Behind the computer screens

The use of keystroke logging for genetic criticism applied to born-digital works of literature

Bekius, L.L.

Publication date
2023

Document Version
Final published version

Link to publication

Behind the computer screens

The use of keystroke logging for genetic criticism applied to born-digital works of literature

Lamyk L. Bekius

'Behind the computer screens'

UNIVERSITY OF AMSTERDAM
Faculty of Humanities

University of Antwerp
Faculty of Arts

'Behind the computer screens'

The use of keystroke logging for genetic criticism applied to born-digital
works of literature

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Aula der Universiteit

op vrijdag 6 oktober 2023, te 11.00 uur

door Lamyk Lidwien Bekius

geboren te Meppel

"I wondered what to do with all the paper that accompanies a writer of my generation. There were scripts for theatre and film, poems, stories, libretti for opera, drafts of novels that had been written on a number of technologies – the manual typewriter, the electric typewriter, early computers. Some of my diaries dated back to 1985" (Levy 2019, 48)

"Je dialogen en taalvondsten verrassen me; ik dacht dat het allemaal veel erger was. Ik heb je manuscript dan ook al elf jaar niet meer ingekeken. Morgen zul je jouw mapje met Exact Neutraal weggooien, over twee jaar crasht de computer waar je nu achter zit, ik dacht dus dat het boek voorgoed verdwenen was. Tot ik het gister in mijn Hotmail-inbox terugvond. Je hebt elk hoofdstuk netjes naar jezelf verstuurd, dank daarvoor." (Hanna Bervoets in Donk and Meer 2016, 63)

CONTENTS

PUBLICATION STATEMENT

Four sections from this thesis have been partially published before:

- Sections 1.2, 'Logging writing processes with a keystroke logger', 1.3, 'Starting to collaborate', 3.1, 'Digital materiality', and 4.2, 'Revisions back on the screen', contain revised parts of the following publication: L. Bekius, "The Reconstruction of the Author's Movement Through the Text, or How to Encode Keystroke Logged Writing Processes in TEI-XML", *Variants* 15-16 (2021). In addition, pages 60, 61, 79, 80, 90, 91, 104, 112, 113, and 116 also contain smaller sections from this same publication.
- Section 4.3.2, 'Triggers and revision episodes', is partly based on the section "Conclusion: Levels of Granularity", and section 5.2.1, 'Creating coherence', contains a revised version of the section "The Keystroke Logging Draft: 'Mondini' by David Troch" in the forthcoming chapter: L. Bekius & D. Van Hulle, "The literary draft in the 21st century: from paper notebooks to keystroke logging", *Toward a Comparative History of the Literary Draft in Europe. Comparative History of Literatures in European Languages Series* (forthcoming). This chapter is co-authored, but the analysis and the writing for these sections was done by Bekius, supported by comments from Van Hulle.
- Section 6.1.3, 'Bogaert: Narrating setting', is based on a section published in: L. Bekius, "De geboorte van een plein. De analoge en digitale tekstgenese van het *incipit* in Gie Bogaerts roman Roosevelt", *Tijdschrift voor Nederlandse taal- en letterkunde* 137.2 (2021)

TRACK CHANGES WEBSITE

Accompanying this thesis is the website Track Changes, which contains the reconstructions of the writing processes central to this study. Whenever the dissertation refers to a writing session, the visualisation of this session can be found on the website, under the respective session of the named author. When cited from the writing processes, the number of the discussed writing action will be given (e.g. n8, or n59-60).

The link to the website is:
https://trackchanges-cmg.uantwerpen.be
For logging in, use the following credentials:
Username: tc_visitor
Password: Keystroke-4-keystroke

ACKNOWLEDGEMENTS

friendship. I also owe a debt of gratitude to Luuk Van Waes, without whom this project might not have started at all.

At the Huygens Institute, thanks to my colleagues from the Computational Literary Studies group and beyond, including Jan Bloemendal, Peter Boot, Peter Kegel, Hannah Busch, Joris van Zundert, Bram Oostveen, Sietske van der Veen, Marc van Zoggel. Apart from my first year in Amsterdam, I have mostly seen you 'behind the screen', but you are a great source of inspiration.

Dank ook aan de adviesraad van het Track Changes project – Saeed Al-Gariri, Willem Bongers-Dek, David Coppoolse, Walter van den Berg, Steven Claeyssens, Nele Hendrickx, Niels 't Hooft, Suzanne Meeuwissen, Annette Portegies, Miriam Rasch, Thorsten Ries, Everdien Rietstap, Fabian Stolk, Xavier Roelens – dank voor jullie interesse in het project, het meedenken en jullie feedback. Dank ook aan het Literatuurmuseum en Letterenhuis, om mij inzicht te geven in hun *born-digital* collecties, en de vele floppydisks die ik heb mogen bestuderen. Isabelle Van Ongeval, dank voor al de interessante en motiverende gesprekken die we al hebben kunnen voeren over het preserveren en ontsluiten van digitaal literair erfgoed.

Een wel hele speciale dank gaat uiteraard uit naar de schrijvers die hebben meegewerkt aan het Track Changes project: Gie Bogaert, Niels 't Hooft, Renée van Marissing, Jens Meijen, Vincent Mercx, Jente Posthuma, Roos van Rijswijk, Aafke Romeijn, Arnoud Righter, Dirk Speelman, David Troch and Ellen Van Pelt. Zonder jullie bestond dit onderzoek niet; jullie durf om de uitdaging aan te gaan om jullie schrijfprocessen vast te leggen en door ons te laten uitpluizen is bewonderenswaardig. Bedankt voor jullie vertrouwen.

Zonder mijn familie was dit uiteraard ook niet mogelijk geweest. Greetje Bijma, je bent de beste moeder die ik me kan wensen – bedankt mij te maken tot wie ik ben. Doeke, hopelijk maakt dit wat duidelijk waar je zus zich al die tijd mee bezig heeft gehouden. Maak jullie borst maar nat om dit eens GOED TE LEZEN. Dit proefschrift draag ik op aan mijn vader Willem Bekius, ik hoop (maar weet zeker) dat je trots zou zijn geweest en dat je het een "mooi stukje" zou hebben gevonden. Mijn schoonfamilie, Piet en Erna, bedankt voor jullie steun en dat ik jullie zoon mocht meevoeren over de verre grens met België.

Jan, bedankt dat je in elke stap die ik neem naast, achter, of – wanneer nodig – voor me staat. Je hebt de perfecte omgeving gecreëerd om dit onderzoek te kunnen volbrengen; dit is ook jouw verdienste. Ik houd van jou.

CHAPTER 1. INTRODUCTION

LITERARY WRITING AND THE OMNIPRESENCE OF DIGITAL DEVICES

> *I once suggested that I had written half a book drunk, the other half sober. Here I estimate that about nine-tenths of the words in this book were written "free," the other one-tenth, hooked up to a hospital-grade breast pump: words piled into one machine, milk siphoned out by another. (Nelson 2016, 100)*

In our present day and age, we are assisted by devices in almost everything we do, from writing to breastfeeding. Although – as Maggie Nelson reminds us – these two can fortunately be done simultaneously, this dissertation focuses on only one of them: writing, or, more precisely, writing on a digital device. In the photography book *Where the Magic Happens: Writing Rooms* (2016), Huib Afman and Cornee Van der Stelt offer a glimpse into the writing rooms of Dutch and Flemish authors, including Saskia De Coster, Tom Lanoye, Gustaaf Peek, Charlotte Mutsaers and Annelies Verbeke. The photographs provide a partial answer to the following question: Where is present-day Dutch and Flemish literature being written? This 'magic' takes place in a variety of writing rooms; some desks have a view of the city, others look out on to the countryside, while other authors stare at a blank wall. Some of these rooms contain stacked bookshelves, while others show no sign of reading. There are modern, industrial, and Bohemian interiors, and some desks are tidy, while others are disordered. One author even has a habit of eating a lot of *speculaas* while writing.

More importantly for the outset of this research, most of the photographs also reveal the media that the authors use for writing. Contrary to the impression given by the cover – it features a photograph of a typewriter – 37 out of the 45 photographs of writing rooms reveal the presence of a laptop or desktop PC. The computer is absent only in the photographs of the writing rooms of A.H.J. Dautzenberg (1967), Midas Dekkers (1947), Kristien Hemmerechts (1955), Elvis Peeters (1957), Jan Siebelink (1938), Peter Terrin (1968), Christophe Van Gerrewey (1982) and Dirk Van Weelden (1957). As indicated by Dautzenberg's printer, however, there must be one somewhere,

and Hemmerechts explains that she has two writing desks: "First I worked upstairs, but after Herman's [Herman De Coninck, her late husband] death in 1997, I moved into the conservatory, where he often used to sit. I work at the computer at his desk and at the desk against the wall I write by hand" (67).[1] Like Van Gerrewey, Peeters has persevered in writing by hand, although he types everything on the computer at a later stage in the process (123). Both Terrin and Van Weelden collect typewriters and use them to write. Dekkers writes mostly on an electronic typewriter: "The only modern objects are the sound system and my electronic typewriter. I'm addicted to my Brother CE-60, but also have three IBM typewriters to hand, just in case. The IBM is the most beautiful machine that God has ever created" (47). In addition, Siebelink states that he has "lost the battle for the computer and the Internet ages ago" (139). These exceptions aside, the assumption confirmed by *Where the Magic Happens* is that most present-day Dutch and Flemish authors indeed use a computer in their writing process.[2]

This should obviously come as no surprise; it would be more startling if the same number of authors were still using typewriters or writing only by hand. Many people cannot imagine writing without recourse to the digital environment at some point. Computers, word processors, email, the Internet and mobile phones, have not only "become common artefacts in our daily lives" (Boomen et al. 2009, 7); they have also been incorporated into the writing practice of present-day literary authors. Given that digital devices are "so widely distributed and used that we take them for granted" (7), scholars should not overlook the consequences of these digital devices for one branch of literary studies in particular: genetic criticism, which is interested in the writing process of literary works. In this study, I examine the consequences of the digital working methods of present-day authors for genetic criticism. This type of research has always been carried out based on the material traces that have been preserved. As authors are now working predominantly within a digital environment, however, scholars are faced with a number of questions. What traces does the digital writing process leave behind? What form do these traces take? What methods can be used to document and analyse these digital traces? What do they have to say about the genesis of the text in question?

A fictional account of the digital writing practices of a present-day literary writer can serve to illustrate the challenges confronting genetic criticism. In 2020, the Dutch author Thomas Heerma van Voss published the novel *Condities* ['Conditions'], in which the main narrator (Vincent Pek) also publishes a novel entitled *De diagnose* ['The

---

[1] In *Where the Magic Happens*, English translations of the authors' reflections on their writing rooms – translated by Helen Simpson – are presented alongside the original Dutch. Throughout this study, all the translations of citations from Dutch are my own, unless otherwise indicated.

[2] The same conclusion can be drawn from Bert Bevers' photographs of Antwerp authors at their writing desks, see http://detafelvan1.blogspot.com.

Diagnosis']. The real novel, *Condities*, describes the writing process of the fictional novel, *De diagnose*, which is written in Microsoft Word. The fictional novel is based on a story that Pek had included in what was to become his new collection of stories. Multiple versions of this collection already exist, under the titles '*Wie houdt ons tegen (nu echt nieuwe boek – elf verhalen)*'[3] and '*Wie houdt ons tegen versie 3 (nu echt nieuw boek – elf verhalen)*'.[4]

Pek submits the manuscript of the collection of stories to his publisher, but his editor is mainly interested in only one of the stories: "Een voorjaarsdag" ["A spring day"], in which the main narrator, Gregor, experiences severe abdominal complaints during a walk. The editor suggests that Pek should expand this story. At home, Pek tries by copying the story into a new Word document, but he is unable to concentrate. In a later attempt, Pek recognises the story's potential, which triggers the creative process: "Nu hij zijn verhaal vaker herleest ziet hij steeds meer uitbreidingsmogelijkheden" (Heerma van Voss 2020, 107).[5] All these possibilities are all listed in the novel. Once on a roll, more and more passages are added: "De nieuwe bladzijden blijven zichzelf aandienen, vijf, zes per dag" (139).[6] The actual novel contains a number of suggestions and questions from the editor as well, and it describes how Pek incorporates them in a file named '*De diagnose (dit is echt het nieuwe boek)*'.[7] One of the revisions concerns the deletion of a scene about Gregor's mother which, on reflection, slows down the narrative. Instead, Pek adds that Gregor has another bout of abdominal pain. He also uses new Word files to experiment with a change from third-person to first-person narration, again on the advice of his editor. He copies the opening chapter into a new Word document and changes 'he' to 'I':

> Later die avond […] opent hij het tweede Word-bestand opnieuw. Ernaast, in een ander venster, bekijkt hij *De diagnose (dit is echt het nieuwe boek)*. Hij tuurt om en om naar de teksten. Hij, ik, ik, hij. En Pek vervangt, met een snelheid die hij van zichzelf kent maar die hem toch bevreemdt, een tempo waarvan hij niet zeker weet of het voortkomt uit gretigheid of uit afkeer, nog enkele keren 'hij' door 'ik', tot hij alles ongedaan maakt, zijn laptop dichtklapt en wil dat er niets meer verandert. (273)[8]

---

3   Translation: "Who is stopping us (now really new book – eleven stories)"
4   Translation: "Who is stopping us version 3 (now really new book – eleven stories)"
5   Translation: "Now that he rereads his story more often, he sees more and more possibilities for expansion"
6   Translation: "The new pages keep presenting themselves, five, six a day"
7   Translation: "The diagnosis (this is really the new book)"
8   Translation: "Later that evening […] he reopened the second Word file. Next to it, in another window, he looks at *The diagnosis (this is really the new book)*. He gazes alternately at the texts. Him, me, me, him. At a speed that he knows he is capable of, but which nevertheless surprises him, a pace of which he is not sure whether it comes from eagerness or aversion, Pek replaces

During the process, Pek changes his mind and, in the end, no traces of this experiment of changing the perspective are saved – everything is undone.

Although the writing process described by Thomas Heerma van Voss is fictional, it seems to exemplify the way present-day authors work. Let us consider this description from the perspective of a scholar interested in the genesis of *De diagnose*. The descriptions of Pek's writing process suggest that the writing process of *De diagnose* may still be reconstructed to some extent, while simultaneously pointing out the lacunae. First, we encounter multiple versions of the collection of stories, even though only one version of *De diagnose* is mentioned. Would it still be possible to reconstruct how this work came into being? The insertion of new passages is frequently mentioned in the novel. If only one version remained, however, the sequence of these insertions might not be so easy to deduce. Nor would it be possible to study immediate revisions. Moreover, although Pek had several versions of his collection of short stories, there is no explicit mention of several versions of *De diagnose*. Are all the passages he deleted lost forever? Finally, the experiment with the pronouns is not documented, while this experiment provides insight into an important decision-making process taking place during the writing of the novel. All in all, the documents saved by Pek do not show the writing operations. As pointed out by the authors of "Digital Materiality: Preserving Access to Computers as Complete Environments", it can be concluded that "this means that a writer working today will not and cannot be studied in the future in the same way as writers of the past" (Kirschenbaum et al. 2009, 105). If it cannot be studied *in the same way*, however, does this necessarily mean that it cannot be studied *at all*?

1.1 THE END OF AN ERA FOR GENETIC CRITICISM?

In an essay with the evocative title "Genetic Textual Editing: The End of an Era" (2009), Marita Mathijsen observes that the computer demands "a completely different way of working with regard to the genesis of a literary text" (233). She states that it would no longer be possible to describe the genesis of a work by a present-day author, "because the process of working is no longer materially recorded" (235). More than a decade earlier, in 1996, Derrida made a similar statement, calling the use of a word processor in the writing process a "provocation for genetic criticism":[9]

> you no longer retain the slightest visible or objective trace of corrections made the day before. Everything – the past and the present – everything can thus be

---

'he' with 'I' a few more times, until he undoes everything, closes his laptop and wants nothing to change."

9 In the research project 'Derrida Hexidecimal' conducted at ITEM (Institut des textes et manuscrits modernes), digital forensic methods are in fact used to recover traces of Derrida's writing process from his old Macintosh (see Crasson, Lebrave, and Pedrazzi 2019; Crasson et al. 2022)

locked, cancelled, or encrypted forever. Previously, erasures and added words left a sort of scar on the paper or a visible image in the memory. There was a temporal resistance, a thickness in the duration of the erasure. But now everything negative is drowned, deleted; it evaporates immediately, sometimes from one instant to the next. (Derrida 2005, 24)

The same is suggested by David Stephen Calonne, who notes that the study of revision has changed dramatically because of word processors and computers: "The tracks of revision, so to speak, will now be covered because much of the process will be 'erased' by the fact that the writer can now make immediate changes to the text on the computer screen and the corrections will normally not be saved" (Calonne 2006, 161). The only way we could still investigate revision, Calonne continues, is when the author makes printouts and revises by hand. Pierre-Marc de Biasi provides a perspective on the issue from the other side of the spectrum. In an article entitled "Pour une génétique généralisée: l'approche des processus à l'âge numérique"(2010), he speaks out against the doomsayers who had declared: "No more manuscripts, no more rough drafts? It is thus the end of genetics!" (De Biasi quoted in, and translated by, Vauthier 2016, 164).[10] According to De Biasi, quite the contrary is true: the computer will keep track of all commands in the given writing process, as long as the memory is not deliberately overwritten (De Biasi 2010, 171; Vauthier 2016).

Bénédicte Vauthier summarises the reactions to De Biasi's optimistic article in "Genetic Criticism Put to the Test by Digital Technology" (2016). In an open letter to De Biasi, Louis Hay and Jean-Louis Lebrave state that two crucial questions must be answered first. The first focuses on the authors: How do authors use the computer, and what is the effect thereof? To realise the full benefit of the possible digital traces, authors must indeed be aware of their digital writing habits and their implications. The second question shifts the focus to genetic scholars: Will they be able to construct scientific objects from the computer data? (Hay and Lebrave 2010, 155). When the objects studied by genetic scholars change, the methodology must therefore change along with it. In addition, Vauthier states that the adoption of digital tools and methods is accompanied by human and technological difficulties. This stance has also been expressed in another article, in which Lebrave (2011) discusses the work of Thorsten Ries – in this case, Ries' analysis of digital files from Michael Speier's poem *exit st. Nazaire* (Ries 2010) – as a possible route for genetic criticism. Using digital forensic methods, Ries investigates the hard disk, exposing recoverable data that could be used to reconstruct the writing process. On the one hand, seeing the potential of this type of research, Lebrave coins the neologism *génétique inforensique* [genetic inforensics] to

---

[10] Original: "Plus de manuscrits, plus de brouillions? C'est donc la fin de la génétique" (De Biasi 2010, 171).

describe it. On the other hand, he acknowledges that such research requires expertise in the field of computer science (Vauthier 2016, 166).

Whether we adopt an optimistic or pessimistic stance on the impact of the digital medium on the continuation of the practice of genetic criticism, it is important to acknowledge the need to anticipate the phenomenon of the digital genesis of texts (Linkès 2017). The digital files created in this process require new techniques, both for collecting and constructing the digital dossier and for studying it (Vauthier 2016; Linkès 2017). As stated by Gabriele Wix, literature whose writing process takes place within a digital environment has "catapulted" genetic criticism "away from traditional philological tools and methods to the needs of a digital analysis demanding software which is not yet at the disposal for general use" (2016, 133). Studies conducted by a number of authors have already revealed a variety of techniques for examining born-digital writing processes, thus proving that the digital writing process still leaves sufficient traces to allow genetic analysis (Ries 2018; 2017; Kirschenbaum and Reside 2013; Kirschenbaum 2012; Crombez and Cassiers 2015; Vásári 2019; Vauthier 2016; Fenoglio 2009).[11] Irène Fenoglio (2009) limits her research to printouts of digital files. Bénédicte Vauthier (2016) takes on the task of investigating the digital files saved by the Spanish writer Robert Juan-Cantavella during the writing of his novel *El Dorado* (2008). She compares digital documents and analyses the tree structure of the folders and other metadata such as the file title and creation date. The aforementioned digital forensic methods also yield promising results. We can therefore conclude that the perspectives of Mathijsen (2009) and Derrida (2005) must be corrected.[12] Regardless, Mathijsen does pose some important questions about genetic criticism and scholarly editing with respect to born-digital works of literature, which therefore return often throughout this study.

With few exceptions, however, immediate revisions and corrections of typing errors often remain irretrievable. The absence of immediate revisions and, more generally, the revisions visible *on* the page, is also mentioned by Vauthier: "It cannot be denied – and this is the main difficulty – that word processing flattens out, even crushes the two-dimensionality of the page, and thereby hides the abundant writing operations which manuscripts genetics seeks to unravel" (2016, 166). This is also the reason why Lebrave remarks that genetic criticism may become the study of variance *between* different versions: "the inforensic genetics thus risks having to give up definitely on being a poetics of processes in order to settle for being a poetics of transitions between

---

[11] In addition, Veijo Pulkkinen currently combines theories and methods from genetic criticism and digital forensic research in the examination of the digital archives of Kalle Päätalo and Christer Kihlman, see: https://researchportal.helsinki.fi/fi/projects/digitaalinen-syntyprosessi-geneettinen-tutkimus-christer-kihlmani

[12] Mathijsen has revised her opinion in private conversations with Thorsten Ries (Ries 2022)

stages" (Lebrave quoted in, and translated by Vauthier 2016, 168).[13] In this research, however, I shift the focus to a method that, when used by the author at the moment of composition, allows genetic criticism to remain a 'poetics of process' rather than becoming a 'poetics of transitions between stages', as envisaged by Lebrave: keystroke logging.

1.2 LOGGING WRITING PROCESSES WITH A KEYSTROKE LOGGER

Four years before publishing *Condities* – with the fictitious account of Pek's writing process – Thomas Heerma van Voss participated in the pilot project entitled 'Het Literaire Werk 2.0' ['The Literary Work 2.0']. This project paved the way for another project, entitled 'Track Changes: Textual Scholarship and the Challenge of Digital Literary Writing', within which my research is integrated. For 'Het Literaire Werk 2.0', four Dutch authors each wrote a short story and recorded their writing processes with Inputlog, a keystroke logging tool. This tool logs the writing operations that take place in a word-processing environment with which the author is already familiar: Microsoft Word (Leijten and Van Waes 2013; Van Waes and Leijten 2020). This software is the most widely used word processor amongst authors in the Dutch literary field (Buschenhenke 2016). Each time an author activates Inputlog to start a new writing session, the Word document in which the author is working is saved in the background, in a folder containing the date and number of the writing session.[14]

The Word document is saved again when the author ends the writing session by de-activating Inputlog. This results in a specific version of the text for each session (i.e., a 'session version'), which shows the gradual evolution of the text. Inputlog does more than simply save Word documents, however, as it records every keystroke and mouse movement with a timestamp whenever the program is running (Leijten and Van Waes 2013). The timestamp makes it possible to reconstruct the order in which the text was typed, how long it took the author to write a specific word or sentence, and the pauses during the text production. Table 1.1 shows the logged data as displayed in the General Analysis, which is the most fine-grained level of output: "it is a linear, event-based, vertical representation of the text production" (Van Waes and Leijten 2020, 29). While writing, authors retain control of the process: they can start and stop the logging whenever they choose, and the data is stored on their local PC or laptop.

---

[13] Original: "[l]a génétique inforensique risque donc de devoir définitivement renoncer à être une poétique des processus pour se contenter d'être une poétique des transitions entre états" (Lebrave 2011, 145)

[14] This description of Inputlog also appears in Bekius 2021a.

The data collected with Inputlog thus provide fine-grained information about the writing process. As Thomas Heerma van Voss tweeted during the pilot project, every little detail and change in the text is recorded: "Toch vreemd, zo'n app die bijhoudt welke komma's ik verplaats en welke voornaamwoorden ik vervang" (Thomas Heerma van Voss, on Twitter, 29 February 2016).[15] Returning to the example of *Condities*: if the character Vincent Pek had followed in his creator's footsteps and had also recorded his writing process for *De diagnose* with a keystroke logger, it would have been possible to reconstruct his experiment with changing pronouns – had it not been fictional.

The pilot project 'Het Literaire Werk 2.0' and the research project 'Track Changes' are not the only projects to make use of a particular software to record *literary* writing processes. Between 2014 and 2018, the English novelist C. M. Taylor collaborated with the British Library to record the writing process for his novel *Staying On* (2018) using the keystroke logging software Spector Pro. Driven by the "lost drafts" in digital writing and the loneliness of the writing process, Taylor contacted the digital curation team of the British Library, and together they decided to record his writing process (Taylor 2018). According to Taylor, this setup helped reduce loneliness while writing: "Somehow the writing felt collaborative, not only because the software was recording me, but also because of the digital curation team who were taking the data" (2018).[16] The data would belong to the British Library and the resultant book to Taylor. On the British Library's 'English and Drama blog', Taylor quotes Jonathan Pledge – a curator of contemporary archives at the British Library – stating that the software used in the project "seem[ed] to have been specifically designed for low-level company surveillance of employees, potentially without their knowledge" (2018). The ultimate dataset is composed of screenshots and keystroke logs. The screenshots can be consulted either individually (in JPG and BMP files) or as an AVI file, which plays the individual captures as a film. The dataset is publicly available under a Creative Commons BY license for research, teaching, art projects or anything else.[17]

---

[15] Translation: "Strange, though, such an app that keeps track of which commas I move and which pronouns I replace". Although the tweet has since been deleted, it can still be found in the educational material from Het Literatuurmuseum related to 'Het Literaire Werk 2.0'.

