grésillon, Almuth. 2016 [1994]. *Éléments de critique génétique*. Paris: CNRS Éditions.
- Grésillon, Almuth, and Jean-Louis Lebrave. 2008. "Linguistique et génétique des textes: un décalogue." *Le français moderne*, numéro spécial "Tendances actuelles de la linguistique française": 37–49.
- Grosman, I., A. C. Simon, and L. Degand. 2018. "Variation de la durée des pauses silencieuses: impact de la syntaxe, du style de parole et des disfluences." *Langages* 211: 13–40.
- Hay, Louis. 2002. *La littérature des écrivains. Questions de critique génétique*. Paris: Corti.
- Hayes, John R., and Linda S. Flower. 1980. "Identifying the Organization of Writing Processes." In *Cognitive Processes in Writing*, edited by L. W. Gregg and E. R. Steinberg, 3–30. Hillsdale, NJ: Lawrence Erlbaum.
- Heilmann, Till A. 2023. "The Beginnings of Word Processing: A Historical Account." In *Digital Writing Technologies in Higher Education. Theory, Research, and Practice*, edited by Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani, 3–13. Berlin and New York: Springer.
- Kruse, Otto, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani (eds). 2023. *Digital Writing Technologies in Higher Education. Theory, Research, and Practice*. Springer.
- Lardilleux, Adrien, Serge Fleury, and Georgeta Cislaru. 2013. "Allongos: Longitudinal Alignment for the Genetic Study of Writers' Drafts." *Lecture Notes in Computer Sciences* 7817: 537–548.
- Leblay, Christophe, and Gilles Caporossi. 2015. "A Graph Theory Approach to Online Writing Data Visualization." In *Writing at the Crossroads: The Process-Product Interface*, edited by Georgeta Cislaru, 171–182. Amsterdam–Philadelphia: John Benjamins.

- Lebrave, Jean-Louis. 1987. “Le Jeu de l’énonciation en allemand d’après les variantes manuscrites des brouillons de H. Heine.” Thèse de Doctorat d’Etat: Université Paris-Sorbonne.
- Leijten, Mariëlle, and Luuk Van Waes. 2006. “Inputlog: New Perspectives on the Logging of On-Line Writing.” In *Computer Keystroke Logging and Writing: Methods and Applications*, edited by Kirk P. H. Sullivan and Eva Lindgren, 73–94. Amsterdam: Elsevier.
- Levy, C. Michael, and Sarah Ransdell. 1995. “Is Writing as Difficult as It Seems?” *Memory and Cognition* 23: 767–779.
- Levy, C. Michael, and Sarah Ransdell. 1996. “Writing Signatures.” In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, edited by C. Michael Levy and Sarah E. Ransdell, 149–162. Mahwah, NJ: Lawrence Erlbaum.
- Lindgren, E., and K. Sullivan (eds). 2019. *Observing Writing: Insights from Keystroke Logging and Handwriting*. Leiden/Boston: Brill.
- McGillivray, B., T. Poibeau, and P. Ruiz Fabo. 2020. “Digital Humanities and Natural Language Processing: « Je t’aime... Moi non plus ».” <https://doi.org/10.17863>.