[16] Authors often reach out to their readers during their writing process. For example, the Australian author Max Barry began a self-published serial on his website, posting one page every day, on which readers commented. An expanded version of this novel was eventually published as *Machine Man* in 2012 (see also Kirschenbaum 2016). A 17-part series is available on YouTube, which shows the author Robert Olen Butler writing a short story in 2012 (https://youtu.be/vIcnmiT0Mc8). In Week 43 of 2020, the Dutch author Lisa Weeda streamed the writing process for her novel on the streaming channel Twitch.

[17] See: https://data.bl.uk/cmtaylorkeylogging/. I also describe this collaboration between Taylor and the British Library in ""The Reconstruction of the Author's Movement Through the Text, or How to Encode Keystroke Logged Writing Processes in TEI-XML" (Bekius 2021a), and it is also touched upon by Van Hulle (2021; 2022).

#ID	Event Type	Output	Position	Document length	Character Production	Start Time (ms)	Start Clock	End Time (ms)	End Clock	Action Time	Pause Time	Pause Location
1	mouse	Movement	0	0	0	1360	00:00:01.360	1625	00:00:01.625	265	0	INITIAL
2	keyboard	LSHIFT	0	1	1	39360	00:00:39.360	39553	00:00:39.553	193	37735	INITIAL
3	keyboard	V	0	1	1	39445	00:00:39.445	39591	00:00:39.591	146	85	BEFORE SENTENCES
4	keyboard	e	1	2	2	39666	00:00:39.666	39770	00:00:39.770	104	221	WITHIN WORDS
5	keyboard	r	2	3	3	39804	00:00:39.804	39892	00:00:39.892	88	138	WITHIN WORDS
6	keyboard	h	3	4	4	40402	00:00:40.402	40477	00:00:40.477	75	598	WITHIN WORDS
7	keyboard	a	4	5	5	40518	00:00:40.518	40593	00:00:40.593	75	116	WITHIN WORDS
8	keyboard	a	5	6	6	40666	00:00:40.666	40765	00:00:40.765	99	148	WITHIN WORDS
9	keyboard	l	6	7	7	41323	00:00:41.323	41396	00:00:41.396	73	657	WITHIN WORDS
10	keyboard	SPACE	7	8	8	42352	00:00:42.352	42449	00:00:42.449	97	1029	AFTER WORDS
11	keyboard	LSHIFT	8	9	9	43144	00:00:43.144	43388	00:00:43.388	244	792	COMBINATION KEY
12	keyboard	G	8	9	9	43316	00:00:43.316	43392	00:00:43.392	76	172	BEFORE WORDS
13	keyboard	i	9	10	10	44033	00:00:44.033	44104	00:00:44.104	71	717	WITHIN WORDS
14	keyboard	d	10	11	11	44224	00:00:44.224	44318	00:00:44.318	94	191	WITHIN WORDS
15	keyboard	s	11	12	12	44463	00:00:44.463	44573	00:00:44.573	110	239	WITHIN WORDS
16	keyboard	RETURN	12	13	13	45848	00:00:45.848	45943	00:00:45.943	95	1385	AFTER PARAGRAPHS

Table 1.1: Example General Analysis from Inputlog

In the United States, the online literary journal *Midst* was launched in 2019.[18] Founded by Annelyse Gelman, *Midst* publishes poems "in the form of interactive timelapses, giving readers and writers unprecedented access to the creative process" (Gelman n.d.). By default, readers will see the finished poem, but they can also rewind the writing process to see how the poem was written: "from the blank draft page to the final draft, including the entire revision process" (Gelman n.d.). *Midst* aims to reach a broad audience, and it is intended for people who love poetry or who "just don't get it", for aspiring poets and poets who "want a better understanding of their own writing process", and for educators in search for examples of the craft and revision, as well as for scholars "researching digital humanities and the archive" and anyone "who ever wondered how technology is impacting how we write and how we think" (Gelman n.d.). The contributions to the journal are promising, as they show the willingness of poets to also share their writing processes and not only their products.

Scholars have previously identified the potential of keystroke logging for text genetic research (Vauthier 2016; Van Hulle 2014). With the exception of a few case studies (Buschenhenke 2016; Johansson 2022), however, no extensive research has been conducted to explore the possibilities of keystroke logging for genetic criticism applied to born-digital works of literature. In this study, my aim is to contribute to the field of genetic criticism in relation to born-digital literature by applying the methodology to keystroke logging material. I describe how the methodology of genetic criticism can be applied to keystroke logging material and how this fine-grained material generates new knowledge about the genesis of texts. As the results clearly show, keystroke logging offers many new opportunities for genetic criticism. It also poses obstacles, however, not least of which being the large amount of data. The abundance of data gathered with keystroke logging is one of the pitfalls addressed by Vauthier: "If these programs allow us to analyse (real of even film) time in a minute way, they will also suffocate the researcher in the abundance of data" (2016, 167). This pitfall has also been observed by Dirk Van Hulle: "we have so much material that the biggest challenge is not the gaps in the archival record but the abundance of data" (2021, 231). For this reason, I set out to investigate how to cope with this wealth of data and interpret it within the framework of genetic criticism.

1.3 STARTING TO COLLABORATE

Most genetic studies work with self-archived born-digital material received directly from authors (Vauthier 2016; Crombez and Cassiers 2015; Vásári 2019). This stresses the importance of such collaboration. With regard to born-digital archives, Catherine Hobbs advocates collaboration between writers, scholars and archivists "to understand

---

[18] See: http://midst.press

the relationship between writers, their documentation and their creative vision" (Hobbs quoted in Gooding, Smith, and Mann 2019, 384). Collaboration with literary authors is also important, as they play a significant role in the preservation of their digital documents (Micunovic, Marčetić, and Krtalić 2016). My present study extends this argument and tries to strengthen such collaboration between literary scholars, authors, and literary and archival institutions.

Collaboration with both authors and literary and archival institutions is made possible by the larger project in which this research is situated: 'Track Changes: Textual scholarship and the challenge of digital literary writing' (hereinafter, 'Track Changes'), funded by the Dutch Research Council (NWO). With the aim of broadening research coverage from short professional writing processes in educational and corporate contexts to include long-term literary writing processes, the team behind the keystroke logging tool Inputlog (Leijten and Van Waes 2013) at the University of Antwerp collaborated with the author Gie Bogaert to log the writing process for his 10th novel, *Roosevelt* (2016). After Bogaert's writing process had been recorded, collaboration was established between the Inputlog team (research domain Professional Communication and Digital Writing Processes), the Centre for Manuscript Genetics (University of Antwerp) and the Literary Department at Huygens Institute (KNAW), to adequately address the analysis of this literary writing process.[19] This led first to the aforementioned pilot project 'Het Literaire Werk 2.0', and later to the research project 'Track Changes'. The latter project thus already embodies the collaborative aspect, as it is a collaboration between the Huygens Institute in Amsterdam (Netherlands) and the University of Antwerp (Belgium), combining expertise from the fields of textual scholarship, textual analysis, genetic criticism and writing process research. The dissertation of my colleague Floor Buschenhenke examines present-day literary writing, primarily within the field of writing process research, with a focus on how dynamics of the writing process (e.g. pauses and revisions) make it possible to describe writing strategies.

Importantly, the project also involves collaboration with authors by asking them to participate and log their writing processes with Inputlog. During the project, 11 Dutch and Flemish authors participated by logging the writing process for a short story.[20] In addition to Bogaert's writing process, this research discusses the writing processes of four of these authors: Jente Posthuma, Roos van Rijswijk, David Troch, and Ellen Van Pelt. In selecting these authors, I have tried to achieve a relatively equal distribution between Flemish (Bogaert, Troch, Van Pelt) and Dutch (Posthuma and

---

[19] This description of the 'Track Changes'-project has also been published as part of Bekius 2021a.

[20] The Dutch authors are Renée van Marissing, Jente Posthuma, Roos van Rijswijk, Aafke Romeijn, Arnoud Righter, and Niels 't Hooft. The Flemish authors are Ellen Van Pelt, David Troch, Dirk Speelman, Jens Meijen, and Vincent Mercx.

Van Rijswijk) authors; between female (Posthuma, Van Rijswijk and Van Pelt) and male (Bogaert and Troch) authors; and between works that have been published (Bogaert, Posthuma and Van Rijswijk) and those that have not yet been published (Troch and Van Pelt). As clearly shown in this study, the writing processes of these authors differ as well. Another distinction is obviously that between the writing process for a novel and that for a short story. These different narrative forms lead to different kinds of writing processes (Heeks 2012, 19). For example, a short story, "does not require designing or 'planning' in quite the same way that a long novel might" (2012, 21). Before moving on to the content of this dissertation, I introduce the authors, without whom this study would not have been possible.

1.3.1 Gie Bogaert

Gie Bogaert (1958) is a Flemish author who made his debut in 1987 with the collection of short stories *Klein Berlijns drama*. This was followed three years later by his first novel *Wat kwaad doen de tovenaars?* He later published *Keizer Doede* (1992), *Wat we doen met de liefde* (1995), *De liefdeverzamelaar* (1998), *Nathan Meyer, vrouwenwandelaar* (2000) and the 'Trilogie van het haalbare': *Hemelstof* (2004), *Opklaringen* (2006) and *Luchtgezichten* (2010). *Noora's dwaling* followed in 2013. From July 2013 to December 2015, Bogaert logged the writing process for his 10th novel *Roosevelt* (2016). His most recent novel, *Onvoltooid*, was published in 2022. Bogaert has always combined writing with teaching, including during the writing of *Roosevelt*. Until June 2020, Bogaert taught Dutch and English at the St Gabriël's College in Boechout, Belgium. He also taught creative writing at the SchrijversAcademie in Antwerp.

   Roosevelt is a kaleidoscopic city novel with a public square – the Franklin Roosevelt Square in Antwerp, Belgium – as its main narrator. The novel is set within the timeframe of one hot summer day, on which a football match between Belgium and the Netherlands is scheduled in the evening. During this day, an anonymous boy is writing a story in chalk on the pavement of the square. This story is printed in the top margin of the novel.

   The novel is divided into 62 sections, each narrated by one of the 13 characters. Franklin, the square, seems to address the reader, speaking directly to them. In fact, however, he is in fact talking to Carla, who must count the number of passengers getting on and off the tram for transport company De Lijn. For this reason, she is in the same spot on the square all day long, and Franklin tries to entertain her by letting her hear the stories of other passers-by. She also tells her own story in six chapters, in which she gradually reveals the secret she is keeping: she has been sexually assaulted on her way to the job the day before.

   Franklin thus alerts Carla to the other characters who have their occupations in the square during the day. One of them is Helge, a female general practitioner who has

spent the night with her childhood love while her husband is away for a conference. Remy, the tram driver, observes the city through the window of his tram and thinks about his disabled son and his wife's complaints about his weight. Another character is Felix, a homeless man who has lost his leg in a tram accident. Janine is the toilet lady at Hotel Terminus, who talks about the people who use the toilet and how she is abused by her husband. Ana is a schoolgirl of Polish origin who fell into the hands of a lover boy. Faraaz is an Iranian waiter at Hotel Terminus, and his little sister was forced into marriage and drowned herself – a secret nobody knows, except his parents. Then, Jella, a dresser at the Opera talks about her unusual youth, her abortion and her hope that her move to a new apartment will offer her an opportunity to start over. The literature professor, Gaard, is having an extramarital relationship with a (former) student named Fran. Fran and Shaun – a football supporter – are both supporting characters. Finally, there is Barry. He has distrusted women since his mother's suicide and the break-up with his girlfriend. This leads him to an act of desperation: he shoots Janine, Faraaz, Carla, Helge, Remy, Felix, Jella and Ana from a window of Hotel Terminus. But nothing is as it appears: Franklin made up all the stories. Nothing really happened.

1.3.2 Jente Posthuma

The Dutch author Jente Posthuma (1974) has published several stories in literary journals, including *De Revisor*, *De Gids*, *Hollands Maandblad*, *Das Magazin* and *Oogst*. She published her debut novel, *Mensen zonder uitstraling*, in 2016. Her second novel *Dingen waar ik liever niet aan denk* appeared in 2020. And in 2023, she published *Heks! Heks! Heks!* Posthuma studied Literary Studies and started as a journalist, doing lengthy interviews for outlets including *De Groene Amsterdammer*, *nrc.next* and *De Volkskrant*. Between the start of August and the end of October 2020, she logged the writing process for her story "En daarom haten ze zichzelf". The story was published in issue 1/2021 of the Dutch literary journal *De Gids* (Posthuma 2021).

The title – "En daarom haten ze zichzelf" ("And that is why they hate themselves") – covers the essence of the story: each character has something they hate about themselves. The female first-person narrator, her husband and son are spending a two-weeks holiday with friends who have a summer house in a southern country. Two incidents happen: 'het vinger-klem-incident' (the stuck-finger incident), in which the finger of her son gets stuck in a pedalo, and the 'het bijna-stikken-incident' (the near-suffocation incident), in which her son nearly suffocates on a piece of an ice lolly. Further, the narrator criticises herself and the world around her.

1.3.3 Roos van Rijswijk

Roos van Rijswijk (1985) is a Dutch author who debuted in 2016 with the novel *Onheilig*. She has published several short stories in a wide range of literary journals, including *De*

Revisor, *De Gids*, *Tirade*, *De VPRO-Gids*, *De Correspondent*, *DW-B*, *De Groene Amsterdammer*, *De SLAA*, *Aicha Qandisha*, *Das Magazin*, *Slang*, *De Optimist* and *Hard//Hoofd*. In 2021, she published a collection of short stories *De dwaler* (Rijswijk 2021), which includes the short story "Zorgvlied". Van Rijswijk logged the writing process for this story in December 2020.

In "Zorgvlied" the first-person narrator takes a walk with her mother through the Zorgvlied cemetery in Amsterdam during one of the COVID-19 lockdowns. During the walk, they are first accompanied by the ghost of her grandmother, after which her great-grandmother and great-great-grandmother also join them. The narrator can see and converse with them, but her mother cannot. This is, however, not the first time she has perceived ghosts; her grandmother has visited her before.

1.3.4 David Troch

David Troch (1977) is a Flemish poet and short-story author. He has published four collections of poetry, the most recent one being *voor jou wou ik een huis zijn* (2021). In 2018, he published his first collection of short stories: *Rue des Regrets*. Other short stories by Troch have been published in literary journals, including *Deus Ex Machina*, *Gierik & Nieuw Vlaams Tijdschrift*, *Plebs* and *Serial Thriller*. From mid-August until mid-September 2020, he logged the writing process of his short story "Mondini".

The story is set in a post-apocalyptic future. A man has fled from the drought and heat caused by climate change and lost his partner and daughter along the way. He is now trying to survive on his own, in a small apartment in the mountain village named Mondini in Italy. The narrator describes his daily whereabouts and how he tries not to lose his mind.

1.3.5 Ellen Van Pelt

The Flemish author Ellen Van Pelt (1980) has published short stories in several literary journals and debuted in 2015 with the novel *Drift*. For more than 10 years, she worked as a psychologist. Her most recent publication is *Deze wereld is geen ergernis waard* (2020), her biography about the Flemish author and journalist Roger Van de Velde. From late November 2020 until mid-January 2021, she logged the writing process for her short story "Dauphin".

In the story the first-person narrator is on holiday with her son in Djerba, Tunisia. They go on a daytrip with a 'pirate ship' to Flamingo Island. It is the first time she must fulfil the role of both mother and father on holiday, as her partner has recently passed away. During the trip, she is unexpectedly helped by Mustafa, who takes over some 'fatherly' tasks, like carrying her son on his back after he refuses to go any further during their walk on the beach. She manages to get through the day and is rewarded by the sight of dolphins on the way back.

1.4 RESEARCH QUESTIONS AND OUTLINE

The writing processes that led to the five literary texts described above are the focus of this study. When the authors were writing these texts, the text was represented as "a continual, smooth textual flow" on their screens (Sullivan 2013, 286). The computer screen is often conceived as *pars pro toto* for computers, both in ordinary speech and in the research field of new media studies. As observed by Marianne van den Boomen, "This tendency could be explained by the ability of the GUI [Graphical User Interface] to swallow up all other components in its visual representations, thus rendering irrelevant what remains invisible" (2014, 34). The three interface devices – screen, keyboard, mouse – "are indispensable for contemporary PC operations" and, according to Van den Boomen, "there is no reason to privilege one above the other" (33). In this research, I want to look beyond the screen, by also considering the mouse movements and keystrokes gathered in the keystroke logging data. This may also affect our notion of the digitally produced text we now see merely as that "continual, smooth textual flow".

'Behind the computer screens': The use of keystroke logging for genetic criticism applied to born-digital works of literature investigates how keystroke logging software can serve as a method to bring all the writing operations 'back to the surface', or to bring what lies behind the screen onto the screen.[21] Drawing on the use of the keystroke logging material of Bogaert, Posthuma, Van Rijswijk, Troch and Van Pelt, this study aims to answer the following main research questions:

1) *How can existing methods and theories of textual scholarship be applied to reconstruct and analyse digital writing processes, and in which ways will the analysis of digital writing processes logged with keystroke logging software enrich the existing methods and theories?*

2) *How can digital and narratological text analysis be applied to the reconstructions of the writing process and intermediate text versions to create new knowledge about the genesis of texts and the creative process?*

To lay the theoretical and methodological foundations for answering these questions, Chapter 2, "The word processor as mediator", addresses the main components

---

[21] The title is also a reference to *Achter De Schermen* (Behind the Scenes) by the Flemish author Willem Elsschot, in which Elsschot analyses the writing process for another of his works, *Opdracht*. Dirk Van Hulle, Peter de Bruijn and Vincent Neyt conducted a thorough analytical exercise in reconstructing the genesis of this work, leading to an electronic edition in 2007. For that edition, they set out to disentangle the sequence of revisions, a difficult editorial task based on handwritten material. In a way, my research builds on this work by studying the genesis of texts by present-day Flemish and Dutch authors at a fine-grained level. By introducing software into the process of writing, however, the reconstruction of the order of revisions no longer has to be hypothetical. I am building on a tradition and taking advantage of digital advancements.

involved: digital writing, genetic criticism, cognitive writing process, and keystroke logging. It begins with an elaboration of the emergence of the word processor and the supposed effects thereof on the writing process (and particularly the literary process). I then discuss the theoretical framework of genetic criticism and models of the writing process as developed in cognitive writing process research. In this chapter, I also describe how the keystroke logging data were gathered, in addition to discussing some of the implications of keystroke logging.

As Marita Mathijsen rightly argues in "Genetic Textual Editing: The End of an Era" (2009), definitions within textual scholarship and genetic criticism should be revised to denote aspects of digital literary writing processes, as they are originally based on traces of the analogue writing process (Mathijsen 2009, 239). The third and fourth chapter therefore reflect on these key definitions and their adequacy for describing born-digital and keystroke logging material.

Chapter 3, "Traces of phases", delves into the question of what constitutes a digital genetic dossier.[22] In this respect, I begin with a discussion of digital materiality and what it means for our notion of the documents that constitute the digital genetic dossier. These digital documents represent different versions and stages of the text-in-progress. I therefore further define what I mean with a 'version' within this study. Finally, I examine the ways in which these versions may still provide insights into specific writing phases.

An important part of the genetic dossier consists of the rough drafts (in French, *brouillons*). Grésillon defines the rough draft as "manuscrit de travail d'un texte en train de se constituer; généralement couvert de ratures et réécritures" (Grésillon 1994, 241).[23] Rough drafts often contain *réécritures*, or 'rewritings', which are "toute opération scripturale qui revient sur du déjà-écrit, qu'il s'agisse de mots, de phrases, de

---

[22] De Biasi defines the genetic dossier as "the whole body of known, classified, and transcribed manuscripts and documents connected with a text whose form has reached, in the opinion of its author, as state of completion or near completion" (De Biasi and Wassenaar 1996, 31). The *avant-texte* in turn, is "the critical discourse by which the geneticist, having established the objective results of their analysis (transcription, relative dating, classification, etc.), reads them as successive moments of a process" (38). The terms 'genetic dossier' and *avant-texte*, however, are slightly ambiguous. Almuth Grésillon defines the genetic dossier as the "ensemble de tous les témoins génétiques écrits conservés d'une oeuvre ou d'un projet d'écriture, et classés en fonction de leur chronologie des étapes successives" (Grésillon 1994, 242). For Grésillon 'genetic dossier' is synonymous with 'avant-texte'. A synonym for 'genetic dossier', the term *avant-texte*, is sometimes also translated as 'pre-text', as in Marita Mathijsen's definition of pre-text: "Material which predates the first edition therefore also counts as a version. For this material, French scholarly editing, *critique génétique* has coined the term *avant-texte* or *pre-text*'" (Mathijsen 2009, 234). Throughout this study, I will use the term 'genetic dossier' for the collection of documents and other relevant material gathered for a text genetic analysis.

[23] Translation: "working manuscript of a text in the making; usually covered with erasures and rewritings".

paragraphes, de chapitres ou de textes entiers" (245).[24] The term 'réécriture' is thus used as an overarching term for variance through revision, instant revision, correction, substitution, overwriting etc. (Dillen 2015, 220). 'Rewritings' show the dynamics of the writing process and, although they may take on a different form in the case of digital writing, they remain the focal point in this study.

Chapter 4, "Reconstructing rewriting", therefore shifts the focus to the concepts of writing and 'rewritings' in digital writing and proposes a method to bring these operations 'back on the screen' to accommodate the analysis of the writing process. The chapter begins with a discussion of the various types of digital revisions, as distinguished in writing studies. This is then compared to the way in which revisions have been described in textual scholarship and genetic criticism. Proceeding from this typology, the next section describes how the keystroke logged writing processes were reconstructed by combining the MS Word document with the keystroke logging data in TEI-conformant XML. The final section provides a closer look at digital revision, with specific attention to the attempt to uncover the possible reasoning behind revisions, as well as behind regular text production.

Chapter 5, "The temporal dimension of writing", provides a closer examination of the way in which the logging of keystrokes makes it possible to study the temporal dimension of writing more effectively within the framework of genetic criticism. First, the metadata of the keystroke logging files enables a close investigation of the chronology of the writing process and certain writing and revision habits of the authors. Second, the possibility of analysing the temporal dimension allows for the analysis of a new aspect of textual genesis, which I have termed the 'nanogenesis' (Bekius 2021a). In a paper entitled "The dual process model of the development of content during writing", given at the Track Changes symposium (2022), David Galbraith asserted that it may be time "to combine the study of process with the study of content". It is precisely this where genetic criticism and cognitive writing process research can complement each other, specifically in the investigation of discovery during writing. By examining the nanogenesis, we may investigate the aspect of discovery in literary writing. Finally, the study of the sources used during writing – exogenesis – also benefits from the detailed temporal dimension. Keystroke logging can generate insight into the use of the internet during writing, as well as into when and how these online sources or analogue sources were (or were not) used. It can even help in determining the exact source when this is not present in the data or the documents.

In Chapter 6, "Nanogenetic narratology", all these elements of the temporal dimension are used in a genetic analysis from a narratological perspective. The combination of genetic criticism with narratological analysis has been advocated by Dirk

---

[24] Translation: "any scriptural operation that goes back to what has already been written, be it words, sentences, paragraphs, chapters or entire texts".

Van Hulle (2022), also together with Lars Bernaerts (Bernaerts and Van Hulle 2013; Bernaerts, Van Hulle, and Martens 2011). For genetic criticism, the narratological insights may help to enhance the understanding concerning how narratives evolve throughout the writing process (Bernaerts and Van Hulle 2013, 285). This chapter examines how genetic narratology could benefit from keystroke logging material, taking the enhanced temporal dimension into consideration. It includes a discussion of the same levels of narrative text (story, narrative, and narration) addressed by Herman and Vervaeck (Herman and Vervaeck 2019) and by Dirk Van Hulle (Van Hulle 2022) to investigate how these aspects arise and change in the course of the writing process for narrative texts. More precisely, the discussion explores how the nanogenesis of and the interaction with the text produced so far can provide clues for understanding the decisions made with respect to these aspects.

Chapter 7, the conclusion, returns to the initial research questions and brings together the findings from the previous chapters to show how keystroke logging opens a new dimension for genetic criticism. In addition, and of equal importance, it argues that keystroke logging – apart from some practical implications (e.g. its reliance on an active contribution on the part of authors) – is one of the ways in which the digital age does not necessarily spell the end of genetic criticism. It is thus not the end of an era – far from it – and, if the right measures are taken, we can celebrate the dawn of a new millennium for genetic criticism.

Throughout this study, my aim is to illustrate how we can work with the abundance of data on the writing process that can be generated by keystroke logging, without losing sight of the human behind the data: the author. As such, it is an exploration of the possibilities of such data for the understanding of literary writing processes. Let us uncover what lies behind the screens.

CHAPTER 2. THE WORD PROCESSOR AS MEDIATOR

DIGITAL LITERARY WRITING, GENETIC CRITICISM AND WRITING STUDIES

Although the quill of yesteryear has largely been replaced by the wafer-thin laptop, the discipline to which all writers submit themselves, as they strive to make the written word sound original and pure, remains unchanged" (Afman 2016, 5)

In *Where the Magic Happens* (2016), 37 of the 45 photographs depict a computer. Of the 55 photographs of authors at their writing desks in Jill Krementz' *The Writer's Desk* (1996), only 10 feature a computer. Matthew Kirschenbaum provides a list of such 'techy-savvy' writers, including the famous photograph of Stephen King with his Wang System 5 word processor, and:

> Amy Tan has moved on from her Kaypro II to a Mac PowerBook,[25] possibly the same model Mona Simpsona is photographed with. Veronica Chambers and Cathleen Schline also both have laptops. Roy Blount Jr. has an IBM-compatible with a great big monitor. Edwidge Danticat has a Macintosh desktop. John Updike leans over a terminal of seemingly indeterminate manufacture, though it too is a Wang product. John Ashbery works at a typewriter, but has a PC alongside. (Kirschenbaum 2016, 75)

Since 1996, the use of a computer in the literary composition process has thus only increased, and the technology has become more and more advanced. Although the computers that were used by the authors in *Where the Magic Happens* are quite different from those depicted in *The Writer's Desk*, they have one thing in common: the texts written on them are born-digital.

---

[25] The photograph actually depicts Amy Tan sitting at an IBM ThinkPad 701, also known as the Butterfly.

The writing environment has thus changed gradually along with the development of word processors since the 1970s. One important point to consider, however, concerns whether the word processor has also changed literary writing (or the experience thereof) and this entails for the theory and methodology of genetic criticism. In this chapter, I set out the foundations of my research. First, I discuss digital writing in more detail, focusing on the word processor. More specifically, I address the advances that have been made in the development of the word processor and the effects that have been outlined from the use of such writing tools. In the second section, I discuss the methodological framework of genetic criticism and models of the writing process as developed in cognitive writing process research. Now that genetic criticism must deal with texts written on the computer, insights from writing studies can help to interpret the writing process – especially when a keystroke logger is used to collect data. This provides a basis for the further exploration of how a word processor can act as a mediator between genetic criticism and writing studies, and particularly, cognitive writing process research.

2.1 DIGITAL LITERARY WRITING

The term 'born-digital' has been applied in a variety of ways with regard to literature. First, it has been employed to refer to literature that is both *written* and *read* on a digital device:

> A born-digital work [...] remains digital at every step of its production, transmission, and consumption: it is one that is composed, edited, and laid out on a digital device; that reaches its readership via the digital medium; and that is designed from the outset to be read and experienced on a screen, not a page. (Hammond 2016, 133)

In this narrow definition, born-digital literature corresponds to what is also referred to as electronic literature or e-lit: literary works that combine "text with the affordances of computers" (Schiller and Grigar 2019, 47). The Electronic Literature Organization (ELO) defines 'electronic literature' as follows: "works with an important literary aspect that takes advantage of the capabilities and contexts provided by the stand-alone or networked computer" (Hayles 2008, 3). This definition emphasises the feature that such texts forsake conventions and medium-specific elements. Born-digital or electronic literature can alter the 'nature' of literary texts by allowing interactivity (e.g. allowing the reader to decide the narrative sequence) or by adding such modalities as sound, video, or animation (Hammond 2016, 133).[26] The computer has made print "a particular form

---

[26] Examples of these aspects can be found in the digital and interactive literature project

of output for digital files rather than a medium separate from digital instantiation" (Hayles 2008, 159; Malvik and Paulson 2016, 1). In recent years, projects have been set up in the Netherlands (the Digital Literature Consortium; Tilburg University) and Belgium (The Laboratory for Electronic Literature (LabEL); UCLouvain and KBR, Royal Library of Belgium) to curate and preserve digital/electronic literature, as well as to explore the possibilities it offers for academic research, public libraries and education.

Despite the possibilities offered by electronic text, most authors still publish in print, and literature continues to be associated with print books (Hammond 2016; Buschenhenke 2016). Most of these printed books are 'born-digital' as well, however, in the sense that they remain digital from the first "keying the text into a computer" up until the moment of publication (Kirschenbaum et al. 2009, 105). Although it is not possible to foresee the future, the cut-off point referred to by Derrida in 1996 persists to this day: "the publishing machine, the market for books, printing, and even libraries – in short, the ancient world – still all play the role of a cut-off point. The book is both the apparatus and the expiration date that makes us have to *cut off* the computer process, put an end to it" (Derrida 2005, 28).

As mentioned by O'Kane Mara, however, few studies have considered "how to perform textual and genetic scholarship on *texts* created in a digital space but intended for and materially re-created as print books" (Mara 2013, 344, emphasis in original). Doug Reside also notes the scarcity of research conducted on born-digital manuscripts of works intended for distribution in print. According to Reside, this also because these manuscripts are still in the possession of their creators, and have thus not yet been donated to public archives (Reside 2014, 70). Nevertheless, Reside considers it a mistake "to limit digital manuscript studies only to works intended to be experienced on a computer" (70).

With this study, I aim to contribute to the existing body of textual genetic research on born-digital works intended for publication (also) in print, but that were created within a digital environment. To this end, I employ the term 'born-digital literature' to

---

PingPong (https://pingpongproject.nl), as developed by Kumi Hiroi (graphic design), Anneke Hymmen (photographer), Basje Boer (author) and van Leeuwen & van Leeuwen (interaction design) as part of the 'Literatuur op het scherm' [Literature on the screen] programme set up by the Dutch Foundation for Literature (Nederlands Letterenfonds). PingPong plays with the interaction between traditional ways of reading and new, alternative ways of navigating through a story. It can be read in 222 different ways, and the readers/users have the choice between reading the stories or listening to them in an audio format with subtitles. Each time, the reader/user can click on one of three pictures, which then leads to the corresponding story. Clicking from one story to the other, the reader can discover how the stories connect. As illustrated in this example, "developments in computer technology facilitate a variety of new creative expressions that keep the parameters of what we 'generally regard as literature' in constant flux" (Malvik and Paulson 2016, 5).

refer to literature that is written on a computer (at least using a word processor), but that does not necessarily grow old exclusively in digital form.

2.1.1 The word processor

As outlined in the previous section, a word processor is a key element in digital writing. In terms of its predecessor, a word processor could be seen as a "sophisticated typewriter", which also functions as Tippex, scissors and paste, as a copier, as a copy-editor (with spelling and style checkers) and as a printshop, with its formatting functions (Chandler 1995, 28). The first word processor, however, differs immensely from what is currently defined as a word processor: "What is very likely the first novel written with a word processor wasn't written on a word processor with a screen and its words weren't 'processed' by the novelist who wrote it" (Kirschenbaum 2016, 166). The question of which novel was the first to be written with a word processor provided the inspiration for Matthew G. Kirschenbaum's *Track Changes: A Literary History of Word Processing* (2016, ix). Although Kirschenbaum notes that the answer to this question is highly dependent on what is meant by word processing, he gives the honour to Len Deighton's *Bomber* (1970), which was written on a machine known as the MT 72 – the European market's name for IBM's Magnetic Tape Selectric Typewriter (MT/ST). It had actually been Deighton's assistant, Ellenor Handley, who had learned how to operate the machine. Kirschenbaum explains the basic principles of the MT/ST as follows:

> at the same instant it was imprinted on the page, each individual keystroke was also recorded as data on a magnetic tape cartridge (each cartridge held approximately 24,000 characters), which could then be played back to have the machine go about the task of automatically printing (and reprinting and reprinting…) a page of text at the rate of some 170 words per minute. Backspacing to correct an error resulted in the usual blemishes on whatever piece of paper was in the Selectric's rollers at the time, but the revised sequence of characters was what got stored on the tape: clean texts could then be produced literally at the push of a button, without the need for time-consuming and imperfect erasures. Sentence spacing, line lengths, even hyphenated words were all automatically adjusted as revisions were introduced, one of the more technically impressive features of the product. (2016, 169)

Although the MT/ST was not referred to as a word processor when it was first introduced, Kirschenbaum regards it as the first word processor. In doing so, he acknowledges that this renders the general understanding of the term 'word processor' less familiar, as it still entails writing on paper: "what is important about word processing turns out to be not the glowing letters behind the glass but a workable mechanism for suspending the act (that is, the moment) of inscription" (2016, 245). In this regard, Kirschenbaum is referring to the common fascination for the screen – the luminous,

glowing "glass square" (44). According to Kirschenbaum, this "aesthetics of luminescence" influenced the way authors thought and talked about their experiences with the word processor (45). Authors were referring to working and writing "with light", which enabled them to write faster and offered freedom and flexibility (45). As Kirschenbaum argues, however, the key feature of word processing – the major shift the new technology brought along – is 'suspended inscription'. Coined by Chandler (Chandler 1992), this term means that:

> the stored record of a text is separate from whatever the medium or surface on which it is ultimately printed or inscribed in more palpable form. When one writes with a pen, creating and composing a text is coterminous with the work of inscribing it; and it is the same with typewriting, the press of a key initiating a simple act of mechanical leverage that sends the type bar hurtling toward the page, its kinetic energy thus impressing the inky fabric of the ribbon it encounters in its path onto the paper behind it in the embossed shake of a letterform […] But word processing is different […] we can see that word processing's suspension of inscription is in fact a suspension both temporal and locative in nature. In other words, there is a gap or delay between the act of writing the text and rendering it in its documentary form; moreover, the record of the text and its documentary instantiation occupy physically distinct media and surfaces. (Kirschenbaum 2016, 47)

The ease offered by such 'suspended inscription' has also been advanced by Jay David Bolter, who notes that word processors "demonstrate the flexibility of electronic writing in allowing writers to copy, compare, and discard text with the touch of a few buttons" (1991, 5). With the emergence of the word processor, the 'manual labour' of writing changed: "The switch from pen and paper to mouse, keyboard and screen entails major differences in the haptics of writing" (Mangen and Velay 2010, 385). The advantages for the author are that it streamlines and simplifies the work of writing; typing on the keyboard is quicker and makes the text more readable. In addition, adding, deleting and moving texts is easier on a computer, and common errors are eliminated by grammar- and spell-checking programs (Haas 1996, 52).[27] For some authors, the speed allowed by

---

[27] WordStar, which debuted in 1979, set the standard for word processing with the marketing strategy WYSIWYG (What You See Is What You Get): "what was on the screen was what would be on the page when you printed it, or such was the claim" (3). This WYSIWYG became "one of the perennial grails for word processing, requiring ongoing innovation not only in software but also in hardware technology" (126). Although WordStar dominated the market in the first half of the 1980s, its position would eventually be taken over by WordPerfect and, later, by Microsoft Word. While Microsoft Word currently remains one of the most widely used word processors, it must now share the market with a wide range of other word processors, including LibreOffice, Apple Pages, GoogleDocs, Ulysses and Scrivener.

typing on a word processor is also "helpful in 'keeping up' with their thoughts" (Chandler 1995, 129). At the same time, however, Bolter provides another explanation for the popularity of word processors amongst authors: they do not interfere with the conventional notion of writing, which results in ink on paper (1991, 5).

From the very onset of word processing as a concept and technology, it has reflected "the quest for flawless efficiency and effectively flawless results" (Kirschenbaum 2016, 34). Using a word processor, it is possible to create a "perfect document" (34). Although the ideal of perfection has never lost its appeal, it is accompanied by the supposed "dematerialization of the written act": no visible sign of its history would be visible on the 'perfect' document, "as though the document did not have a history, but rather emerged, fully formed in its first and final iteration, from the mind of the author" (36).

2.1.2 Does the word processor affect writing?

The word processor thus did not challenge the purpose of writing, and most books continue to be published in print – or as e-books, but this format usually does not challenge the reading conventions of print. Although the output format has not been drastically affected, but one question that has often been raised is whether the text – the writing style – has been changed by writing on the computer. In an interview in 1996, Derrida mentioned that he was often asked this question, but that he did not know which criteria should be taken into account: "There's certainly a change but I'm not sure that it affects what is written, even if it does modify the way of writing" (Derrida 2005, 25). In this respect, in *Writing Technology: Studies on the Materiality of Literacy* (1996), Christina Haas discusses two myths about technology within literacy studies: the 'technology is transparent' myth and the 'technology is all-powerful' myth.

The 'technology is transparent' myth sees "writing as writing as writing" (Haas 1996, 22). The essential nature of writing itself would be unaffected by the way it is produced and presented. This myth is thus positive about the acceptance of computer technology without considering its possible negative effects, believing as it does that writing is not shaped by the use of technology (22). While the 'technology is transparent' myth acknowledges that writing with computers is different, it limits this effect to an increase in efficiency: "writers can compose, revise, edit, and produce texts more quickly and with less effort with computers; therefore, using computers increases writers' efficiency, but makes no profound difference in how writing gets done" (34).

The pitfall of this assumption is that it implies that we do not need to study the effects of technology at all. As Haas argues, however, "believing that technology is transparent does not in fact make it so, and does not preclude technology having powerful effects on literacy, effects that we are not prepared to examine or understand if we are operating with a belief that technology does not matter" (34).

In the 'technology is all powerful' myth, computer technology is regarded as highly influential, with effects that are mainly positive and always inevitable (Haas 1996, 22). Within this myth, "individual practices and motives, as well as cultural habits and beliefs, take a subordinate position to technology, which is seen as determining itself" (22). One consequence of this myth is that it suggests the necessity of re-establishing the theory and practice of literacy: "Existing theories, practices, and rhetorics will be useless in the new age of this new literacy tool, the computer" (35).

According to Haas, the consequences of technologies (e.g. print) should be regarded as "the result of a complex of technological, cultural, and historical factors" (1996, 36). Moreover, technologies are not static. On the contrary, they are shaped by how we think and talk about them, as well as through the ways in which we use them: "technologies continue to evolve, not just because of technological breakthroughs but because their contexts of use, and their users, continue to shape them" (36). What is clear is that one ought to be cautious in making claims about the influence of a particular writing tool on one's writing style. Moreover, as Chandler warns, it is important to avoid adopting a hard deterministic stance (the 'technology is all powerful' myth): "Writers are seldom completely enslaved by their tools. This is not to deny that influences related to the use of writing tools may well be profound, but they are also likely to be extremely subtle and difficult to capture in tidy generalizations" (1995, 141). Cultural practices, individual thoughts and technologies all have an effect on the act of writing (Haas 1996, 27).

The importance of tools can be traced in various ways, including through the statements of literary authors about the rituals they have for selecting and using specific writing tools (Chandler 1995, 132). Chandler refers to this as 'resonance', defined as: "any kind of significance which may be attached to the use of one medium rather than another" (1995, 12). As evidenced in several quotations from *Where the Magic Happens*, authors do indeed have profound preferences for writing tools, all of which are highly personal. For example, Elvis Peeters (1957) continues to write by hand, as he has a need to feel that he is creating literature: "Only when I can physically delete a passage, does a text begin to live for me" (Peeters quoted in Afman 2016, 123). Jan Siebelink's (1938) approach has not changed in years: "I write by hand on the left-hand side of my desk, and type it up on my Triumph-Adler on the right-hand side" (139). And he also uses different colours of pen "for corrections, different voices or comments" (139). Geert Kimpen chooses whether to write by hand or on the computer depending on the complexity of the text to be written: "I write difficult sections with a fountain pen in a notebook but type straightforward passages directly onto my computer" (79). Jan Vantoortelboom (1975) draws a distinction between the inner world and the outside world, which "nourish each other" during the writing process: "This is also reflected in the fact that I work on two computer screens: one for my manuscripts, the other mainly

for research" (175). Annelies Verbeke (1976) has a preference for the position of her screen, and this influences her experience of the act of writing:

> When I write, I place my big screen as far away as possible and increase the font size to large. Not because my eyesight is bad, but because I find it less visually taxing. I'm so attached to the screen now that I miss it when I have to go somewhere else and type on my laptop. Yet for a long time, I didn't work on anything else. It's as though the writing somehow feels less real. (Verbeke quoted in Afman 2016, 183)

The quotations presented above thus reveal "the way in which small, seemingly insignificant details of a particular technology ended up mattering a great deal to a writer, often for inexplicable or inarticulate reasons" (Kirschenbaum 2016, 10). We can see this in the statement that writing is 'less real' when the text is not displayed on a particular screen or that the text becomes alive only when written by hand.

Haas observes that technologies (e.g. screen size and other aspects of the graphical interface) influence the author's 'sense of the text', which is "a mental representation of the structure and meaning of a writer's own text" (1996, 188). Closely related to this are the reading difficulties that Haas identifies as one of the main problems of computer writing. The writers whom Haas interviewed complained about difficulties as they attempted to "access large parts of their writing or move quickly to a specific space in the text", to reorganise or to detect errors in the text. Some even referred to a lack of 'intimacy' (55). Although the results reported by Haas are based on interviews held between 1983 and 1991, the difficulties of reading on a computer screen are still relevant today. For example, the Dutch author Robert Anker (as quoted by Mathijsen) said that he revises on printouts for this reason: "I do corrections on the print-out. You can see better on paper whether something is too soft or swampy. And minor flaws, such as word repetition, you can see those better on paper as well" (Mathijsen 2009, 236). Like most of the writers in Haas' survey, Anker attempts to alleviate his online reading problem by printing out his texts. In a survey of Dutch authors, 75% of the participants considered it important to make printouts, primarily for revising and to read the text (Buschenhenke 2016, 12).

Van Waes and Schellens (2003) identify differences in the level and distribution of revisions and the degree of fragmentation in the writing process, between writing with pen and paper and writing on a computer. At the level of revisions, computer writers revised more at the level of the letter than pen-and-paper writers did. Computer writers had a tendency not to delete the part of the word that did not need to be changed, and they tended to delete words before they were completely typed (Van Waes and Schellens 2003, 848). The experience of writing on the computer was also characterised by the high number of revisions made in the first writing stage, while the units in which computer writers revised their text were smaller. Another difference between computer

writing and writing with pen and paper was the degree of fragmentation: "The large number of short pauses within the sentence which occurred in rapid succession in the writing of computer writers resulted in a writing process in which planning, formulation, and revision were strongly focussed on relatively small units of text" (848). According to Van Waes and Schellens, this might be explained by what they propose to call 'word processing comfort': "During the writing process, writers are aware of the fact that, at any moment during writing, they can alter their text without creating an illegible jumble of crossed out and inserted words" (848). Chandler observes that some writers feel that "the word processor encourages *over-editing*", which can feel like a "loss of spontaneity", while other writers report being 'sloppier' and less critical when writing with a word processor (1995, 150). In some cases, the latter kind of *under-editing* is related to "the printed text looking more 'finished' than it is" (Chandler 1995, 150).[28]

In relation to editing and revision, a word processor "obscures its own evolution", deleting words leaves no visible trace (Chandler 1995, 150). This is the central problem of this study with regard to the perspective of textual scholars. As demonstrated by Chandler, however, it has implications for authors as well: "The handwritten text maps paths not taken in a way that enables them to be re-explored if necessary. For tentative, exploratory writing such a function may matter" (150). Chandler also notes that, while many writers make a printouts or save their files, few would "preserve a complete draft every time a few words were changed" (150). According to Chandler, this is due to several problems, including limited storage space and a desire not to waste paper. More pressingly, to actively save or print documents, an author must be continuously conscious of the act of writing: "Doing so would not be compatible with preserving spontaneity or momentum in writing" (151).

In my opinion, most of these problems no longer seem to be an issue, given the abundance of space available on hard drives or in the 'cloud'. Moreover, with AutoSave, there is no need to worry about saving files. Even now, however, it is unlikely that every change will be documented, or that we will have easy access to all the changes we have made. What is preserved from the writing process is also highly dependent on the

---

[28] Kirschenbaum describes an anecdote about a cover story in *Writer's Digest*, that is worth mentioning in this context. In 1982, the journal published an article entitled "Writing Made Easier with Personal Computers". This cover story would attract a considerable amount of critique through letters to the editor, in which somebody stated: "Typing made easy with personal computers, not writing" (Kirschenbaum 2016, 37). In addition, the 'perfect' document that could be created with a computer often led to suspicion, scepticism and anxiety. The polished appearance of the text could also be deceptive, possibly misleading the author to think that the text was better than it actually was (37). The word-processed document therefore became "a stand-in for a whole set of much deeper anxieties related to authenticity in the writer's craft – originality and creativity, truth and beauty" (40).

preferences and views of the author, in both the choice of writing software and whether they see the value in saving the changes – I return to this aspect in Chapter 3.

I now return to the question of writing *style*, given that some have claimed that they could also pinpoint the effects of a writing tool on the author's personal writing style. In a discussion or analysis of style, the focus is often on "the characteristics or features of texts" (Hoover 2021, 6). Don Ihde, a phenomenological philosopher, asserts that different tools, and their relative speeds of composition, can affect the style of writing (Chandler 1995, 137). Chandler quotes an editor who stated that "texts produced on word processors often have a recognizable style – tendency to repetition, over lengthiness, failure to establish logical steps between 'blocks' of argument" (147). Authors have reflected on this as well, one noting that he was able to see how the writing tools he used altered the way he wrote:

> As for a word processor, because all its erasures are instantly eradicated, it grants the writer easy access to a syntactical complexity that he would probably, unconsciously, recoil from if it were to result in the ugly scorings-out of a handwritten page or the equally ugly xxxxxxxs, those little kisses-to-make-it-better, on a typewritten page. (Chandler 1995, 147)

In addition, the essayist Anne Fadiman said she could detect the traces of word processors in books as well, which tended to be too long and wordy: "The writers – no longer slowed by having to change their typewriter ribbons, fill their fountain pens, or sharpen their quills – tend to be prolix" (Fadiman 1998 quoted in Kirschenbaum 2016, 38). Sceptical about the use of a computer in the writing process, the Dutch author Willem Brakman stated that he would not like to be able to type faster: "als je langzaam typt, dan heb je misschien weer kans op een inval" (Brakman in Hoffmann 2003).[29] When asked if he thought it was noticeable that a book has been written on a computer, he answered that it would be hard to answer, but that he sincerely believed that this would indeed be the case: "Ik kan daar niet direct op antwoorden, maar ik zou d'r naar loeren of daar geen verschil was. Ja, ik eh, ik acht de activiteit van het schrijven zo hoog, dat het mij zou spijten als er geen verschil te merken was. We moeten zoeken naar fouten" (Brakman in Hoffmann 2003).[30] He even opined that this should be mentioned on the cover of the book: 'made on a computer'.

It would be difficult to verify these claims scientifically, even (according to Kirschenbaum) with computational methods: "Computational textual analysis can show

---

[29] Translation: "if you type slowly, then you might have a change of getting new inspiration". See: https://www.youtube.com/watch?v=SIJfmffun4c&list=WL&index=1&t=1s

[30] Translation: "I can't answer that directly, but I would peruse it to see whether there was any difference. Yes, I uhm, I hold the activity of writing in such high esteem that I would be disappointed if there were not to be any difference. We must look for mistakes'"

us much, but it would take a lot of convincing for me to believe that [George R. R.] Martin's sentence structures (for example) are tied in any significant degree to the specifics of WordStar's keyboard commands" (2016, 6). Nonetheless, this is exactly what David Hoover tried to do in *Modes of Composition and the Durability of Style in Literature* (2021). In this way, computational methods are used in the attempt to pinpoint changes in style caused by a change in writing tool. Whereas an author's 'sense of text' is subjective, Hoover argues that a change in style caused by a change in composition would be measurable "with a fair amount of objectivity and confidence" (12). In such stylometric research, style is defined as "a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively" (Herrmann, van Dalen-Oskam, and Schöch 2015, 16). To answer questions concerning the effect of the word processor on style, Hoover performed computational stylistic analyses on the works of four authors: Arthur Clarke, Octavia Butler, Stanley Elkin, and Ian McEwan. Clarke and Butler moved to word processing from an electric and a manual typewriter, respectively, and Elkin and McEwan both shifted from handwriting to word processing. For Clarke, Butler and Elkin, the change in the mode of composition did not significantly affect their style. For McEwan, however, Hoover identified differences between the handwritten novels and the novels written with a word processor, although these differences seem more semantic than stylistic:

> study of the characteristic vocabulary of the novels in the two modes casts additional doubt on a mode effect. It is difficult to imagine that changing from handwriting to word processing would cause a decrease in past tense verbs or an increase in abstract nouns. (Hoover 2021, 189).

In combination with the results of other shifts in modes of composition (e.g. from handwriting to dictation, from handwriting to typing), this led Hoover to the following conclusion: "My conclusion is not that nothing happens when authors change how they write their texts but rather that nothing *dramatic* happens and that literary style is surprisingly durable in the face of such changes" (2021, 191). Louise Doughty still remembers the discussion about the effects of the word processor on style, which, according to her do not seem to hold today:

> I remember the arguments about whether using a 'word processor' was good for a writer's prose style. When they were first launched, there was serious discussion about whether, because it made writing easier and quicker, it would lead to 'baggy prose'. Such arguments seem silly now, but I do remember that at the time I wrote first drafts longhand and thought I really needed the physical connection to the paper. Now, I wouldn't dream of it. (Doughtly quoted in Johncock 2012a)

The shift to a word processor as a writing tool thus may not have a dramatic effect on the literary style, but what effect does it have on the study of the writing process in which this literary style emerged?

2.2 INVESTIGATING THE DIGITAL LITERARY WRITING PROCESS

Suspended inscription makes authors experience writing as flexible, fluid and fast, and it enables the creation of a 'perfect' document. At the same time, however, it has implications for scholars wishing to study precisely those actions made between the first keystroke on the keyboard and the 'perfect' final document in order to gain a better understanding of the text's genesis. This description of the implications of suspended inscription reflects two perspectives on writing: writing as process (writing towards the 'perfect' document) and writing as product (the 'perfect' document itself). Writing as process and writing as product are studied in various domains. Psychology, textual genetics and anthropology are mainly interested in writing as process, while discourse analysis, text analysis and corpus linguistics focus mainly on the product (Cislaru 2015, 1). Genetic criticism already proceeds from the assumption that the concepts and methods of the study of process and product are porous (2015, 2). I am nevertheless convinced that an interdisciplinary approach – even amongst the various traditions in writing-process research – could provide new insights into both the process and the product of writing. This is especially likely now that pen and paper have largely been replaced by the computer (Fenoglio 2015, 131).

Suspended inscription also contributes to the problem of the metaphorical first, outer black box within the *double black box* – a term used by Grésillon and Perrin to describe the two problems that researchers encounter when investigating writing (2015, 36). The first, outer black box is used to describe the problem "that written language is mostly presented as a finalized product, detached from all traces of genesis such as insertions and deletions" (Grésillon and Perrin 2015, 36). The second, inner black box presents itself the moment researchers manage to retrace the process behind the writing: only the material activity, the result of physical behaviour, is captured (36).

The first black box largely "stopped researchers [within the field of linguistics] for a long time from investigating writing processes in natural contexts" (Grésillon and Perrin 2015, 37). Grésillon and Perrin describe how, in the introduction to their collection of early approaches to writing processes, the linguists 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 quoted in Grésillon and Perrin 2015, 37). What they had in mind were the analyses of literary processes as practiced

within genetic criticism (37), which regards the 'definitive' text of a published work as "the result of a process, that is, a progressive transformation, an investment of time that the author has devoted to researching documents, writing, correcting and recorrecting, etc." (De Biasi 2004, 37). The literary manuscript, which bears the traces of this process of invention offers the opportunity to bypass the first black box, as it sheds light on the "backstage process of writing" (Grésillon and Perrin 2015, 36).[31]

The development of keystroke logging made it possible to analyse the geneses of non-literary texts, as it facilitates the study of the dynamics of digital writing (Grésillon and Perrin 2015, 36). With the electronic recording of all operations (e.g. keystrokes, deletions, cursor movements) made by the writer during the writing process, keystroke logging offers insight into a large number of features of the writing activity, like the fluency of writing or the sequence of actions during writing (Spelman Miller and Sullivan 2006; for an overview of research involving keystroke logging, see Sullivan and Lindgren 2006; Lindgren and Sullivan 2019). Genetic criticism thus offered an early methodology for understanding processes of text production (Wengelin et al. 2019). This inspired linguistic research on the writing processes underlying non-literary texts, which itself was made possible by the development of observational tools, such as keystroke logging (Grésillon and Perrin 2015). In turn, keystroke logging now creates the opportunity for genetic criticism to study all the writing operations in the digital literary manuscript, thereby circumventing the implications of the suspended inscription of word processors. Furthermore, recordings of the digital composition process can now help to comprehend the phenomenon of text production (Leblay and Caporossi 2014, 9). Although several joint contributions between cognitive psychology and genetic criticism have been published over the years (Alamargot and Lebrave 2010; Plane, Alamargot, and Lebrave 2010; Leblay 2011; Leblay and Caporossi 2014; 2014; Cislaru 2015), the literature contains only a few explorations of keystroke logging as a means of collecting data on literary writing processes. For example, Victoria Johansson (2022) studies the keystroke logging data of two professional poets to investigate the planning and revision processes involved in the compositional process, and Anne-Marie Butzek is conducting a doctoral project at Aix Marseille Université to study literary writing from a cognitive perspective, using keystroke logging data. To investigate what the rapprochement

---

[31] Throughout the 19th and 20th centuries, literary drafts were carefully preserved by the authors themselves (De Biasi and Wassenaar 1996, 28). With its emphasis on the past, individualism, the 'genius' and his mind, the Romantic Period provided an impetus for the change in authors' attitudes towards the status of their rough drafts: "these documents were being regarded as traces of a brilliant mind or 'genius' at work" (Van Hulle 2014, 7). In addition, once paper became more affordable, authors could revise their works more often (Sullivan 2013, 23). As a result, there are sizeable manuscript collections that represent the literary production of the 19th and 20th centuries: the material traces of literary invention.

between genetic criticism and the cognitive writing process research might entail, an overview of the key aspects of both disciplines is presented below.

2.2.1 Genetic criticism

Genetic criticism was established in the mid-1960s, and its history relates to the acquisitions of the Bibliothèque Nationale de France (BnF). After the national library acquired the manuscripts of Heinrich Heine, Louis Hay established a research team for the analysis of the material, funded by the French National Centre for Scientific Research (CNRS). Several research groups that were working on authors' manuscripts joined forces in 1974 to create the Centre d'Analyse des Manuscrits Modernes (CAM), which became the Institut de Textes et Manuscrits Modernes (ITEM) in 1982.

The aim of genetic criticism, or *critique génétique*, is to investigate the genesis of texts in order "to construct a series of hypotheses on the operation of writing" (Grésillon 1997, 106). In doing so, "textual genetic research takes place on a sliding scale of manuscript analysis, critical interpretation, and scholarly editing" (Bleeker 2017, 22). In "Toward a Science of Literature: Manuscript Analysis" (2004), Pierre-Marc de Biasi describes the two objectives of genetic criticism. The first consists of the constitution of readable and analysable manuscripts. The second consists of the critical dimension: the reconstruction of the logic of the genesis. This cannot succeed without the application of a selective critical procedure that reconstructs the genesis from a specific point of view – for example, from the perspective of psychoanalysis, socio-criticism or narratology (De Biasi 2004, 42). To achieve these two objectives, De Biasi further distinguishes five phases of research, which are: 1) the constitution and organisation of the whole genetic dossier; 2) the organisation and arrangement of the dossier of rough drafts and other draft documents; 3) the specification and classification of each document page in the dossier; 4) the preparation (e.g. transcription) of the material to serve the genetic analysis and 5) the establishment and ultimate publication of an 'avant-texte' (2004, 44).

The first four stages contribute to the first objective of the constituting a readable and analysable manuscript, and the final stage contributes to the second objective of reconstructing the genesis. Therefore, these two objectives seem to roughly correspond with genetic editing (objective one) and genetic criticism (objective two) (Bleeker 2017, 26). According to De Biasi, however, "the essential objective of textual genetics is to produce genetic editions", and both objectives thus belong to genetic editing (De Biasi 2004, 61; see also Bleeker 2017, 26). Nevertheless, one important function of this genetic edition is to support the critical analysis of the manuscript. To this end, Peter Shillingsburg and Dirk Van Hulle distinguish between genetic editing and genetic criticism: "Whereas in scholarly editing, manuscript analysis is often seen as a means to an end, that is, a tool to make an edition, genetic criticism reverses these roles and sees

the making of an edition as a tool to facilitate manuscript analysis" (2015, 37). Given that genetic editing and genetic criticism both focus on the writing process and use the same methods, however, the line between them is largely superficial. One strategy for overcoming this division is the concept of the 'orientations to text' as proposed by Shillingsburg and Van Hulle, as its definition "is not necessarily geared towards the production of an edited text, but conceived as a way of identifying how we read the materials in order to understand the work" (2015, 37).

In "Orientations to Text, Revisited" (2015), Shillingsburg and Van Hulle distinguish six different orientations to text that a scholarly editor might adopt: the material orientation, the causal orientation, the genetic orientation, the performance orientation and the aesthetic/commercial orientation.[32] These orientations shape the ways in which works are identified and understood, and they subsequently guide the approach to the textual material: "differentiating ways to edit, and determining how the narrative of composition, revision, and publication is framed" (Van Hulle and Shillingsburg 2015, 28). In line with genetic criticism, the primary focus of the genetic orientation is creative invention and the trajectories of the creative development. The main aim is therefore to "produce a narrative of genetic development" and it may involve a form of scholarly editing, but it does not have to (37). If it does, "it implies an editorial strategy that displays creativity in motion rather than settling on a final version as the main object of editing" (37). Within this study, I adopt the genetic orientation to text to investigate the textual development present in the keystroke logging data. In addition, I employ an editorial strategy to prepare the keystroke logging material to facilitate genetic analysis (see Chapter 4).

Within the genetic analysis, the study of the writing process is commonly divided into three levels: the examination of the exogenesis, the endogenesis (Debray-Genette 1977) and the epigenesis (Van Hulle 2014). In turn, these levels can be studied through a microgenetic analysis or a macrogenetic analysis (De Biasi and Wassenaar 1996). Exogenesis encompasses the interaction between the text and 'external' sources. The creative process of literary writing is seldom an exclusively internal process. More often, it is an "interaction between books and notebooks" (Van Hulle 2016, 37). Raymonde Debray-Genette coined the term exogenesis "to denote external source texts relating to a creative process" (Van Hulle 2014, 14). De Biasi borrowed these terms and gave them a more general value: exogenesis "designates any writing process devoted to research, selection, and incorporation, focused on information stemming from a source exterior to the writing" (De Biasi and Wassenaar 1996, 44). Exogenetic material can therefore

---

[32] In this article, Shillingsburg and Van Hulle offer "a re-orientation to orientations to text" as presented in Shillingsburg's *Scholarly Editing in the Computer Age* (3rd ed., 1996). See also Dillen 2015.

include reading notes, reports, newspaper cuttings, comments on things seen or heard, marginalia and bibliographical references, amongst other things (44).

In contrast to exogenesis, endogenesis takes place within the actual composition of texts. De Biasi defines endogenesis as "the process by which the writer conceives of, elaborates, and transfigures pre-textual material, without recourse to outside documents or information, through simple reformulation or internal transformation of previous pre-textual data" (De Biasi and Wassenaar 1996, 43). However, endogenesis need not be purely textual. For example, a drawing can be endogenetic if it does not depict an external source, but rather a fictive entity produced by the writing (43). As noted by Debray Genette and De Biasi, however, the distinction between endogenesis and exogenesis is blurred within the text itself (De Biasi and Wassenaar 1996, 47). To be able to study the 'après-text' as well, Van Hulle coined the term 'epigenesis' to be used alongside exogenesis and endogenesis. Epigenesis describes the continuation of the genesis after a text has been published (Van Hulle 2014, 14).

The exogenesis, endogenesis and epigenesis of a text can be studied from a microgenetic and a macrogenetic perspective. Microgenesis is the analysis of a restricted part of the text: "the total compositional development of a short textual fragment" (De Biasi and Wassenaar 1996, 27). According to Van Hulle, it includes all intra-textual processes: "the processing of a particular exogenetic source text; the revision history of one specific textual instance across endogenetic and/or epigenetic versions; the 'réécritures' or revisions within one single version" (Van Hulle 2016, 50). In turn, macrogenesis analyses "large-scale phenomena" (De Biasi and Wassenaar 1996, 27). It thus embodies "the genesis of the work in its entirety across multiple versions" (Van Hulle 2016, 50). In Chapter 5, I argue that keystroke logging makes it possible to analyse the genesis with an unprecedented granularity, known as nanogenesis (Bekius 2021a; see also Van Hulle 2021 and 2022). To fully comprehend this aspect of nanogenesis, however, it is necessary to take the insights of the cognitive writing process research into consideration.

2.2.2 Cognitive writing process research

The cognitive approach to writing has focused on the process in writing since the late 1970s. It is concerned with "what the writer does (planning, revising and the like) instead of [on] what the final product looks like (patterns of organization, spelling, grammar)" (Applebee in Spelman Miller and Sullivan 2006, 2). In writing research, the cognitive dimension is based on the interests and empirical methods of cognitive psychology (MacArthur and Graham 2017, 24). Research on cognitive processes involved in writing began in the late 1970s with the work of John R. Hayes and Linda Flower (1980), who applied methods of cognitive psychology to the study of expertise in writing (MacArthur and Graham 2017, 24). This research led to a model of writing as a problem-solving

process (Flower and Hayes 1980), which continues to be influential in writing research. Several models of the writing process have since been developed (e.g. Bereiter and Scardamalia 1987; Hayes 1996; Kellogg 1996; Zimmerman and Risemberg 1997; Hayes 2012), and the initial Hayes and Flower model (Hayes and Flower 1980) has been revised several times by Hayes (Hayes 1996; 2012; Leijten et al. 2014).
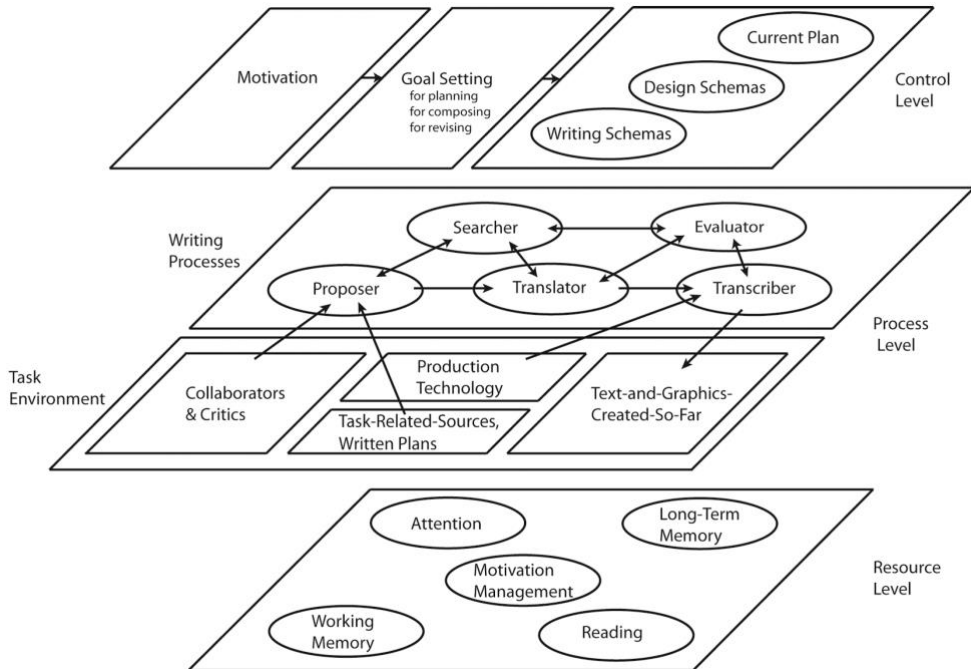
Figure 2.1: Model of the writing process, adapted from Hayes 2012 and modified to include activities of skilled professional communicators (taken from Leijten et al. 2014, 324).

The original model developed by Hayes and Flower (1980) consists of three components: the task environment, the long-term memory, and the writing process. The task environment includes "everything outside the writer's skin that influences the performance of the task" (Hayes and Flower 1980, 12). These external factors include the writing assignment (the intended audience and the topic), the motivation and the text produced so far. The latter is especially important, "because the writer refers to it repeatedly during the process of composition" (12). The text produced so far can refresh memory of the text that has already been produced and prompt new text production as well as revision (Wengelin, Leijten, and Van Waes 2010, 736–37). As pointed out by Wengelin and colleagues (2010), the text produced so far can also be used as "visual external storage", to "decrease the cognitive load of the writer" (737).

The writer's long-term memory encompasses the knowledge that the writer has concerning the general plan, the intended audience, or schemas for accomplishing writing tasks (MacArthur and Graham 2017, 26). The writing process is divided into

three major processes: planning, translating and reviewing. Planning is divided into the subprocesses of generating, organising and goal setting. The translating process involves turning ideas into text. Reading and editing the text are subprocesses of reviewing, but reviewing and editing are regarded as two distinct modes of behaviour. Editing "is triggered automatically and may occur in brief episodes interrupting other processes", while reviewing "is not a spur-of-the-moment activity but rather one in which the writer decides to devote a period of time to systematic examination and improvement of the text" (Hayes and Flower 1980, 18). A writer can control these cognitive processes, as they can be applied flexibly: "any subprocess can interrupt or incorporate any other subprocess during writing" (MacArthur and Graham 2017, 26).

Hayes' (2012) most recent model accounts for the "processes that individual writers engage in as they plan, compose, and evaluate their texts" (Leijten et al. 2014, 322). Although the model focuses mainly on the cognitive processes, it considers aspects of social processes as well. The model has been elaborated by Leijten, Van Waes, Schriver and Hayes (2014) to include the phenomena observed in studying a skilled professional communicator (322). The elaborated model comprises three levels: control, process, and resource.

The control level includes motivation, goal setting, the current plan and writing schemas. Motivation is seen as "essential to sustain the writing activity" (Leijten et al. 2014, 322). Writing activities are determined during goal setting, including the creation of planning, the composition of the formal text and the revision of the text. The current plan comprises a set of goals for the writing of the current text, and the writing schemas consist of the writer's ideas on how the writing processes and resources should be used to obtain the planned text (322). Leijten et al. elaborate this control level with design schemas to account for the creation of visual texts, graphics, drawings, and photographs by professional writers (324). These additions seem less applicable to literary authors, as they mainly produce textual output. Some authors nevertheless do include various kind of visual material in their literary writings. One extraordinary example in this respect is the Norwegian author Johan Harstad. He decided to 'reproduce' the paintings created by one of the main characters Mischa in his novel *Max, Mischa & Tetoffensiven* (2015), as well as to make the posters for the exhibitions, plays and films mentioned in the novel. He also created a document including photographs of places and houses where the characters went, lived or had lived (Koenders and Stevens 2017). This document even includes a floor plan of an apartment in which the characters had lived. This material served to support the writing process.

At this point, I briefly jump ahead to one of the writing processes discussed in this study – that of Gie Bogaert – as it provides an excellent example of how literary authors can make visual designs to support them in their writing processes. Bogaert had already considered the ultimate layout of his novel at the start of his writing process. Inspired
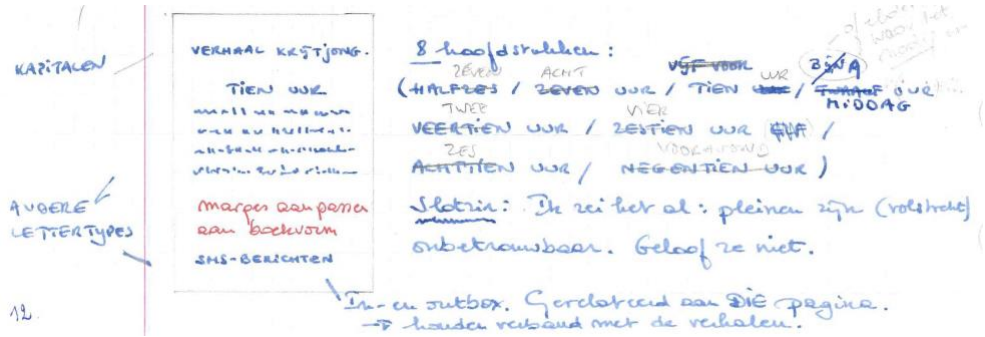
Figure 2.2: Bogaert's sketch in the Atoma notebook for the layout of Roosevelt.

by Louis Paul Boon's *Menuet* (1948), he planned to add a sentence in the top margin of each page, which together would form a different story. He already planned this advance in the notebook in which he sketched the general layout (Figure 2.2).

The process level consists of two parts: the writing process and the task environment. All "the internal mental processes that the writer uses to compose" belong to the writing process: a 'proposer' generates (non-verbal) ideas, which are transformed into language by a 'translator'; a 'transcriber' creates written text out of this, and an 'evaluator' reviews the output of the other three processes (Leijten et al. 2014, 323). These processes interact with physical, social and cultural contexts, which constitute the task environment. This includes "critics and collaborators who may provide suggestions and criticisms for the text", as well as the writer's culture (norms, ways of knowing) and physical source materials (books, articles, written plans or outlines). It accounts for the transcribing technology (e.g. pen or keyboard) as well (323). Another important part of the task environment is the text produced so far, which is consulted frequently during writing. Leijten et al. expand this aspect to include text and graphics created so far (325). They also noted the absence of an account "for the search process during content development" (325). Since most writing is an interplay between the long-term memory (LTM) and external sources of information, Leijten et al. add a 'searcher' to search for information in external sources: "the searcher operates for any type of writing, whether academic, creative, or professional" (325).

The resource level "includes resources that are used in writing but are also widely available for carrying out other activities", and these resources "include long-term memory, working memory, reading and the ability to focus attention" (Leijten et al. 2014, 323). At the resource level, Leijten et al. add 'motivation management', given that writing projects often have a long time frame (325). Downtime – activities that cannot be related to the primary task – is considered as well, as it is at least partly a "result of professionals' meta-knowledge of their own motivational limits"; it should thus not be regarded as 'noise' (331). Voluntary downtime accounts for more productivity at a later stage. For example, it could promote "incubation processes that might yield new ideas

and improved quality" (332). A writer can take pauses to reduce fatigue and boredom or to improve the ability to concentrate and stay on task. Upon returning to the task, the writer may do so with new energy or a fresh outlook (331). Non-voluntary downtime is caused by unexpected visitors, mail and technological issues. Both voluntary and non-voluntary downtime may involve restarting costs, although they could also have a positive effect on the productivity.

The model of the writing process discussed above may also help to describe the literary writing process. In research on writing processes from a cognitive psychological perspective, little attention has been paid to the writing process of literary authors, despite the existence of differences between literary and technical writers (Alamargot and Lebrave 2010). First, at the rhetorical level, literary authors are not restricted to any clearly-defined, circumscribed framework – with the possible exception of the literary genre, which is nevertheless subject to more flexible rules that can even be ignored (Alamargot and Lebrave 2010, 13). Second, literary authors can define the aim of the text and the resources used as the work unfolds. At least in the initial phases of the writing process, this leaves the problem space relatively open (13). Third, in pragmatic terms, the communicative aim of a literary author is more diffuse, as there is no well-defined intended reader. The author's choices are more likely to be motivated by the intrinsic quality of the form and content of the text, as well as by the ideal aesthetic form. The constraints on literary authors are therefore primarily stylistic, semantic, linguistic and aesthetic (13). Finally, in temporal terms, literary authors do not need to write within a given scope of time, and they generally take on lengthy writing projects: 'This difference in temporality (spending just a few hours or days on a document, as opposed to several months) inevitably induces highly contrasting methods of overall task management' (13). As argued by Alamargot and Lebrave, the study of creativity, time management and memory in creative writing processes would enhance the study of writing processes in professional writing.

2.2.3 Keystroke logging

"That bloody cursor blinking at me on the word processor screen is awful. I mean, it's blink, blink, blink – well, screw this bastard, it's telling me to get on!" exclaims the author Tom Sharp (Hammond 1984, 213; quoted in Chandler 1995, 31). In the same way, the Dutch poet Pieter Boskma experiences the cursor as extremely pressing and demanding of speed:

> Het beeldscherm zuigt te veel. Die knipperende cursor zegt voortdurend: 'Kom maar, kom maar, kom maar.' Met de hand schrijven is veel organischer, fysieker. Op papier werken geeft zo'n rust. Het valt me op dat er nu minder

gedichten afvallen. Ik werk wat voorzichtiger, trager, ik proef de woorden eerst goed voordat ik ze opschrijf. (Boskma in Hasselt and Sistermans 2021, 31)[33]

The demanding blinking cursor seems to have replaced the infamous blank page to describe one of the difficulties of writing: finding the 'right' words to communicate what is on our minds. The blinking cursor – or rather the pause it indicates – can indeed have something to say about the writing process: "Pauses are, for instance, widely assumed to be associated with either monitoring processes, where the accuracy and appropriateness of the emerging text is evaluated, or with planning processes, where the content and form of future text segments are reflected upon" (Wengelin et al. 2019, 31). Therefore, the "main rationale behind keystroke logging is that writing fluency and flow reveal traces of the underlying cognitive processes" (Leijten and Van Waes 2013, 3).

As stated in the introduction, the keystroke logging data for my study – and for the Track Changes project in general – were collected using the keystroke logger Inputlog (Leijten and Van Waes 2013). Other keystroke loggers include Scriptlog (Wengelin et al. 2009) and GGXLog (Usoof, Leblay, and Caporossi 2020). Developed at Lund University, Scriptlog is more customised to research on experimental writing processes. Although the actual editor is very limited, it does contain a frame in which pictures can be shown to elicit certain writing action. The GenoGraphiX-Log 2.0 (abbreviated GGXLog), which was developed in a collaboration between HEC Montréal, University of Turku, ITEM (L'Institut des textes et manuscrits modernes) and GERAD (Groupe d'études et de recherche en analyse des décisions), has a primary focus on graph theory for the visualisation of the writing process. GGXLog also works with a built-in text editor, which has the advantage of offering more precise data, especially with regard to the location of a modification in the text. It nevertheless raises one prime concern: it could force the writer to write in an 'unnatural' environment, without the 'look and feel' and other functions of the word processor with which they are familiar. Inputlog works with Microsoft Word, a common, if not the most popular word processor. This allows authors to write in a familiar writing environment, thus hopefully reducing the feeling of being under surveillance while writing.

Although Inputlog and other keystroke logging programs (e.g. Scriptlog and GGXLog) have been developed specifically to observe writing processes for research purposes, keystroke logging is most notorious for being a form of malware. For example, keystroke loggers on public computers can be used to steal passwords or credit card information (Creutzburg 2017, 139). Keystroke logging has also been applied as a method for monitoring employees without their knowledge (Swaya and Eisenstein

---

[33] Translation: "The screen sucks too much. That blinking cursor is constantly saying, 'Come on, come on, come on.' Writing by hand is much more organic, more physical. Working on paper is so peaceful. I notice that fewer poems get rejected now. I work a little more careful, at a slower pace, tasting the words carefully before writing them down.

2005, 9). In contrast, keystroke loggers developed for research have been used to describe writing processes and to understand writing from a sociocultural and cognitive perspective (Lindgren, Knopse, and Sullivan 2019, 1). The further development of keystroke logging tools has also made it possible to combine it with eye-tracking, speech recognition and other technologies. In addition, keystroke logging technologies are also increasingly being used in educational settings to encourage students to reflect on and discuss writing. Further uses of keystroke logging tools include research on the effects of writing development and education; writing in first, second, and foreign languages; writing difficulties and ageing; writing in the workplace and translation. (For an extensive overview of the range of themes in research using keystroke logging software, see Lindgren and Sullivan 2019; Van Waes and Leijten 2020). Usoof and Lindgren (2008) have even investigated the possibility of distinguishing unique 'fingerprints' based on the fluency of transitions and found individual micro-level writing patterns.

One objection to genetic criticism concerns the privacy of the author: "like biography, its nosing around in discarded textual matter is an invasion of places that are essentially private and were never intended to be looked at" (Fordham 2010, 20). I acknowledge that this seems an even more pressing issue for the recording of every keystroke and every window opened during a writing process. Even more so, in contrast to most genetic research, my approach can also be categorised alongside studies researching the archives of living authors – examples of which include the dissertations by Andrea Davidson (studying the archive of Aidan Chambers) and Vincent Neyt (studying the archival material of Stephen King) – which must also take into account the author's preferences regarding the use of archival material. The method of logging the writing process within the 'natural' writing environment is undeniably subject to privacy concerns. In a theoretical exploration of what the digital draft will mean for text genetic research, Serge Linkès briefly refers to software that would be able to save all the changes in the text. In the same sentences, however, he questions the value of writing that is placed under surveillance and whether it would lead to self-censorship (2017). I will therefore briefly reflect on the aspect of writing 'under surveillance'.

It is important to note that, in contrast to when keystroke logging is used as malware, the keystroke logging system used in this research allowed the author to have complete control over the registration of the writing process. At the moment the data were collected (i.e., the moment of writing), the authors are anonymous and logged their stories under pseudonyms. The authors were also completely in charge of their own writing processes, as they could switch the software on and off when preferred. They could therefore decide which parts of the writing process would be logged – although they apparently did not 'censor' the data in any consequential way, except for not logging some of their online 'downtime' activities. They could end their participation at any time and could decide afterwards whether they did or did not wish to share the data.

The data were stored locally on the authors' devices, and they would have been in the position to manage their data themselves before submitting it to the researchers in the project. Inputlog also offers the functionality of filtering the data to show only the keystrokes made within the Word document. The entire Track Changes project was approved by the University of Amsterdam ethics committee.

Despite the aforementioned precautionary measures, this project did involve collecting data from the participating authors (including some very personal data). There was thus also a great deal of trust between the first author to log his process (Gie Bogaert) and the person who collected the data (Luuk Van Waes) before, during and after the data collection. The same applies to my colleague Floor Buschenhenke, who collected keystroke logging data from participating authors (some of whom she knew personally). She was also completely open about which data Inputlog would collect. The authors were well informed before logging their processes, and they signed contracts confirming their willingness to participate in the project. All participating authors agreed for their data to be analysed within the Track Changes project and published in this thesis. They also had the opportunity to read the results.

Although the authors would certainly have been available to answer any questions, I deliberately chose to minimise contact with the authors regarding questions about their writing processes as I conducted my analyses. One reason had to do with the 'reliability' of their responses to questions concerning details of their writing processes, as it is quite possible that they would not remember exactly why they had made a certain revision.[34] This issue relates to the reflection of Andrea Davidson, who studied the construction of age in the genesis of the *Dance Sequence* series by the living author Aidan Chambers. Davidson questions whether

> the octogenarian Aidan Chambers whom I met is a reliable source of information about projects he worked on nearly half a century ago. It seems obvious to conclude that he is not, that any person's memory is a faulty source of factual information, and so to mean to disregard information from him, but I found that Chambers' reflections on his lifelong literary work were so thought-provoking that they nevertheless influenced my thinking. (Davidson 2022, 65)

Like Davidson, I could not always ignore some of the additional information provided by the authors. I was occasionally able to present some of the results of my analyses to

---

[34] One possible way to overcome this problem would be to use of think-aloud protocols as a methodology to observe writing. However, this methodology would be too intrusive for the literary writing process, which we sought to log in a natural setting. The think-aloud method might thus have disrupted the behaviour that we aimed to study. Moreover, the method has also been criticised in terms of its reliability and validity (see Spelman Miller and Sullivan 2006, 4).

the authors. Throughout this study, therefore, I sometimes refer to information provided to me by authors after my initial analyses. The decision to include this additional information was made because I felt it was important to the proper understanding of the writing process. These instances are clearly marked.

The second reason why I did not actively collect additional information from the authors is that my study is concerned with the implications of keystroke logging for genetic criticism in the future (e.g. as a method of preserving the writing process). I therefore sought to examine hypotheses that could be drawn from keystroke logging material without recourse to information from the authors themselves – as if I were to encounter this material in the archive many years from now.

To clarify exactly what is logged with keystroke logging, I again briefly address how the data are stored and which options Inputlog offers for the analysis of the data. The 'raw' keystroke logging data are given in an idfx-file, structured in XML format. These files can be used to generate a variety of analyses: general analysis, summary analysis, pause analysis, process graph, fluency analysis, linear analysis, revision matrix and S-notation. Each of these analyses offers a different perspective on the writing process.

As stated before, the most comprehensive view of the writing sessions is found in the *general* output format (i.e., General Analysis). In Table 1.1 (in the Introduction), each row represents one log event. The first column shows the number of the event (consecutively). The second column shows the 'event type', which encompasses 'keyboard', 'mouse', 'speech', 'focus', 'insert' and 'replacement'. The column next to the event shows its 'output'. For a keyboard event, this represents the letter that has been typed. The position in the fourth column represents the 'cursor position'. The fifth column shows the 'length of the document' in characters. This differs from the character production represented in the sixth column, which displays all characters produced during all writing sessions so far. The Start Time and Start Clock (Columns 7 and 8) show the time of the 'key in' – in milliseconds and clock time, respectively – and the End Time and End Clock (Columns 9 and 10) of each 'key up'. The actionTime (Column 11) represents the time between each key in and key up; the pauseTime (Column 12) represents the time between two key ins. The location of the pause is shown in Pause Location (Column 13) (Leijten and Van Waes 2013). Overall, the General Analysis thus provides information about what was written where and when, and it therefore provides all the details needed for a fine-grained reconstruction of the writing process.[35]

This fine-grained level of event recording has both advantages and disadvantages: "On one hand it allows for very detailed analyses, on the other hand the huge amount of data is sometimes hard to interpret" (Leijten and Van Waes 2013, 7). The most useful

---

[35] This description of the General Analysis in Inputlog has also been published in Bekius 2021a.

analyses for textual genetic research are the Summary Analysis, the Revision Matrix, the S-notation, and the Source Analysis.

The Summary Analysis can be used to study the general development of the text. It comprises five main sections. The first is 'process information', which contains data on the number of keystrokes, words, sentences and paragraphs produced during the writing session. The second section contains 'product information', which provides details on the number of characters, words, sentences and paragraphs in the final text. The relationship between product and process information (e.g. product/process ratio) is presented in the third section ('product/process'). The product/process ratio is "the total number of characters in the final text [divided by the] total number of characters produced during the writing process: a ratio of 1 means that no revisions took place" (Leijten and Van Waes 2013, 7). The fourth section provides information related to writing and pausing times, and the fifth section contains data on writing modes (e.g. keyboard, mouse).

Another useful analysis for genetic criticism is the Revision Matrix: a linear representation of all insertions and deletions that have occurred in the text (Leijten and Van Waes 2013). Each revision is numbered sequentially, and it identifies three types of text production: normal production (new text produced at the end of the text produced so far), deletions (text deleted from the text produced so far) and insertions (text inserted within the text produced so far). The normal-production type is somewhat problematic for long-term literary writing processes. In longer writing projects, normal text production is likely to occur at locations other than solely at the end of the text produced so far. For example, Bogaert had already written some preliminary sentences for the last chapter in Session 13 (at the end of the document), after which he continued writing the first chapter (and so on) above this text fragment. As a result, all normally produced text was also labelled as 'insertion'. In addition to the number and content of the revisions, the revision matrix contains information on the number of edits the writer needed to complete the action, the start and end time of the action, the length of the action and the position of the first and last character of the action, which also allows for the identification of substitutions (in case the operation occurred at the same place). Another problem associated with the revision matrix is that the keystrokes are not always logged accurately. Given that Inputlog records the position of the event according to its position on the x and y axes of the MS Word document, the text (including the position of deleted text) is not always given correctly.[36]

The S-notation offers a method for analysing the revisions within the textual context (Kollberg 1998). This represents changes in the text at their specific locations, and it provides information about the range, order and structure of the revisions

---

[36] This shortcoming of the S-notation and the description of the S-notation below, has also been published in Bekius 2021a.

(Kollberg and Eklundh 2002). It is important to note, however, that the S-notation was initially developed to visualise revisions of short writing processes in experimental settings. As such, it could not be used for analysing the material from the writing processes of Bogaert, Posthuma, Van Rijswijk, Troch and Van Pelt (see Chapter 4).

The digital component of the exogenetic material is gathered in the Source Analysis, which is based on the 'focus events' in the general analysis. Focus events are switches between different applications or other actions performed on the computer, but outside of the working document. For example, when the writer opens a notes-application different from MS Word, or opens a folder to search for a specific file. When a writer goes online, the web browser used is identified, as well as the active URL, page title, keywords used in the search and the URL of the webpage accessed after the search. In addition to this information about the window statistics, source analysis generates statistics on window transitions, which are the switches between the windows that were used (e.g. from Wordlog.docx to TASKBAR). It also provides information about the time spent in a specific window/source.

Genetic criticism and cognitive writing process research share an interest in the writing *process*, and keystroke logging ensures the availability of data with which to combine the disciplines. This combination may offer new perspectives on the literary (or other) writing process.

2.3 CONCLUDING REMARKS

Throughout this study, I employ the term 'born-digital' in its broadest sense, to refer to literature that is written within a digital environment – thus born digital – but that is meant to be published in print eventually. My choice to adopt this broad definition was prompted primarily by the fact that Inputlog was developed to capture those writing processes (written in MS Word). A secondary reason has to do with the sparse body of research on those literary writing processes using a word processor. This is not because word-processed documents are not worth studying, but because the challenges associated with studying them, as the dynamics of writing are less visible in such documents. This is due to the key feature of a word processor, as pointed out by Chandler (1995) and Kirschenbaum (2016): the 'suspended inscription'. This denotes the temporal and locative delay between the act of writing and its final documentary form. It makes writing with a word processor both flexible and fast. Although it is difficult to discern general effects of word processing for both the writing process and the writing style (given that all authors have their own individual preferences), some of the main influences that have been pointed out include difficulties in getting a 'sense of the text', reading difficulties, more revision at the level of the letter and during the first stage of composition, and more fragmentation during writing (e.g. Van Waes and Schellens 2003). Although some people claim to be able to spot whether a text was

written in a word processor, no research to date has been able to verify the effect of the word processor on the actual writing style.

The word processor also encourages a rapprochement between genetic criticism and cognitive writing process research. De Biasi (2004) distinguishes two objectives for genetic criticism: establishing an analysable manuscript (mostly in the form of a genetic edition) and the critical aspect of analysing the writing process from a specific perspective. Both objectives are still important for keystroke-logged literary writing processes, and both can benefit from insights generated by research on the cognitive writing process. First, keystroke logging data must be made suitable for genetic analysis. For a genetic edition of keystroke-logged writing processes, the way in which revisions have been analysed in writing studies may prove valuable for the encoding of revisions. This is discussed in Chapter 4. Second, for the critical dimension – the reconstruction of the logic of the genesis – the models of the writing process that have been established within the field of cognitive writing process research may help to address the cognitive processes involved in the creation of literary texts. This is the topic of Chapter 5. As noted before, two important levels of genetic analysis are the endogenesis (the composition of the text) and the exogenesis (the use of sources). In the model of the writing process under discussion (Leijten et al. 2014), exogenesis denotes the interaction between the process level and the resource level, while endogenetic analysis focuses on the process level alone. Both endogenesis and exogenesis could potentially benefit from the detailed temporal dimension made available by keystroke logging. For example, an examination of the temporal dimension could allow for examining questions such as: What internal processes could be discerned by analysing the development of the text? At what moment did the author feel the need to consult external sources?

Kirschenbaum identifies the MT/ST as the first word processor. This system stored the keystrokes so that they could be revised and printed as a clean text. The most prominent feature of word processors is that they enable the creation of a 'perfect' text. With the use of keystroke logging software, I undo this function (at least temporarily), given my aim of retracing the 'hidden' history of the 'perfect document' and storing the keystrokes for the reconstruction of the 'messy' text. It is this messy, word-processed text that emerges in this study. First, however, I focus on the materiality of these digital documents and texts.

CHAPTER 3. TRACES OF PHASES

THE GENETIC DOSSIER IN THE DIGITAL AGE

"[C]lose up against the chasm between composition and inspiration we find – always – material things and technologies." (Kirschenbaum 2016, 9)

Textual genetic research obviously begins with the constitution and organisation of the genetic dossier. The documents and other objects belonging to the work in question "materially condition the very possibility of analysis" (De Biasi 2004, 45). Given that documents are the prerequisite for genetic criticism, the shift towards the use of the computer in the writing process led Marita Mathijsen to express her fear of the demise of the genetic method of editing. According to Mathijsen, the preference of present-day authors for using computers to write will eventually lead to the point at which "no new authorial versions of works will become available in material form" (Mathijsen 2009, 243). Moreover, she presupposes that authors working on computers are less inclined to distinguish between the various phases of their texts: "An author who works electronically can no longer tell the phase in which something was finished, or when he revised it" (235). Without documents, there can be no examination of phases, as documents "connect the writing space and the writing gesture with the content (invention and organization) and verbal expression" (Mahrer et al. 2015, 154). This raises questions concerning the possibility of constituting genetic dossiers now that authors are writing their literary works within a digital environment. If it is possible, what would such a digital genetic dossier contain, and what could these digital documents have to say about the writing process? The concept of writing phase, which lies on the intersection between the material documents and the writing process, seems an ideal starting point for analysing digital literary writing processes.

This material prerequisite is the subject of this chapter. As I already argued elsewhere, "[t]he digital environment [within] which present-day literature is composed significantly changes the materiality of the sources available for textual scholarship and genetic criticism" (Bekius 2021a, 6). I therefore begin by providing an overview of what

has been said about the implications of digital materiality for the prevailing notions of the document and the text. This paves the way for a discussion of the digital genetic dossier. The content of this dossier depends largely on the author's way of working and file-saving habits. The second section therefore begins with an examination of what authors have said about their working methods and their views on the value of preserving different versions. These insights have implications for the methods that geneticists can use to study their writing processes. Discussing the preservation of versions requires defining what is to be understood as a version within a digital context. In other words, it is important to define what the "scope and range of a version" in the study will be (Bleeker 2017, 81). I argue that the concept of the session version is most useful for the present study – and for the genetic analysis of keystroke logging data in general. In the final section, I present the genetic dossiers of Gie Bogaert, Jente Posthuma, Roos van Rijswijk, David Troch and Ellen Van Pelt. Based on the MS Word documents in this dossier, I then examine the extent to which discrete phases can be distinguished within these digital writing processes. In other words, I consider whether genetic dossiers still include 'traces' of phases.

3.1 DIGITAL MATERIALITY

As discussed before, scholars are accustomed to talking about a genetic dossier as "the *physical* collection of documents […] pertaining to the work one wishes to study" (Van Hulle 2014, 11; emphasis mine). Within this context, a document is understood according to the definition advanced by Peter Shillingsburg: "the *physical* material, paper, ink, bearing the configuration of signs that represent a text" (Shillingsburg 1996, 47; emphasis mine). A document contains the 'text', which is defined as "the actual order of words and punctuation as contained in any one physical form" (46). The 'work' is thus "the imagined whole implied by all differing forms of a text that we conceive as representing a single literary creation" (43). It is possible that multiple 'versions' of a 'work' exist – in other words, multiple specific forms of the work, which the author intended at some particular moment (44). Shillingsburg thus defines 'text', 'work' and 'version' in relation to the 'document', and only the 'document' is regarded as material. Questions nevertheless remain concerning how to describe digital documents in terms of their materiality, the implications of such descriptions on the notion of the text and what exactly digital materiality is understood to be.[37]

---

[37] The following literature review of digital materiality and digital text contains parts of the following publication: L. Bekius, "The Reconstruction of the Author's Movement Through the Text, or How to Encode Keystroke Logged Writing Processes in TEI-XML", *Variants* 15-16 (2021).

3.1.1 Digital documents

In the mid-1990s, electronic technology was characterised as immaterial. This understanding was incentivised by a binary opposition that had emerged: "Inscriptional technologies were being divided into 'old' media (print, ink, manuscript) and 'new' electronic and digital ones without careful examination of the bases of these distinctions" (Drucker 2013, §3). According to this reasoning, which is also known as *digital mysticism*, the "new media marked a shift from the material to the immaterial" (Boomen et al. 2009, 8). In such discussions of materiality, 'material' is regarded as having a physical substance (Leonardi 2010). Whether it is done on paper or within a digital environment, however, writing is always *situated* in the material world (Haas 1996; Bolter 1991). At all times, writing requires at least some material implement: from pencils, fountain pens and ball-points to laptops, desktop computers and tablets with speech-recognition software.

When using a screen, keyboard and mouse to write, the human-readable text and symbols written on the alphanumerical keyboard are translated into digital code; more specifically, binary code. Contemporary computers work with bits – binary digits, 1 and 0 – to compute (Boomen 2014, 28). The language of digitality is therefore "an artificial language of digits": numbers are assigned to discrete entities, which can be manipulated through computation (28).[38] These numbers can both be assigned "to electronic circuit states as well as to human-readable symbols" (28).

For this reason, Mats Dahlström notes that digital documents as such cannot be defined materially. In digital documents, works are not constituted by the alphanumeric notation of their text, but "by the pattern of signals and tensions at the binary level of the material carrier" (Dahlström 2000, 'Digitalics'). Like writing itself, however, digital documents are always bound to *a* material carrier, in which "data files, programs that call and process the files, hardware functionalities that interpret or compile the programs, and so on" are required to produce the digital document (Hayles 2003, 274). A text document cannot be read without a word processor, and the word processor cannot work without the proper hardware. Such "hardware embodiment" can easily be overlooked, however, while focusing on the digital object itself (like the digital document). It is thus also easy to forget "the numerical base of digitality", given that this principle is attached to sophisticated human-readable and human-controllable input and output (Boomen 2014, 29). Computer screens do not display the binary level by default. Instead, they display "visual representations of the inner state of the machine,

---

[38] However, the computer also does not recognise numbers as numbers, but as different voltage states: "Th[e] pattern of binary digital code is in fact an analogical mapping of the machine state as expressed by the levels of the numerous tiny voltage circuits on the microprocessors (chips) inside the machine. Usually a lower voltage level is represented as 0 (0 to 1 volts) and a higher voltage level as 1 (2 to 5 volts)" (Boomen 2014, 29).

ordered by pictorial icons, textual menus and sub-screens (the planes we have come to call windows)”, otherwise known as the graphical user interface (GUI) (Boomen 2014, 29).

In *Transcoding the Digital* (2014), Van den Boomen argues that it is through metaphors that we can make sense of “the various nested layers of numerical representations and its translations” and make the digits accessible as interfaces, programs and data objects (32). According to Van den Boomen, interfaces, programs, data objects, and files exist only *through* metaphors, they have no prior existence: “Metaphor is what makes them legible, articulated, delineated, operative, and operable” (32). As observed by Lev Manovich in *The Language of New Media* (2001), metaphors from print, cinema and HCI (human-computer interface) are indeed employed in articulating computer interfaces, programs and content. Metaphors emerging from print – which are the most relevant to this research – include such notions as documents, files, pages and browsing, as well as the standard graphical icons for editing (pencil), cutting (scissors) and pasting (glue brush and clipboard) (Boomen 2014, 32). Throughout this study, it would not be possible to describe the digital writing practices of the authors without using these metaphors as well.

Manovich uses the term ‘cultural interfaces’ to “describe a human-computer-culture interface – the ways in which computers present and allow us to interact with cultural data” (Manovich 2001, 70). According to Manovich, many cultural interfaces “rely on our familiarity with the ‘page interface’ while also trying to stretch its definition to include new concepts made possible by the computer” (74). Like the print layout in MS Word, the GUI creates the illusion of materiality in digital documents (Van Hulle 2019, 468). In reality, however, it merely “imitates the layout of the typed page” (Bolter 1991, 17). In User Interface design, this is also known as skeuomorphism: the imitation of “the design of a similar artefact in another material or technique” (Page 2014, 131). Skeuomorphism can be helpful to familiarise users to new technologies, but can also be confusing when the skeuomorphic design element resembles a physical object without having the same features (e.g. using the visual metaphor of a book, without offering the ability to flip the pages by swiping left to right) (Page 2014, 134). When, in other words, the stretching of the definition has gone too far.

The stretching of the definition is reflected in the possibility of distributing the ‘same’ document across different computers. The everyday use of the term ‘document’ also seems to complicate its use in textual scholarship. As observed by Ries, we “speak of the ‘same’ digital document when we save ‘it’ after changing its content, after copying ‘it’ to a pendrive and ‘open it’ on a different computer with a different word processor which might display the content in a different way” (Ries 2018, 397). The definition of ‘the document’ has been stretched so far that it is tempting to think of it as immaterial. As stated by Kirschenbaum: “a digital environment is an abstract projection supported

and sustained by its capacity to propagate the illusion […] of immaterial behavior: identification without ambiguity, transmission without loss, repetition without originality" (Kirschenbaum 2008, 11). Moreover, as pointed out by Ries, digital documents "are not bound to *a single* physical entity, not even to *a single* processing system context or display application" (Ries 2018, 397; emphasis mine). It is therefore very tempting to classify digital objects as immaterial.

Digital mysticism thus lingers even to this day: "The complexity of digital code is necessarily black boxed in user-friendly interfaces, and this makes assumptions of mysterious immateriality hard to exorcize" (Boomen et al. 2009, 9). It is indeed difficult to list the materials that are part of a piece of software (e.g. Microsoft Word), given that the software itself has no physical substance (Leonardi 2010). In a discussion of digital materiality, Paul M. Leonardi focuses on 'affordances' – the ways in which we perceive and use objects – and questions the importance of *physical* matter. For Leonardi, the characteristic that renders digital artefacts 'material' is that they provide capabilities that can afford certain actions while constraining others (Leonardi 2010).

Digital writing thus requires two forms of *physical* materiality: peripheral devices (e.g. keyboard, mouse and other input devices) and display devices (e.g. screen), along with internal components (as contained by the computer case), including the CPU (central processing unit), HDD (hard disk drives) or SSD (solid-state drives) and RAM (random-access memory). These forms of materiality are also prerequisites for word-processing software to operate and to display the digital document, which – in its own way – is also material, as it affords certain actions. Digital materiality could thus be regarded as 'in-material': "as stuff which may defy immediate physical contact, yet which is incorporated in materiality rather than floating as a metaphysical substance in virtual space" (Boomen et al. 2009, 9).

Jean François Blanchette uses the term 'distributed materiality' to refer to the dispersed fashion of the digital document. As summarised by Johanna Drucker, distributed materiality is "a way to describe the co-dependent, layered contingencies on which the functions of drive, storage, software, hardware, systems, and networks depend" (2013, §6). These elements are confined to relationships with each other "that are governed by their material design and constraints in ways that have an effect on the costs and efficient operation of the system" (§6).

To describe what "the overall term materiality does and does not mean in an electronic environment", Kirschenbaum distinguishes between 'forensic materiality' and 'formal materiality', which should be understood in relation to one another (Kirschenbaum 2008, 10). Forensic materiality emerges in the physical traces: in other words, in the "variety of surfaces, substrates, sealants, and other matériel that have been used over the years as computational storage media" (10). Like "traces of magnetic poles on a hard disk that have been changed or overwritten with data", these traces can be

made visible with digital forensic methods (Boomen 2014, 151). These, traces, however, can show only whether something has happened or not. Overall, as Kirschenbaum notes, "forensic materiality rests upon the principle of individualisation (basic to modern forensic science and criminalistics), the idea that no two things in the physical works are never exactly alike" (2008, 10).

In contrast to forensic materiality, formal materiality refers to the "imposition of multiple relational computational states on a data set or digital object" (Kirschenbaum 2008, 12). This means that different software environments expose different, specific layers of the data object (13). In programming the "behaviors, features, and functionality of the word processor", the "documents were formally modeled to behave in certain ways in certain environments" (152). Opening the document with another application (e.g. a hex editor) could expose different layers of the document. Kirschenbaum's point is that the word processor is viewed as the normative environment, and the traces found upon exposing it in a hex editor are seen as a "dramatic discovery". Kirschenbaum argues that this is wrong, given that both are "programmatic computational environments applying some particular logic – a certain formal materiality – to the string of bits in question" (133). Different types of software can thus expose different instantiations of the same 'text'.

3.1.2 Digital text

Digital texts are thought to pose some issues for textual studies, including (as noted by Graham Barwell) "their ease of reproduction, alteration, corruption, and transmission" (2005, 418). Moreover, it is a challenge to determine "the relationships between apparently identical copies" of digital texts (418). These assumed characteristics of digital texts could nevertheless be categorised as part of "the common assumptions about electronic textuality" that characterise what Kirschenbaum refers to as the 'medial ideology'. These assumptions include the following: "that electronic text is hopelessly ephemeral, that it is infinitely fungible or self-identical, and that it is fluid or infinitely malleable" (Kirschenbaum 2008, 50). With regard to the *ephemerality* of the electronic text, Kirschenbaum states that a deleted file it is not immediately wiped off the hard drive. The space is only made available for overwriting, and it may take some time before this actually happens (51). Moreover, each time a file is opened and modified, its physical storage location is changed. The earlier 'incarnations' of the file thus also persist until it is overwritten (51). At the same time, multiple instances of a file may exist. Word processors may include an auto-save function, "that writes a snapshot of an open file to the disk at set intervals" (51). These files often have opaque names (e.g. ~WRL0005.tmp), and they do not appear in standard directory listings (51). Other possibilities for extracting remainders of deleted files may be found on the swap space – a portion of the hard disk as an extension of the RAM – and the slack space. In short,

Kirschenbaum states that there is simply no way of knowing the number of instances and states of a single file that the operating system houses in all its various locations (52). If these instances of the file are found, they may expose variants of the text that were initially deemed lost.

The second assumption of the medial ideology is the supposed *fungibility* of an electronic text, meaning that digital files would be interchangeable with and indistinguishable from other digital files (of the same type). As noted by Kirschenbaum, however, for a forensic investigation, it is not enough simply to copy the files, as they would not be the 'same'. Instead, it is necessary to create a bitstream image of the original file system: "every bit recorded on some original, physical instance of storage media is transferred in linear sequence to the copied image, whether it is part of a file currently allocated in the FAT [File Allocation Table, file system] or not" (53).

The final common assumption is the *fluidity* of electronic texts. According to Kirschenbaum, although we often fixate on the fluidity and malleability of electronic texts, as well as on their putative instability, "there is nothing essentially fluid about data in an electronic environment" (56). With the use of tools and technologies, documents can be electronically signed and sealed (57). More specifically, a file may be secured in such a way that it could not be altered by anyone except the author.

When an author uses a word processor, the document visualised by the GUI remains neat with every modification to the text: additions are always represented as inline insertions, and deleted text 'disappears' from the surface. In the words of Irène Fenoglio, "on the computer, the word is 'erased' and not literally crossed out; it is immediately 'replaced', if the replacement takes place during the course of the writing; it is immediately dispensed with" (2009, 49). As demonstrated above, however, while the deleted text may be 'erased' from the GUI, it may not be entirely erased from deeper levels of digital materiality – the materiality of the non-arbitrary inscriptions and patterns. The word processor is designed only "to facilitate the production of a single manuscript that conceals through its very process of composition the errors, emendations, and edits associated with material (paper) based technologies" (Alexander 2015, 90). They are meant to cover our traces.

The formal materiality in which a document is opened may thus influence how the text is rendered. It is not obvious, however, whether the meaning of the text that is displayed is also affected by its materiality. The definition formulated by Shillingsburg – "the actual order of signs that represent the text– implies that it is not (1996, 46). A text has no material existence, "since it is not restricted by time and space" (46). According to Shillingsburg, when the text is transported (e.g. from print to electronic media), the text remains the same:

> it is possible for the same text to be stored in a set of alphabetic signs, a set of Braille signs, a set of electronic signals on a computer tape, and a set of

magnetic impulses on a tape recorder. Therefore, it is not accurate to say that the text and the signs or storage medium are the same. If the text is stored accurately on a second storage medium, the text remains the same though the signs for it are different. Each accurate copy contains the same text; inaccurate or otherwise variant copies contain new texts. (1996, 47)

This concept of 'text' has been criticised by N. Katherine Hayles, specifically with regard to electronic literature. According to Hayles, Shillingsburg's concept implies that the text is bound to a specific physical location, while electronic text comes into being as a *process*, which "includes the data files, the programs run, as well as the optical fibers, connections, switching algorithms and other devices necessary to route it from one networked computer to another" (2003, 267). By itself, the digital text does not exist; what will actually appear on the screen depends on the formal materiality in which the text-bearing file is opened. Another objection advanced by Hayles is that the definition implies that 'text' does not include such qualities "as color, font size and shape, and page placement, not to mention such electronic-specific effects as animation, mouseovers, instantaneous linking, etc." (267). For electronic literature, as well as for literature in print, these qualities may be significant to the meaning of the text. For this reason, Hayles advocates discussing texts '*both* in terms of their conceptual content and their physical embodiments', as well as in terms of how these elements work together (278).

It is exactly this focus on choice in the editorial process that is present in Shillingsburg's 'formal orientations', as presented in *Scholarly Editing in the Computer Age* (*SECA*).[39] By 'formal orientation' Shillingsburg means "a perspective on forms which leads to the selection of one set of formal requirements over another" (1996, 6). The orientation most relevant with regard to Hayles critique is the 'bibliographic orientation', as added by Shillingsburg in the third, revised edition (1996).[40]

---

[39] The first, limited, edition was published in 1985, which was succeeded by a second edition in 1986. In 1996, Shillingsburg published a third, revised edition. Shillingsburg continued finetuning the orientation in the article entitled "Orientations to Text" that appeared in *Editio* in 2001, and again in 2015, together with Dirk Van Hulle in "Orientations to Text, Revisited". For an extensive discussion of the orientations, see the dissertation by Wout Dillen (2015).

[40] Hayles makes one reference to the bibliographic orientation: "Although he admits, some hundred pages later, that "Proponents of the bibliographic orientation have demonstrated beyond the argument, I believe, that the appearance of books signifies a range of important meanings to their users," he apparently does not thin this imbrication of physical form with meaning requires a different notion of textuality" (Hayles 2003, 266). She therefore does not agree to include material aspects only in the biographical orientation to text, instead arguing for a more general recognition of material aspects. In "Orientations to Text, Revisited" Shillingsburg and Van Hulle replaced the 'bibliographic orientation' and the 'documentary orientation' with the 'material orientation'. This 'material orientation' is divided into two

Shillingsburg describes that this orientation broadens the definition of text "to include all aspects of the physical forms upon which the linguistic text is written", including the type of font, the texture of the paper, the colour and the binding style, amongst many others (1996, 23). But what is the role of the material aspect in the genetic orientation to the text? As demonstrated previously, the 'genetic orientation' – which was added to the 'orientations' in 2015 – has an affinity with the 'material orientation', as it also concentrates on the documents, which are regarded "as sources of evidence of textual development and change through time" (Van Hulle and Shillingsburg 2015, 36). As mentioned by Van Hulle and Shillingsburg, manuscripts and typescripts can exhibit "clear signs of different writing tools or different handwritings", which allow for the further subdivision of the version into several stages (39). Moreover, it is widely acknowledged that not only "the order of words and punctuations" in these documents are essential to textual genetic research, but also the material features thereof, which include:

> changes in the features and dimensions of the handwriting in the course of writing, changes of ink, and writing materials, the specific geometric constellation of the text on the physical page, of lines, deletions, corrections, graphical signs, and drawings to one another, chemical characteristics of the writing materials and paper. (Ries 2018, 398)

To understand the text and its genesis, the documentary context is regarded as a crucial source of information (Pierazzo and Stokes 2011; Dillen 2015). Complex handwritten draft materials often consist of chaotic pages that contain multiple textual fragments that were written at different positions and in different directions on the page. Detailed consideration of elements of the document – the layout, the stratification of writing and the disposition thereof – in combination with an understanding of the text, is required in order "to gain insight about the composition, time of revisions, and flow (flux) of the text" (Workgroup on Genetic Editions 2010, §1.3). In analogue documents "the dialectic between a document's physical limitations (as a two-dimensional surface of limited size) and the internal structure of its different writing zones on that surface often contains important clues in the investigation of the text's writing process" (Dillen 2015, 71).

With regard to born-digital literature and the creative process, a focus on the material aspects has also been emphasised. The files of most interest for genetic criticism go beyond the simple text file, as the textual content is encoded by means of

---

subsets: the lexical and the bibliographic (Shillingsburg and Van Hulle 2015, 20). Both of these orientations "consider the physical document to be the basic unit of textual evidence, but where the lexical focuses on the text, the bibliographic focuses on the material object" (Shillingsburg and Van Hulle 2015, 31).

a proprietary language of the word processor, in which the text is not merely represented as a string of characters, but also formatted (Linkès 2017). Authors have remarked on their use of fonts throughout their writing processes. For example, Nikesh Shukla states that she likes "playing with font sizes to indicate levels of tension in dialogue, and Word is at the tail end of being so ubiquitous that it makes it the easiest thing to use" (Shukla in Johncock 2013). In addition, Joanne Harris changes the font from Perpetua to Times Roman when editing her texts, which helps her to spot things that she has missed (Johncock 2012b).

These material aspects have also been acknowledged by geneticists like Fenoglio, who considers the formal aspects so significant as to regard the same text in different forms as different versions: "when the printed pages do not have any handwritten corrections, they are presented in a different typographical form: different or larger print, wider margins, altered paragraph breaks, title specified or not" (2009, 51). The authors of "Digital Materiality: Preserving Access to Computers as Complete Environments" note that it is important to take account of the "uniqueness of individual instances of both hardware and data objects, coupled with an awareness of how the affordances of particular systems, environments, and technologies" affect the creative process (Kirschenbaum et al. 2009, 110). For instance:

> knowing how much of a document would be visible on a screen at one time – knowledge that depends on the physical size of the display hardware, its screen resolution, and preferences as defined within particular application software – can be critical to understanding aspects of an author's composition process. (Kirschenbaum et al. 2009, 110)

Kirschenbaum et al. therefore call "for preserving the integrity of the original hardware and storage media" (111). Although I acknowledge the importance of these material aspects of hardware and software in the overall creative process, authors are not always willing to hand over their entire computers – only the files stored on them – and it is not always possible to access the hardware. In this research, therefore, I explore how keystroke logging can be a way of investigating the genesis of a text without having access to the author's computer, but also without having to acquire the skills of digital forensics.[41] To this end, I consider only the formatting options offered by the word processor. These material aspects make it possible to study how formatting settings are used in the writing of the text, which I regard as crucial to understanding both the meaning of the text and its genesis.

A keystroke logger can be used to reveal what would otherwise remain hidden, with all the information about the changes (i.e., the messy part of the writing) stored in

---

[41] However, in the case of keystroke logging a prerequisite is of course that the author used a keystroke logger during their writing process, which is, naturally, exceptional.

a separate file. Most of the crucial information about the genesis of the text is thus moved to the keystroke logging data. The reconstructions of the writing process as logged with the keystroke logger – and the visualisations thereof – represent a text that did not exist as such, nor did its formal aspects appear as I visualise them (see Chapter 4.2). While I am fully aware of the importance of including the material aspects of text in a genetic analysis, in this study my approach is largely restricted to examining the text in terms of the actual order of words and punctuation, as represented by the GUI of the word processor. This does not diminish the fact that the material aspects would play a pivotal role in the investigation of other genetic dossiers, especially those without keystroke logging data.

3.2 WHAT TO EXPECT IN A DIGITAL GENETIC DOSSIER

For Kevin Molloy, manager of the Manuscripts Collection, State Library of Victoria in Melbourne, the "ideal literary collection captures the full gamut of a writer's work", thus including "background notes and research; annotated books and critical editions; literary drafts; photographic components; audio material; personal journals; literary logs; objects like keepsakes or awards; correspondence with publishers, editors and friends; editors and printers' proofs, and final copies" (2019, 328). The material characteristics of a genetic dossier nevertheless depend on the author's working method and preferences for particular hardware (e.g. Mac or Windows; backup in the cloud or on a physical device). Moreover, an author's computer skills also influence "the composition and the degree of preservation of [their] personal digital documents" (Weisbrod 2016, 138). Richard Rorty had an automatic file backup application enabled, but he had "a lack of interest in file management and deletion", which created a large digital archive (Schmitz 2010, 3).

According to Mathijsen, the working methods that authors employ to preserve different versions depend on whether they started their writing careers by writing with a typewriter. Mathijsen argues that authors who started writing *before* the computer age tend to make more printouts. For example, the Dutch author Harry Mulisch (1927–2010) continued to make daily printouts of his work, as he did not "trust it [the computer] completely" (Mathijsen 2009, 235). Mulisch used the computer not as an archive, instead keeping printouts for bits of text that he could use later. In contrast, Mathijsen discusses Robert Anker (1946) and Connie Palmen (1955) as examples of Dutch authors who have worked predominantly in the computer age. Anker writes prose on a computer and does not save intermediate versions of the text: "I don't keep anything. I just write over things that I'm not satisfied with" (Anker, quoted in Mathijsen 2009, 236). The same applies to Palmen, who makes printouts for corrections but destroys these versions after completion. As she clarified to Mathijsen: "I do corrections by hand, and then enter them into the computer. Only when the book is finished do I

throw away the versions I've printed out. I don't see the importance of all those versions" (Palmen, quoted in Mathijsen 2009, 237). After discussing the working methods of these authors, Mathijsen concludes: "While in the case of Harry Mulisch it is only impossible to reconstruct the *Sofortkorrektur*, in the other two cases, which are illustrative of the new generation of writers, the whole pre-text has materially vanished" (2009, 237).

In short, Mathijsen assumes that authors who write on the computer will not save any trace of their writing process, let alone any versions. To test this assumption, therefore, I review what can be learned about the saving practices of other authors. I also consider how these practices will eventually lead to different approaches from scholars who are interested in the writing process. In turn, I address how this affects the prevailing conception of what a 'version' is and how I approach the notion of the version throughout this study.

3.2.1 Saved for later

The fact that, in the computer age, a genetic dossier depends on the author's working method does not represent a drastic change from the analogue period. As noted by De Biasi, the set of manuscripts "obviously varies in quantity and type according to the work and the author under consideration" (2004, 38). Palmen might thus not be especially illustrative of 'the new generation of writers', given her strong stance on the value of draft versions in general. Instead, Palmen's statement that she does not keep her drafts and old manuscripts says more about her own view as an author of the value of these documents than it does about the impact of the digital environment on her habit of deleting previous versions. This becomes apparent in *Gekrenkt & hongerig* (2016), a special jubilee issue of the literary journal *Das Mag*. The editors of *Das Mag* asked Dutch and Flemish authors to send in early work that had not been published before (e.g. a passage from a rejected manuscript). Palmen's contribution was a non-contribution – in her opinion, she had nothing that was worthy of publication. In the introduction to this non-contribution, Palmen wrote: "Ik heb een map met oud werk geopend en vervolgens integraal verscheurd. Te veel walging" (Palmen in Donk and Meer 2016, 29).[42] Here again, Palmen propagates her stance on old work and older versions: they are not worth preserving, let alone sharing. There is thus apparently no specific causal relationship between Palmen writing within a digital environment and the absence of drafts and earlier versions of her work.

What about other authors? For the pilot project *Het Literaire Werk 2.0*, an online questionnaire was designed to obtain a general idea of the working and storage habits

---

[42] Translation: "I opened a folder containing old work and then completely destroyed it. Too disgusting".

of Dutch and Flemish authors (for an elaborate discussion of the results, see Buschenhenke 2016). The questionnaire was open from September to December 2015, and it was completed by a total of 161 participants. The results point to the predominance of a hybrid working method – using both a computer and pen and paper. The majority (78%) of respondents identified the digital part as the most important. Although nearly all (95%) reported using pen and paper to take notes, they did not do so exclusively: 22% said that they also used mobile devices to take notes. For 88% of the respondents, the computer (laptop and desktop) was the most important writing tool, although many (74%) sometimes still used pen and paper for the actual writing (Buschenhenke 2016, 11). In response to the question 'Do you keep different versions during the writing process?', the majority (84%) responded positively. The ways in which the respondents interpreted the term 'version', however, were quite ambiguous: answers like 'around the clock' and 'before every pause moment' point more towards the act of saving the file (pressing 'save') than towards the practice of saving it under a different file name (12). In-depth interviews with six authors indicated that versions mostly serve to demarcate different stages in the writing process, such as sending a first rough draft to an editor or editing the text on paper (12). All writing in the intermediate periods takes place in the same file, without saving different versions of the document (12). In contrast, however, 70% of the participants reported keeping a file containing previously deleted text elements. Half of the participating authors said that they kept their working documents after finishing a book, while one third keep only part of this material (12).

Unlike Anker and Palmen, some present-day authors certainly do keep multiple versions of their work, albeit usually only a few (see also Vásári 2019). In *eSchrijven* ('eWriting', 2014), a guide to help writers learn how to use their digital tools efficiently, Louis Stiller includes a subsection entitled "Maak elke dag een nieuw bestand",[43] in which he asks, "Hoe krijg je die schitterende formuleringen terug die je drie dagen geleden per ongeluk om zeep hebt geholpen?" (2014, 99).[44] By means of an answer, he advises writers to make a new document every day, or every part of the day. According to Stiller, some authors indeed do this:

> Sterker nog: ik ken verschillende schrijvers die elke dag een nieuw bestand aanmaken via *Bewaren als (Opslaan als)*. 28032014Verhaal_Sophie is op deze manier de voorganger van 29032014Verhaal_Sophie. Zo weet je in elk geval

---

[43] Translation: "Make a new document every day".
[44] Translation: "How do you get back those dazzling formulations you accidentally deleted three days ago?"

> zeker dat je werk in alle versies bewaard en toegankelijk blijft. En zo hebben je latere biografen en andere geïnteresseerden er ook iets aan. (2014, 99)[45]

This points towards an awareness that it is indeed problematic to 'overwrite' text within the 'same' digital document. This applies both to authors – as it prevents them from returning to deleted sentences – and to future researchers – as it leaves them no material to study. One writer who practices what Stiller preaches is the Flemish author Bart Moeyaert (1964). He kept 24 MS Word documents containing different versions of his book *Het paradijs* ('Paradise', 2010) – the second book in the Haydn trilogy – varying from first drafts to the final version. Most of these documents were saved using the *Save as* function in MS Word. As a result, Moeyaert's files have only two different creation dates: 20 February 2009 and 20 February 2010. The 'last modified' dates do differ, however, and they provide an indication of the chronological order of the files. Moreover, the names of the files are also of interpretative value. For example, the filename 'HET_PARADIJS_BARTMOEYAERT_020310GOD kopie.doc', indicates the date Moeyaert had worked on this file, in addition to providing a clue about the content of the draft: it is written from the perspective of God. As this example illustrates, not all draft material of the 'new generation of writers' will eventually vanish.

The work of Bénédicte Vauthier (2016; 2019) also offers assurance that different versions of born-digital works will continue to exist, although their form may not always correspond to what geneticists are used to – or comfortable – working with. Vauthier interviewed several Spanish writers about their working methods. In terms of versions, Vincent Luis Mora (1970) seems to have a working strategy that is similar to Moeyaert's:

> When I spoke about versions, I was referring to the same base document with some changes; sometimes with many changes, sometimes with fewer. The base document, however, is always the same and therefore keeps the same creation date. Each one of these digital copies that you [Vauthier] have is a backup copy of the same document in progress, in continuous change: I enter the document and I add new things, correct or delete some of the old ones, so the document is not the "same" anymore when I finish. This is why I said it is a version, while you called it a "changed draft". When I create a digital copy, it is because I think that I have made enough changes in the original to save it separately, and the chronological order is not marked by the date of the file, but by [...] [the date of the backup]. (Vauthier 2019, 41)

---

[45] Translation: "In fact, I know several writers who create a new file every day using *Save as*. 28032014Story_Sophie is thus the predecessor of 29032014Story_Sophie. This way, you can be sure that your work will be saved and accessible in all versions. Your later biographers and other interested parties will also benefit from it".

Although such a strategy may thus be quite commonly employed by authors, Vauthier "gave up studying them", as she "did not find a satisfactory form of exploring the genetic dossier composed of versions which, in addition, as the author suggests, all have the same date of creation" (2019, 41). This again indicates that the adequate study of born-digital writing processes requires geneticists to step out of their comfort zones.

Vauthier nevertheless did take on the task of investigating the digital files that the Spanish writer Robert Juan-Cantavella had saved during the writing of his novel *El Dorado* (2008). After comparing digital documents and analysing the tree structure of the folders and other metadata (e.g. file title and creation date), Vauthier concludes that, "although the dossier does not contain the normal traces of writing – cancellations, additions, shifts – whose absence […] would appear to make our analysis practically impossible, collating and comparing the digital documents and files gives us more than a sound basis to allow a meaningful genetic investigation" (Vauthier 2016, 175).[46] And Irène Fenoglio (2009) even studied the 'word processed manuscripts' of the short story *Fête des Chants du Marais* by the French author Pascal Quignard by focussing solely on the corrected printouts. After conducting in-depth interviews with nine Croatian authors, Micunovic, Marčetić, and Krtalić also highlight different ways of documenting phases during the text development: while some authors numbered and dated the versions, others did not document any phases (2016, 9)

Like Kirschenbaum, the only conclusion that can be drawn at this point is that it is wise to avoid making generalisations about the working methods of authors: "The reality, of course, is that *every* writer's individual habits and practices are deeply personal and idiosyncratic, and it is difficult, if not impossible, to extract patterns in support of generalizable conclusions – beyond the intensive intimacy and commitment that the act of writing invariably demands" (2016, 23). What remains important is to make authors aware that whatever digital legacy they create is worth preserving, and that they can and should play an active role in doing so (Micunovic, Marčetić, and Krtalić 2016).

3.2.2 From the harddrive, by way of Google Drive to the keystroke log

Despite the calls for authors to help preserve their digital legacies, it must be acknowledged that some authors will always keep writing in the same document until the moment the final version is completed. Even in these cases, however, some traces of the writing process may continue to exist. As demonstrated above, digital materiality consists of different levels, and traces of the digital document may be "scattered across the hard drive" (Ries 2018, 412). Mathijsen's assumption that no traces will remain of

---

[46] This description of Vauthier's investigation of the digital genetic dossier of Juan-Cantavella also appears in Bekius 2021a.

the digital writing process, is – according to Ries – a 'screen essentialistic' perspective, which only focuses on the formal materiality of the GUI (2010, 152). To echo the words of Ries: hard disks do not forget so quickly (152). As evidenced by Ries' work, different traces can be 'uncovered' using digital forensic tools: "In many cases, large amounts of deleted textual draft versions can be recovered from archived storage media, depending on the specific material transmission and historical technological setup" (Ries 2017, 131). Given that the digital materiality is inherently a distributed materiality, the genetic dossier changes significantly. Ries points out:

> The digital document and the digital trace are not self-sufficient pieces of evidence, as they have to be read against the backdrop of their original system context, a specific historical ensemble of hardware, operating system, application and usage data. Files, textual data, metadata and contextual information relevant for the born digital *dossier génétique* are to be found distributed throughout several layers of the archived system. (Ries 2017, 131)

Instead of looking solely to the saved documents, attention should be paid to the specific hardware and storage media that authors have used (Kirschenbaum et al. 2009, 111). Hard drives contain crucial information about the genesis of the text: "they are likely to contain draft snapshots and traces of writing processes preserved in digital documents, recoverable temporary files, residuals of deleted files and metadata in numerous locations" (Ries 2017, 132). Based on these files, various applications (e.g. hex editors, binary parsers, undelete tools or file carvers) can be used to reveal revisions and intermediate steps in the writing process (Ries 2018, 418). Exposing a digital document to a hex editor means exposing it to another programmatic computational environment. In other words, it places the document in a formal materiality other than the default GUI of the word processor. In doing so, the data object may manifest additional, or hidden data. For example, the 'scratch file' [~WRS0003.tmp] containing the first paragraph of the first chapter of Kling's essay *Herodot* (2005) "contains an almost complete protocol of the first writing phase of this paragraph in the form of text additions and textual variants from the first written line on to a point of time between [~WRL3681.tmp] and [~WRA1775.wbk]", which are two other temporal and backup files (Ries 2017, 141). After extracting the fragments, Ries was able to reconstruct the nonlinear development process of this paragraph, with the inclusion of editing phases and correction of typing errors (142). Ries nevertheless notes that, "although the relative, layered sequence of edits can be determined […] due to textual fragmentation, it is not in all cases possible to determine a consistent text status at any given time with certainty" (142).[47]

---

[47] This description of Ries' digital forensic method and his results also appears in Bekius 2021a.

A more modest data carrier than a hard drive (e.g. a floppy disk) can also contain more information about earlier versions of the digital text than might be expected. Doug Reside's research on Jonathan Larson's floppy disks containing drafts of the script for *RENT* also demonstrates how 'formal materiality' has an effect on what can or cannot be seen (Reside 2014, 71). The disks contain Word files that show the development of the script for *RENT* "from a summary with a few song lyrics to the show presented off-Broadway just before Larson's death" (71). As Reside explains, however, some of the drafts were not visible by the default representation:

> Using a text editor (such as Notepad++ for Windows or TextWrangler for the Macintosh), it is possible to open these same files on a modern operating system and uncover layers of work saved prior to the last save. Microsoft Word 5.1 included feature called "fast save. […] The "fast save" feature sped the process somewhat by replacing the entire file only once every 14 saves or so. In most cases, the software would simply append changes to the end of a file with information about where they belonged in the original document. When the file was opened in Word 5.1, the software would integrate these changes back into main text; however, by opening this file with a simple text editor, the original ("base") text could be recovered and with all of the later emendations revealed at the end of the document. (71)

Opened within a different formal processing context, each of the Word files thus contained several versions. This once again stresses that the digital text is not quite ephemeral, and that textual genetic research might benefit from looking further than the default processing context.

Melina Vásári describes another way in which the genesis of a literary work can be analysed: through access to Google Drive. If the author works in Google Drive, the work is automatically saved on a server, which "makes possible the arrangement and rearrangement of the lines, corrections, additions and deletions without having to rewrite the whole poem, and still, the older version(s) can be revisited, nothing is lost and it can be easily read, not like in the case of the paper" (Vásári 2019, 302). The program completely maps the writing process, which allows for an analysis of all the modifications. In turn, the Google Chrome extension 'Draftback' can replay the complete writing process, showing every revision and the corresponding time and date. This software can also create graphs and statistics of the writing process (304). Contrary to a forensic investigation, for which the geneticist must learn how to use all these forensic methods, the work of the geneticist has been completely taken over by the software developers of 'Draftback', as 'the design is defined by the program' (305). Related to acquiring access to Google Drive, which is a cloud-supported way of self-archiving, is Dirk Weisbrod's proposal for archives to set up a cloud-based self-archiving system, in which authors have the opportunity to upload files – which would

thus be a copy of the original file, which can still be modified – as well as to work directly within the cloud environment (2015; 2016). In this way, the "time span between an object's creation and its preservation, this critical factor of digital preservation, reduces to a minimum" (Weisbrod 2016, 148).

In this current study, I investigate yet another possibility for gathering material on the genesis of a literary work: keystroke logging. The invitation to authors to log their writing processes with a keystroke logger can be seen as a form of a 'pre-custodial' archiving – which begins in the period before the actual acquisition of the literary archive – as advocated by Weisbrod. Given that the keystroke logging tool Inputlog records every keystroke and mouse movement while the software is running – thus throughout the entire composition process of the literary work in question – the original hard drive of the laptop or PC is not a prerequisite for the constitution of the genetic dossier. Rather than consisting of various 'traces' of the writing process, the genetic dossier now consists primarily of 'recordings' of the writing. One prerequisite in this case, however, is that author must indeed use keystroke loggers in their writing process. This means that they may be asked by the geneticist to record their writing processes for research purposes, or they may choose to do so for their own archival practices or work habits. Either way, authors must incorporate the activation of the keystroke logger into their working methods. This could also be beneficial for the author. For example, it could help to retrace 'old' versions of sentences that were previously deleted – like the aforementioned issue pointed out by Louis Stiller.[48] When the program is running, authors do not have to actively save the file under a different name, as the software does it for them.

The developments discussed above suggest that a genetic dossier can still be constituted in the digital age, but that it can no longer consist exclusively of *physical documents*. Amongst various other possibilities, it can be assembled using the hard disk (and other hardware), which will also contain deleted files and temporary files; with exclusively digital documents saved and provided by the author; with access to Google Drive or other cloud-based storage and file-sharing services; or with all the recordings of the writing process made with a keystroke logger. There are different levels of digital materiality that can provide information on the genesis of the digital text, from the hardware to the binary code, to the digital objects represented by the GUI. Each of these digital genetic dossiers demands a different method of investigation, and each will generate a different genetic analysis. In this research, I describe only the possibilities of

---

[48] Part of the 'Track Changes' project is the development of a version of Inputlog that works with Libre Office, which aims to improve the program's usability and convenience for authors, so that it can be used for their own archival practices. The development and promotion of this keystroke logger are intended to help spread awareness of the existence and usefulness of such a tool for writing practice, as well as for future textual research.

keystroke logging, even though this is unlikely ever to be the default case in contemporary writing processes. The aim of this research is to investigate how keystroke logging can serve as a method for analysing textual genesis without having access to the author's computer and without resorting to digital forensic methods. For this reason, I draw solely on the material gathered with the keystroke logger Inputlog in the constitution of the digital component of the genetic dossier.

3.2.3 Versions in keystroke logged writing processes

As born-digital materials become increasingly present within literary archives, one of the basic questions a scholarly editor must consider has to do with the relationship between the digital documents, analogue manuscripts and printouts contained within a single genetic dossier, as Kirschenbaum did with regard to the files of Toni Morisson's *Beloved*: "Were the contents of the Word files duplicated in the page scans of hard copy materials […]? Or did the recovered files represent unique states of the text? And at what stage of the composition process?" (Kirschenbaum 2021, 18). He found that one file – BELOVED3.DOC – contains a "minor variant state of the text not otherwise represented in the collection materials" (18). The digital file represents a unique version of the text in which Morisson still had to decide on the final sentence of the novel, "A hot thing" or "Beloved" (18). According to Roger Lüdecke, "versions are textual unities that show linguistic variation, adding that they are of 'interpretative importance'" (Lüdecke quoted in Bleeker 2017, 81). Even though the digital file of *Beloved* exhibits only a minor variant state, it certainly fits within Lüdecke's definition of a version. The same is true for the relationship between the digital files and printouts of the American philosopher Richard Rorty (Light 2014). After comparing the digital and paper files in Rorty's archive, Dawn Schmitz – an archivist at the Department of Special Collections and Archives at UC Irvine Libraries – observes that, while "some of the printouts in the files reflected the content of the floppies", most of the digital files contained unique content, which made them "absolutely necessary for Rorty's scholarly work" (Light 2014, 20). In addition to establishing the relationship between digital and analogue versions of a text, however, it is obviously important to specify exactly what is to be regarded as a version and to determine whether this concept is still applicable with regard to born-digital literature (Van Hulle 2019; 2022).

Deciding on what is to be understood as a version is one of the first decisions a scholarly editor must make (Bleeker 2017, 81). It is perhaps even more pressing with respect to born-digital material. After having studied the digital files of Robert Juan-Cantavella's novel *El Dorado*, Vauthier emphasises that genetic criticism and scholarly editing in the 21st century "will have to focus on the question of digital versions and variants – and on the complexity of the problem of *versions*" (2019, 37). The concept of

'version' is highly prone to interpretation, and it has been the subject to various debates (see Pereira 2021, for a literature review of the definitions of version). Peter Shillingsburg defines a version as "one specific form of the work – the one the author intended at some particular moment in time" (1996, 44). With intention, Shillingsburg refers to "a single or at least a coherent overall intention" (44). For example, "the manuscript may represent the version produced in the heat of creativity, while the revised galleys may represent the author-edited version, and the final edition may represent the author's most mature intentions" (44-5). I return to the term 'intention' in Chapter 4.3.

In her interviews with three writers of the "New Spanish Narrative" about their digital working methods, Vauthier uses the term 'version' to refer to a text "that slightly differs from another text, saved by the author before or after the first text", which would show variation and allow for the reconstruction of the writing process (2019, 40). For Vauthier, one important aspect of a version is "textual identity" (40). She further states that, due to the ease of making a new 'version', the concept of version – and the process it denotes – is fundamentally different for born-digital material than it is for analogue material:

> Even if we content ourselves with the version saved voluntarily by the authors, we could see that they do not hesitate to duplicate the most complete version of the text in order to avoid regrets in case they have to "come back" to a previous one. Although this duplication is not merely mechanical, it is – from a *critique génétique* perspective – a fundamentally different process compared to the isolated revision and the revision by rewriting of a text in the analogue medium. (Vauthier 2019, 43)

That 'version' is indeed an "editorial construct" (Shillingsburg 1996, 44) becomes clear upon comparing Vauthier's definition is compared to the one Fenoglio employs in her genetic analysis of the 'word-processed manuscript' of Pascal Quinard. In that analysis, Fenoglio 'limits' the definition of version to the printouts of the digital document: "by "version" we define any sentence of a text on a new paper format, in other words any printout of a computer input, whether modified or not by any handwritten corrections, preserved by the author as part of his dossier" (Fenoglio 2009, 51). This decision was informed by editorial considerations of the way in which Quinard utilised the word processor, the fact that these printouts are dated, and that even versions that do not exhibit any textual variance differ from one another in their formal features. These printout versions can thus have "as many states as stages of correction, which are in general distinguished by the colour of the pen used in the re-reading" (51).

Fenoglio thus opts to define a digital version by considering only non-digital instances of the work: the printouts. As stated by Mathijsen, however, the status of printouts is difficult to determine with born-digital works of literature. For example, if

an author has ten identical printouts, which one is the original? (Mathijsen 2010, 48)? For Mathijsen, only the digital document would be the 'original', unless it has been deleted, or unless the printout contains handwritten corrections or notes (48).

As in Fenoglio's definition, 'versions' can be linked to specific material entities, and they are often – but not always – distinguished from 'layers' and 'phases'. Elsa Pereira summarises the discussion of the concept of versions and observes that "the discussion of whether versions must correspond to different material documents has involved three main contentions" (2021, 110). The first is that "each document represents one version, irrespective of any layers within it" (110). This stance is adopted primarily by scholars taking on a documentary, bibliographical, or sociological approach to editing (110). Second, some scholars argue that layers represent different versions. A revised draft thus contains different versions: "the text that was initially written and the successive texts created by intradocument layers combined with the invariant parts" (111). The third position outlined by Pereira is complementary to the second and groups different documents and individual layers into a single version: "[i]nstead of a physically embodied text, versions become abstract entities with a certain range of variation" (112).

Van Hulle clearly indicates that, for Mathijsen, each layer of revision corresponds with a new version (2022, 194), but that it may be impossible to distinguish between writing layers (e.g. when the revisions are made in the same ink):

> Take for instance a manuscript in black ink which contains revisions in the same black ink, and this first revision campaign was followed by a second one, again in the same black ink, so that the two cannot be distinguished: they may be separate versions in theory but it seems impossible to treat them as such in practice. (2022, 194)

According to Van Hulle, this also applies to born-digital literature: "the material link with writing 'layers' in the document disappears in born-digital literature" (195). With keystroke logging, however, these layers can become visible again, as addressed in the following chapter. However, this possibility introduces another issue for the definition of versions. For example, consider Siegfried Scheibe's definition of 'version': "Textual versions are achieved or unachieved elaborations of the text that diverge from one another. They are related through textual identity and distinct through variation" (1995, 207). As Van Hulle points out, this implies that even a single change made at the level of punctuation, or a single keystroke would create a new version. For born-digital material, and especially for keystroke-logged writing processes, this would quickly become unworkable (Van Hulle 2019, 475). Within the field of writing studies, this has also been acknowledged by Mahlow et al. (2022), who state that "[t]echnically, we could create a version after each keystroke, which would result in a large amount of versions to be parsed, but two adjacent versions would differ in one character only" (Mahlow,

Ulasik, and Tuggener 2022, 'Version'). The consequence of this was previously illustrated by Mahlow (2015): for a sample text of 160 words in the final version written within 23 minutes, this would result in 2130 versions.

For this reason, Mahlow et al. consider a version to be "a point in the production history of a text that is deemed relevant based on particular criteria" (2022, 'Method'). It is therefore a "specific text-produced-so-far" (2022, 'Method'). The criterion they apply is a *change in production mode*, which is a switch between one of the following modes: a) "continuous writing at the leading edge" of the text-produced-so-far; b) "continuous deletion of something"; or c) "continuous insertion of something into existing text" (2022, 'Version'). They then introduce 'text history' for "the collection of versions that yield the final text product" and it is meant to capture "the text production process at large and is intended to gain insights into general writing habits of the writer" (2022, 'Text history [...]'). Mahlow and colleagues further developed a stand-alone open-source application known as THEtool for visualising text history (at a macrolevel), which "offers a compact overview of the flow of the text production and enables a quick visual comparison of writing habits between writers" (2022, 'Creation [...]'). For a genetic analysis, this visualisation may also yield promising results for a macrogenetic analysis of nanogenetic actions, although it is not very suitable for the examination of the genesis of the content of the *text*. Distinguishing versions based on different modes of writing at the level of the whole text does not seem workable either: for literary writing processes logged with a keystroke logger, the number of versions will quickly approach several hundreds.

Within the context of genetic criticism, Van Hulle argues that the size of the version should be adapted to the size of the textual unit that is of interest to the geneticist or scholarly editor. Van Hulle clearly indicates that the term 'text version' is usually assumed to denote a version of the text in its entirety. From a chronological point of view, a version of a novel (on which the author has worked for more than a year) has a meaning different from that of a poem (which may have taken the author only one day to write). For this reason, Van Hulle argues that it would be preferable to work with smaller units such as the version of a chapter, paragraph, sentence or even a word (2019, 473; 2022, 197). According to Van Hulle, the 'sentence version' is useful for writing processes logged with a keystroke logger: authors can work on a sentence as long and hard as they like, but if the revisions stay within the boundaries of the sentence, it can still be regarded as a single version of the sentence (2019; 2021). When an author decides to start working on other parts of the text, the first sentence version is completed; when the author returns to the sentence at a later stage, this creates a second sentence version. In accordance with digital scholarly editions of analogue documents, which usually offer both a static digital facsimile of a scan and a static transcription, Van Hulle proposes to combine a static transcription with a dynamic visualisation of every

sentence for a digital scholarly edition of a born-digital work: "In this way, a scholarly editor can combine stasis with movement, a transcript of every version and a dynamic (filmic) visualization of all the keystrokes constituting a sentence" (Van Hulle 2021, 236).

Mahlow et al. (2022) also apply the unit of the sentence to allow for a more textual analysis of the writing operations, referring to this output – which is derived from the text history – as the 'sentence history'. Automatically created with THEtool, the sentence history "contains all versions of a particular sentence in chronological order", and it therefore explains its evolution, even when they are not part of the final product (Mahlow, Ulasik, and Tuggener 2022, 'Text history [...]').

Working with sentence versions also seems to provide at least a partial resolution to the issue that Ries addresses with regard to digital versions that have been reconstructed with digital forensic methods, including keystroke logging material: "authors writing with a word processor are in general unaware which traces and variants the writing process leaves on the system next to the document versions they deliberately 'saved' to the storage medium" (Ries 2017, 152). According to Ries, the textual stages that can be recovered can therefore not be regarded as 'versions', as they are not intentionally saved as such. In this regard, Ries refers to Shillingsburg's definition of version, which includes that a version of a work is a specific form of the work that "the author intended at some particular moment in time" (Shillingsburg 1996, 43). According to Shillingsburg, this intention cannot be recovered, "but the sequence of words and punctuation produced by the author at the time he was controlled by that intention is recoverable in the manuscript or in revisions" (45).

As argued by Ries, however, the fragments that can be uncovered with digital forensics, are "not the result of a deliberate decision on the part of the author to *save* the document" (2017, 152). For this reason, '[t]hey have to be considered as unintentional snapshots of the writing and revision process, which raises methodological as well as potentially ethical questions for archivists and researchers" (152). Especially when considering using the concept of sentence versions, however, it is conceivable that these recovered or logged textual units – even though they largely represent instances of the text in unfinished state – will also contain sentences that were intended to be there, and in that exact formulation, at least for that particular moment. Taking it one step further, it is also conceivable that these unfinished or premature textual units are the traces of the author's quest to find the right words to represent the intention of the moment – or that they were written with the intention of spurring creativity. Overall, I believe that snapshots and keystroke logging data are largely reflections of "intentions to do" (Shillingsburg 1996, 35; see Chapter 4.3).

As a starting point for the analyses in this study, I use the session versions of each text. The term 'session version' denotes the state of the text *after* a logged writing session.

In most cases, the sessions reflect the author's choice to stop writing, and they therefore provide an indication of the working method of the author in question. Given that several writing sessions can take place on a single day, it is possible for multiple session versions to have been written on a single day. This might indicate that the author prefers working in multiple, shorter writing sessions. Alternatively, some session versions could have been the product of very long writing sessions. Although a very long pause could also be interpreted as demarcating a session and subsequently a version, the nature of the pause remains unclear. Just as the reason the author starts and stops multiple writing sessions in a single day, it is unclear what the author did in the meantime. If a session contains many long pauses, it says something about the author's specific working method as well. Likewise, a considerable amount of time may lapse between the session versions if the author had decided not to work on the text for a long period. The session versions may also contain different versions of smaller units of the text. Recall that session versions always represent a specific text produced so far, which changes as the author works towards the first version of the complete text, such that they have the 'identity' of a 'work in progress'. Session versions therefore seem an appropriate starting point for an investigation of the writing process.

3.3 DIGITAL DOCUMENTS AND WRITING PHASES

With regard to versions of a born-digital work of literature, Van Hulle also considers "whether a version needs to be defined in time, which would mean that a version is linked to a writing 'phase' (or writing 'stage'), rather than to a writing 'layer'" (2022, 195). In turn, this raises the problem of what constitutes a phase in a digital writing process (195). Are writers who write digitally still inclined to distinguish different phases in their writing process? Alternatively, if they are not consciously involved in dividing their work into phases, can phases still be detected in this flow of writing? Before a tentative answer to these questions can be formulated, it is necessary to establish how phases have been distinguished in analogue material.

De Biasi distinguishes several phases and stages that can be found in textual genetic material. According to De Biasi, the genetic dossier will usually show four main phases: pre-compositional, compositional, pre-publication and publication (De Biasi 2005; De Biasi and Wassenaar 1996). Each of these phases can be broken down into several processes and functions, and each is related to individual types of documents (De Biasi and Wassenaar 1996, 31). As De Biasi notes, this typology "should be seen not as the depiction of the actual genetic unfolding along the axis of time, but as the abstract diagram of the logical links that allow us to name and classify genetic documents relative to their function" (36). Three of the phases De Biasi distinguishes belong to the *avant-texte* – documents designating successive moments of the writing process before publication: the pre-compositional phase, the compositional phase and the pre-

publication phase. These sequences correspond to one or several partial processes (from the perspective of the pre-textual phase) and to one or several operational functions (from the perspective of the operational functions) (42).

First of all, the *precompositional phase* relates to a "provisional, exploratory and preparatory *process* whose *operational functions* are: orienting, exploring, decisionmaking, conceiving, and initial planning" (De Biasi and Wassenaar 1996, 42). This phase designates the start of the writing which prepares for the actual writing (Grésillon 1994, 245). Second, the *compositional phase*, is "a *process* of structuration, documentation, and composition whose *operational functions* are: structuring, researching, and textualising" (De Biasi and Wassenaar 1996, 42). This phase is the heart of the genesis of the text in which the text is set up (De Biasi 2005, 39; Grésillon 1994, 245). Third, the *prepublication phase* is "a postcompositional *process* whose *operational functions* are: adding finishing touches and preparing for publication" (De Biasi and Wassenaar 1996, 42). The prepublication phase includes the final reworking of the text, and it results in a clean copy or typescript (Grésillon 1994, 245). The chain of the aforementioned "partial processes" relates to a series of stages that constitute the *avant-texte*. After analysing the documents according to their operational functions and placing them in one of the three phases, the stages offer concepts for a generic classification of the results in their logical concatenation (De Biasi and Wassenaar 1996, 38).

As they are present in the documents, phases "connect the writing space and the writing gesture with the content (invention and organization) and verbal expression" (Mahrer et al. 2015, 154). They therefore belong at the intersection between the writing process and the material traces. In the section below, I examine whether these phases – or other manifestations of phases – can still be distinguished in the digital genetic dossiers of the five authors. First, however, I provide a brief outline of the contents of these genetic dossiers.

3.3.1 The genetic dossiers within this study

In this study, I examine the digital genetic dossiers of Gie Bogaert, Jente Posthuma, Roos van Rijswijk, David Troch and Ellen Van Pelt. As indicated above, my approach is to focus on the method of keystroke logging for genetic criticism, meaning that I also consider only the data provided by Inputlog: the keystroke logging data in idfx-files (and the associated analyses offered by Inputlog) and the Word documents. As a result, the genetic dossiers consist mainly of such files, and the exact quantity is described below. Inputlog provides a Word document at the start and at the end of each session. In most cases, however, the start document of a session coincides with the ending document of the preceding session – provided nothing was written in the document without the software being enabled. If there were no changes between these documents, the starting document of the session will not be included in the genetic dossier. Any additional

material provided by the author (e.g. notes and photographs of printouts) is included as well.

Gie Bogaert wrote *Roosevelt* in 266 days – from July 2013 to December 2015 – during which 447 writing sessions took place, each logged with Inputlog. Of these sessions, 422 were dedicated to the writing of the novel. Luuk Van Waes assisted Bogaert in installing Inputlog and was available for questions or when problems occurred. The genetic dossier consists of the 422 writing sessions that resulted in 453 Word documents (session versions) that show the gradual evolution of the text, along with 277 hours, 14 minutes and 22 seconds of keystroke logging data.[49] In addition to this digital material, the dossier includes the Atoma notebook (as a PDF file), which consists of more than 100 pages in which Bogaert prepared the novel.

The genetic dossiers of the short stories, which had been logged for the Track Changes project, are obviously less substantial. With the exception of Ellen Van Pelt, all of the authors wrote their short stories on laptops that they were able to borrow specifically for this project. Floor Buschenhenke was available for assistance, troubleshooting and questions about Inputlog during the period in which the authors were writing their stories.

Jente Posthuma wrote "En daarom haten ze zichzelf" in 29 days between early August and late October 2020, resulting in 38 logged sessions. The genetic dossier consists of the keystroke logging data, with a total duration of 64 hours, 41 minutes and 55 seconds, along with 38 Word documents.[50] The genetic dossier also includes a Word document with a revised version of the text based on the comments given by an editor of the journal *De Gids*.

Roos van Rijswijk wrote "Zorgvlied" in eight sessions across five days in December 2020. The genetic dossier consists of seven writing sessions – one during which no actual writing actions were logged – with a total duration of 11 hours, 48 minutes and 1 second. The dossier also includes seven Word documents.

David Troch wrote "Mondini" in eight days – from mid-August until mid-September 2020 – during which he logged 30 writing sessions. Only 20 sessions are part of the genetic dossier, however, as the remaining 10 sessions were 'false starts' caused by an error in Inputlog. The genetic dossier thus comprises 20 Word documents and the keystroke logging data, with a total duration of 27 hours, 48 minutes and 24 seconds.

---

[49] In some sessions, the initial document did not correspond to the ending document from the previous session. There are therefore more session versions than there are sessions.

[50] This description of Bogaerts writing process also appears in Bekius 2021a. During the logging of the writing processes for the short stories, Inputlog also made a copy of the working document in Word every minute. These copies are not included in the genetic dossier, but they did serve as a resource when the logging was ambiguous. For the preservation of the keystroke logging material, these one-minute versions are obviously valuable, and they should therefore be included in the archive of the author in question.

Ellen Van Pelt wrote "Dauphin" in 21 sessions over a period of 13 days between late November 2020 and mid-January 2021. The genetic dossier consists of 18 sessions (excluding 'false starts'), with a total duration of 15 hours, 19 minutes and 26 seconds, along with 18 Word documents. The keystroke logging data are also complemented with two PDF files containing pictures of printouts of the Word documents with handwritten revisions.

Starting from these genetic dossiers, my approach is to search for phases in the Word documents, as a first exploration of the material. At times, the actual keystroke logging data are addressed as well, but the manner in which the keystrokes were actually made visible in the text produced so far are discussed in Chapter 4.

3.3.2 Bogaert's cyclic writing process

The aforementioned phases distinguished by De Biasi can roughly be found in the genetic material of Bogaert. First, Bogaert divides his writing process into two stages. The first part – to which Bogaert refers as the 'creative process' – consists of making notes in a paper notebook.[51] In this stage, he develops the concept and structure of the novel, writes character descriptions and collects additional material. The preparation of the notebook roughly corresponds to the pre-compositional phase, although it also contains notes relating to the compositional phase (e.g. the research he conducted) and the pre-publication phase (e.g. a list of elements to check in the typoscript). For example, to account for realistic descriptions of the public square, Bogaert went to the actual Franklin Roosevelt Square in Antwerp to observe and make notes of everything that caught his eye. The writing of the Atoma notebook must be completed before he can start the second part of the writing process: the 'linguistic creative process', which relates mainly to the compositional and pre-publication phase. During the linguistic creative process, Bogaert relies heavily on the Atoma notebook, which therefore informs the digital part of the writing process.

Bogaert usually starts writing a chapter by typing his notes from the Atoma notebook in the Word document. This seems to correspond to what De Biasi terms the initialisation process, which negotiates the transition to the actual writing (De Biasi 2005, 37). According to De Biasi, when the project has been explored in several preliminary phases, the author tends to go back over old notes to see how the ideas may have evolved, with or without the hope of recycling them. To illustrate how this initialisation process is reflected in Bogaert's Word documents, and how the digital writing process then develops through various phases, I focus on the chapter narrated by the character Fran: the former student of the character Gaard – she had taken his

---

[51] I also address this working method of Bogaert in Bekius 2021a.

literary classes at the university before starting at the art academy – with whom she is having a relationship (which, for Gaard, is extra-marital).

The initialisation process of the chapter narrated by Fran becomes visible in the Word document at the end of Session 264. This document contains the text almost verbatim as it was written in the Atoma notebook. All of the sentences have a telegraphic style:

> Op het plein gebleven. Haar versie van haar 'liefde' met Gaard (die is anders! cf leugen) en over haar kunst. Portretteert ook Faraaz.
> Twijfelt aan haar gevoelens voor G. Moet ze hem dat vertellen gezien de waarheidseed?
> Hij komt om zes uur terug. Verwacht wellicht een antwoord van haar.[52]

This document thus shows only 'retyped', previously written, general sketches for this chapter. It therefore exhibits the close interaction between the Atoma notebook and the writing in the Word document. For this specific section of the novel, Bogaert repeats with this retyping what Hayes and Nash refer to as 'abstract text planning'. It includes conceptual planning in which "the writer proposes ideas for the text without specifying the particular language to be used" (Hayes and Nash 1996, 43).

In addition to retyping some of his notes from the Atoma notebook, Bogaert generally also adds new notes, mostly related to the structure or style of the chapter. An example of such a note was included by Bogaert in the Word document as an instruction for himself: "(nieuwe stem, andere, nieuwe stijl)" (Session 317).[53] This note about the character Fran is aimed at the compositional process, as it gives a direction for the style to be used in this chapter. After the inclusion of such notes, Bogaert seems ready to embark on the compositional process, and he starts to include his research as well. At first, this can still take the form of a note. Bogaert found inspiration for the character Fran in an item in the newspaper *De Standaard* (28 March 2014) about the illustrator Sabien Clement, which can be found in the Atoma notebook. In this newspaper clipping that Bogaert, he had marked and underlined certain sentences that inspired him for her character. Bogaert used Clement's experience with and view on figure drawing for Fran's experiences of drawing, as can be seen in the marking of the following sentence: "Modeltekenen is de ultieme oefening voor het oog. Het is 80 procent kijken en 20 procent tekenen" (Steyaert 2014).[54] Bogaert included these sentences in the Word document as a note "Het is veel meer kijken dan tekenen. Veeleer de kunst van het

---

[52] Translation: "Stayed on the square. Her version of her 'love' with Gaard (which is different! cf lie) and about her art. Also portrays Faraaz. Doubts her feelings for G. Should she tell him given the oath of truth? He returns at six. Might expect an answer from her".

[53] Translation: "(new voice, different, new style)".

[54] Translation: "Model drawing is the ultimate exercise for the eye. It is 80 per cent looking and 20 per cent drawing".

zien!" (Session 322).[55] While telegraphically styled notes are still present in the Word document, Bogaert then gradually starts to textualise them.

This process of textualisation is characterised by transforming the notes into syntactically correct sentences, contextualising them and changing their narration. For example, the notes about Gaard expecting an answer from Fran were still formulated from Bogaert's perspective (e.g. "van haar"), but after Bogaert has rewritten them, they become part of Fran's thoughts: "Gaard verwacht een antwoord. Ik heb nog twee uur. Weet niet waar ik aan toe ben" (Session 322).[56] Once the notes have been transformed into proper sentences, Bogaert also starts revising them – mostly at the level of the word. More elements (e.g. from his research) can still be inserted into the text at this point of the process. Now, instead of preserving the identity of the notes, they are mostly incorporated directly as proper, syntactically correct sentences. For example, for Fran's chapter, Bogaert included more of his observations of the square. The observation from the Atoma notebook "Iemand draagt een brug op zijn neus. Gevochten. Pleister"[57] is implemented in the text as an external feature of someone observed by Faraaz, which is in turn noted by Fran: "Hij kijkt naar de man met een brede brugpleister over zijn neus" (Session 331).[58] Once the text has reached an advanced stage – when there are no more structuring elements to be processed – the complete text is subjected to finishing touches. This is accompanied by the deletion of some sentences, correction of typographical errors and replacement of words with synonyms.

The text in Bogaert's Word documents prove that, at a microgenetic level, different phases can still be demarcated in his writing process. He starts by typing over his notes from the notebook and gradually adds more notes to the text that indicate its direction. The addition of notes then takes place simultaneously with the formulation of previous notes. If necessary, research is included and textualised. Such textualisation is interspersed with revision of the text produced so far. After there are no longer notes present in the text – and after the text is written entirely from the perspective of the character – Bogaert tends to return to it a few times to make revisions. At a macrogenetic level, this same working method can be found in the writing process of every chapter (at least to a certain extent), as clearly shown in the visualisations in Figure 3.1, Figure 3.2, and Figure 3.3. In line with macrogenetic research, which "looks at one or several complete collections of genetic documentation, studying large-scale phenomena" (De Biasi and Wassenaar 1996, 27), these macrogenetic visualisations of Bogaert's writing process show which sections of the novel were altered during each of

---

[55] Translation: "It is much more about looking than drawing. Rather, it is the art of seeing!"
[56] Translation: "Gaard expects an answer. I still have two hours. Don't know where I stand".
[57] Translation: "Someone is wearing a nasal strip on his nose. Fought. Band-Aid".
[58] Translation: "He looks at the man with a wide nasal strip on his nose".

Figure 3.1: Visualisation of a part of the macrogenesis (Session 267–323): Indicating the sections on which Bogaert worked during each session.

Figure 3.2: Visualisation of a part of the macrogenesis (Session 324–380): Indicating the sections on which Bogaert worked during each session.

94
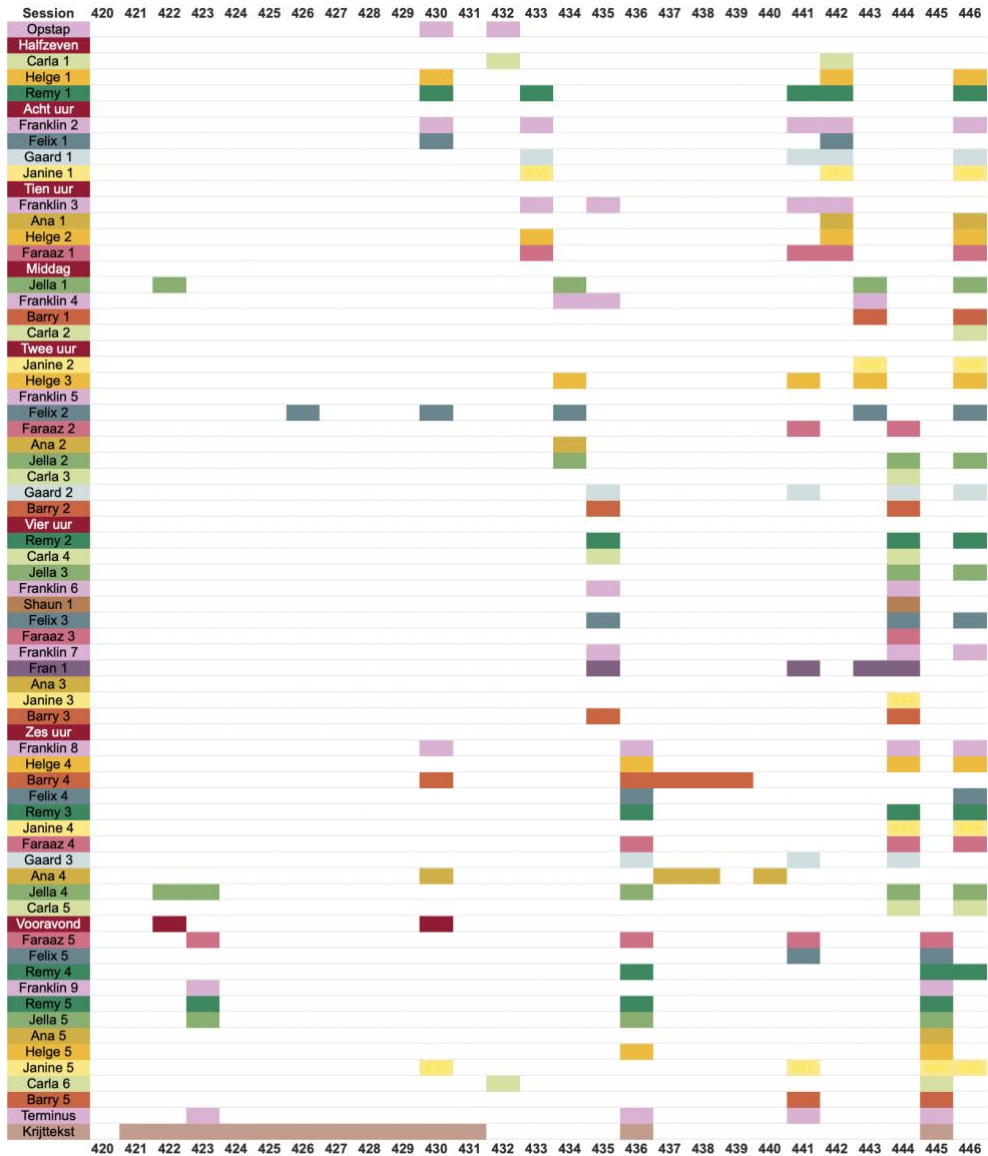
Figure 3.3: Visualisation of a part of the macrogenesis (Session 420–446):
Indicating the sections on which Bogaert worked during each session.

the writing sessions. On the vertical axis of the table, the chapters narrated by one of
the 13 characters are listed in the order in which they appear in the published novel,
starting at the top with the first chapter, "Opstap". The horizontal axis represents all
writing sessions. This visualisation reveals another characteristic of Bogaert's working
method: he worked on the characters in chronological order. In Session 247, he
prepared to write the sections in the overarching chapter "Vier uur" ('Four o'clock') by
inserting notes for each of these sections. He then began to elaborate one of these

sections at a time, in the order in which they were listed in the Word document for that moment: starting with the chapter on Remy, then those of Jella, Frankin, Shaun, Carla (the location of this section was changed later), Felix, Faraaz, Franklin again, Fran, Janine and Ana. The separate blocks indicate that Bogaert made some adjustments or notes for previous or upcoming chapters as well, although these adjustments were often less substantial. After all these sections had been written, he returned to each of the sections in Session 339 and 340 to make revisions as needed. He then repeated this process for the sections in the overarching chapter "Zes uur" ('Six o'clock'). Towards the end of the complete writing process, the pre-publication phase is clearly visible in the visualisation: from Session 432 to 436, he systematically went through the complete text to make revisions, and again from Session 442, while he had made some additional changes.

Different phases are thus present in the genetic material from *Roosevelt*. For Bogaert, it seems that writing this novel was a cyclical process, in which he repeated the phases from inserting notes to the final revisions for each section within each chapter. This raises questions concerning whether the presence of the different phases in his writing process is not just the result of Bogaert's very specific writing strategy. Although short stories may require different writing methods in general, because they are less lengthy, phases can still be distinguished in these writing processes as well.

3.3.3 Idiosyncratic phases in the short stories

The session versions of the short stories indicate that phases can also be demarcated for these writers, albeit more for some than for others. In this section, I discuss some phases visible in the writing processes for the short stories, starting with the writers who relatively planned the most. The writing processes of Posthuma and Troch both reveal a clear pre-compositional phase. Posthuma starts the first session (Session 0) with writing notes – "Aantekeningen" – in which she describes certain elements that she plans to incorporate in the story. Many sentences start with "Hoe" ('How') and "Over" ('About'), including "Hoe machteloos ze was"; "Hoe zij mensen haat"; "Hoe alert ze is op de haat van anderen jegens haar, vooral die van haar reisgenoten"; "Hoe a-relaxed ze omgaat met geld"; "Hoe ze gefocust is op het eten en of ze wel genoeg krijgt"; "Over geen begeerte meer voelen omdat ze niet meer het standaard begeerlijke uiterlijk heeft"; "Over de schaamte die ze voelde toen ze zichzelf voor de grap a hot woman noemde"; "Over haar woede toen haar zus haar om 6 uur 's ochtends ophaalde van Schiphol maar in de verkeerde aankomsthal stond" (Session 0).[59] This indicates that these notes have

---

[59] Translations: "How powerless she was"; "How she hates people"; "How alert she is to the hatred of others towards her, especially those of her travelling companions"; "How a-relaxed she is about money"; "How focused she is on the food and whether she is getting enough";
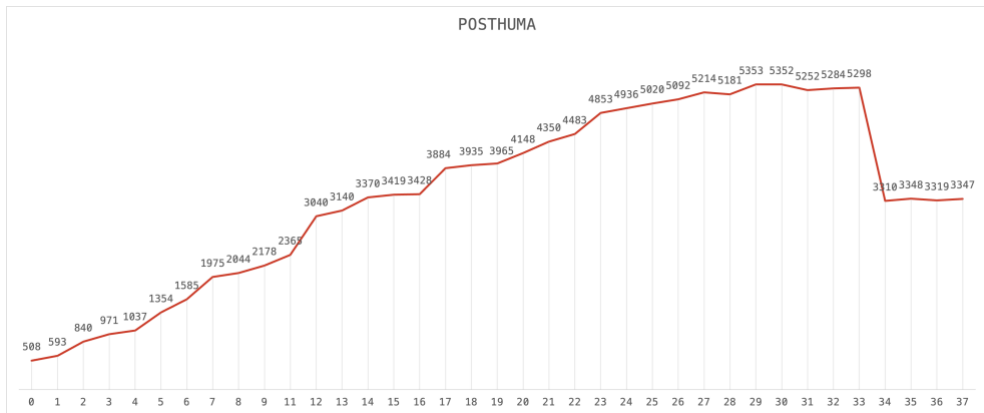
Figure 3.4: The number of words after each writing session of Jente Posthuma.

the function of reminding her of what she wants to describe in the story. After writing these notes, she starts to textualise them, while also incorporating unplanned passages. The notes remain present at the end of the document and are complemented by text elements that she had deleted from the main text but had not yet wanted to discard permanently. While writing, she also includes these writing directions, along with new ones, into the textualised text. Overall, there is constant interaction between textualising, structuring and revising throughout the entire process.

The evolution in the length of the text is displayed in Figure 3.4. In Session 34, Posthuma deletes the notes and the accumulation of unused text fragments. Viewed from within the writing process, this moment marks an important phase: the decision to remove all the notes indicates that Posthuma has decided that the story would not need any of these unused elements.

The writing sessions taking place after these sessions may therefore be regarded as belonging to the pre-publication phase; Posthuma does not add new text and makes only minor, stylistic changes. The pre-publication phase also extends beyond the keystroke logging phase. The story was published in the literary journal *De Gids*. Before publication, the text had been through another editing phase, in which Posthuma added a few new sentences, and – more importantly – a new ending.

For Troch, the planning consists of typing a sentence with a length of 700 words, in which he sketches the concept of the story: a man hiding in a small apartment in a

---

"About not feeling desire anymore because she doesn't have the standard desirable look"; About the shame she felt when she jokingly called herself a hot woman"; "About her anger when her sister picked her up from Schiphol airport at 6:00 a.m., but stood in the wrong arrival hall".
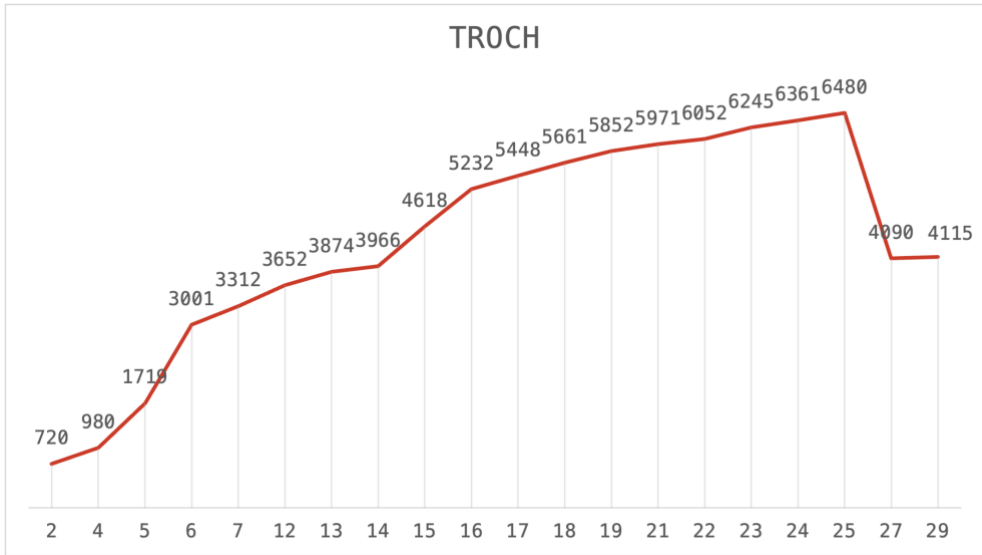
Figure 3.5: The number of words after each writing session of David Troch.

deserted mountain village. It is unclear whether there was a phase before the typing of this sentence – the actual writing of the sentence, perhaps on paper – that is not part of the genetic dossier. This sentence functions as a kind of draft version for the story, and it consists of the basic elements he wishes to incorporate. With this sentence as starting point, Troch then starts to build the story 'from the inside out', repeatedly taking a part of the initial sentence to elaborate it. Most of his new text production also takes place between segments of text that have already been written, and not at the end of the text produced so far. The 'notes' at the end of the text are gradually elaborated with newer text fragments that still demand the writer's attention.

One key moment in Troch's writing process, and one that indicates a more personal phase, takes place after Session 19 (the last session on 21 August). Troch ended this session by writing down the date and time: "21 augustus 2020, 16u20", at the end of the text, but before the notes and unused text elements. Although Troch has never dated his versions up to this point, he eventually dates the final version, thus possibly indicating that, for him, this version marks a specific moment in the writing process. Moreover, the writing actions occurring during Session 19 also seem to suggest that he is working towards a temporary finished version (temporary, as the notes are still included in the document): he goes through the text, more or less linear from top to bottom, in order to make revisions. Six days later, on 27 August, Troch returns to the text and starts by revising at the top, where he includes a 'possible addition': "Mogelijke toevoeging: Zonne-energie. Er zonnepanelen zouden nogal renderen. , maar er is geen
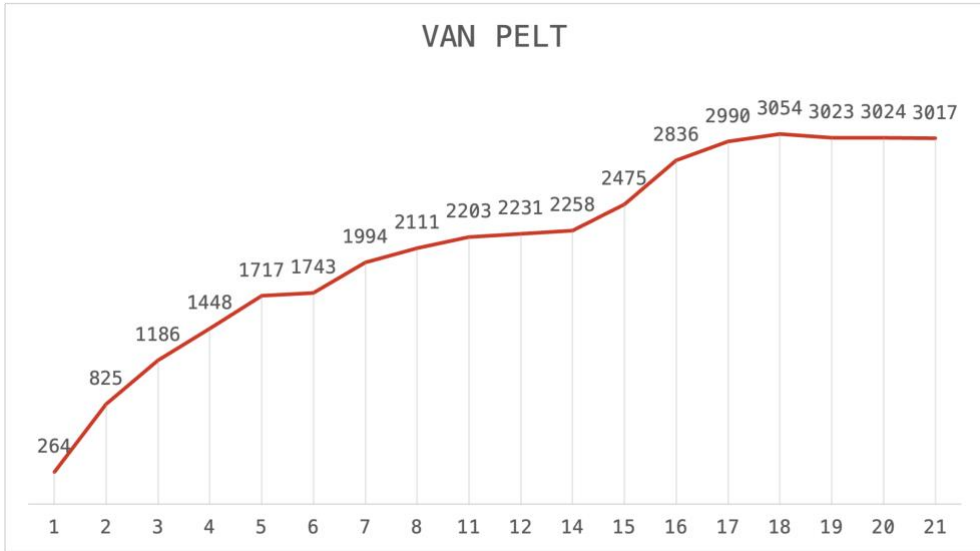
Figure 3.6: The number of words after each writing session of Ellen Van Pelt.

mens meer om ze te bedienen. Alles heeft het begeven" (Session 21).[60] In the following session, on the same day, it becomes clear that Troch is most likely incorporating suggestions from a 'first reader', as one of the documents he has open on his desktop is entitled: "Mondini_[name of first reader]_25082020". Within the writing process, the moment of sending a version of to a first reader denotes a clear phase, a demarcated moment that probably also allowed Troch to re-vise the text. The following, and final, phase for Troch then consists of deleting the notes at the end of the document in Session 27, as can be seen in the drop in the word count. As was the case for Posthuma, this launched the pre-publication phase.

Van Pelt starts her first writing session by writing down some notes. It consists of a preliminary title ("Dolfinarium"), a note on the main character – who may be divorced and is perhaps on vacation with her parents – and describes some possible settings and situations. At the end of the session, she has also included a general outline of the excursion to Flamingo Island. This note, which consists of text taken from a website, will serve as the general structure of the story. In Van Pelt's compositional phase, the operational functions of researching and textualising alternate with each other (see Chapter 5.4). In contrast to Posthuma and Troch, Van Pelt does not collect extensive notes at the end of the document. Moreover, most of the text that she produces during the compositional phase is written at the end of the text produced so far. Although Van Pelt is not an excessive note writer during this writing process, some notes do feature

---

[60] Translation: "Possible addition: solar power. There solar panels would pay off quite a bit. , but there is no human left to operate them. Everything has stopped functioning".

in the writing process. These notes are more reminders and brief ideas, however, than they are elaborate notes on the content. For example, she includes the following note in Session 8:

> Mustafa weer?
> Terugtocht - dolfijnen
> Einde?? -Ander begin? (Session 8)[61]

The note indicates that Van Pelt has not yet planned an ending for her story, and she will not settle for an ending until Session 19. While still indecisive about the end, Van Pelt prints her story to make revisions – after Session 14 and after Session 17. She includes these revisions in the keystroke logging sessions following these offline revision phases, albeit not always verbatim. The offline revision phase also gives her new leads to explore for the end of the story. As indicated by the number of words after each session (Figure 3.6), Van Pelt gradually expands the text, although the number of words remains stable towards the end of the writing process: the moment the text is nearing completion.

Van Rijswijk writes her story primarily as a continuous process of textualization (compositional phase), and the genetic material does not include any evidence of a pre-compositional phase. It may nevertheless have taken place, but it was only not documented. This does not mean that all the text in the Word document has the same function. The text contains elements that mark phases in the creative process: XX's are used for undetermined names, and asterisks are used to mark uncertainties or words that the writer wishes to alter. In some cases, she also inserts notes for clarification, such as:

> […], misschien hebben we allemaal wel een onaards talent meegekregen. (Van wie? Hoezo onaards? Wat is er aardser dan mens zijn, en wat is er menselijker dan zelfs na je dood nog je stempel op het leven willen drukken?*) (Session 2)[62]

Other notes seem to be used as a reminder for 'text to be produced'. Van Rijswijk ends two writing sessions with a quick summary of elements she intends to include in the following writing sessions, as if she is afraid to forget these ideas in the meantime:

---

[61] Translation: "Mustafa again? Return trip - dolphins End?? -Other beginning?"

[62] Translation: "[...], perhaps we have all been given an unearthly talent. (From whom? Unearthly how? What could be more earthly than being human, and what could be more human than wanting to leave your mark on life even after death?*)"
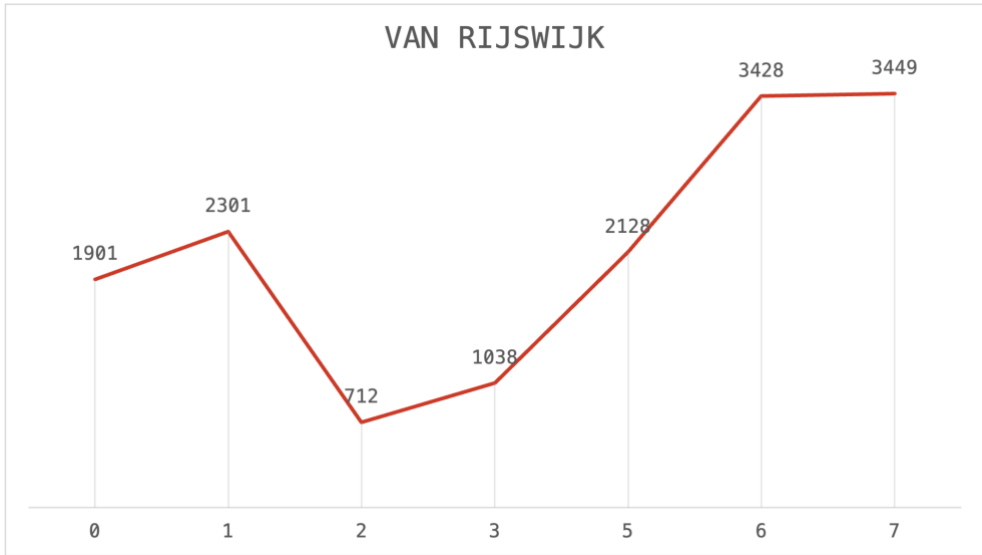
Figure 3.7: The number of words after each writing session of Roos van Rijswijk.

> Mijn moeders schater galmt --- over hele begraafplaats enz enz mooi want ze leeft nog en misschien blijft die galm wel lang hangen omdat de dood m al eens gehoord heeft en bordje verboden roken enz (Session 0)[63]

> [.....]/haar leven/dat ze mee mag lopen/neemt steeds meer mensen mee/niet als ik medicijnen slik. (Session 3)[64]

Moreover, the number of words in the session versions also indicate more idiosyncratic phases, which are not general but specific to this writing process of Van Rijswijk. In the first session (Session 0; 3 December 2020), she writes a first draft of 1901 words. The draft takes place within the same context – the Zorgvlied cemetery – but the essence of the story (i.e., that the I-narrator can communicate with ghosts) is not yet included. In the second session (Session 1), Van Rijswijk begins by revising the draft of the previous day. This revision process at the start of the writing session is most likely part of the process of rereading the text produced so far. After 54 writing actions, consisting predominantly of contextual revisions, Van Rijswijk copies the first five sentences of the draft and pastes it at the top of the document. From there, she starts to write a new beginning. The third session (Session 2) thus demarcates a new phase. The 'Focus

---

[63] Translation: "My mother's laugh reverberates --- all over cemetery etc etc beautiful because she is still alive and maybe that reverberation will linger for a long time because death has already heard m once and sign no smoking etc".

[64] Translation: "[.....]/her life/that she is allowed to walk along/is taking more and more people along/not if I take medicine".

events' (switches between different 'windows' and apps) show that Van Rijswijk starts this writing session by opening a new file in the 'Kladblok' app. She then cuts a large part of the draft, probably to paste it into this new file. This makes space for her to re-conceptualise the story and introduce the aspect of talking to ghosts.

As demonstrated in these examples, phases can still be distinguished in digital writing processes, starting with the examination of session versions, and they manifest themselves in the various functions of the text. For some of these writers – especially for Bogaert – the phases correspond to some extent to the phases distinguished by De Biasi. Moreover, a pre-publication phase can be observed for every author. The analysis also deviates from De Biasi's typology, however, as all the phases can be accumulated within the same document. The phases that can be observed in these born-digital writing processes are even more oriented towards the process. They indicate certain moments in the writing process (in some cases, highly temporal) in which the author has the urge to include possible ideas or writing directions, by way of a reminder, and hesitations about which direction to take or which word to choose.

3.4 CONCLUDING REMARKS

Whereas it has always been possible to constitute genetic dossiers with physical documents (Van Hulle 2014; Shillingsburg 1996; De Biasi 2004), genetic criticism must now deal with digital documents. In themselves, these documents are no more than binary code – 0's and 1's – and metaphors are used to make sense of them. For example, the word 'document' is used, and the GUI renders the file as a sheet of paper. Although the files obviously do not come into being through the use of metaphors only, the metaphor shapes the ways in which we interact with the files. As noted by Drucker, digital files have a distributed materiality. They come into existence only with the use of hardware and specific software (Drucker 2013). While digital documents seem to be immaterial, they actually rely on physical objects. More specifically, they come into existence through the peripheral devices (e.g. keyboard and screen) and internal components of a computer. In addition, the software makes the files accessible and human-readable. Moreover, digital objects are material, as they afford certain actions. The ways in which authors make use of the affordances of word processors or other devices are discussed in subsequent chapters. The two types of materiality distinguished by Kirschenbaum – forensic and formal – address this distributed materiality as well. Forensic materiality denotes the actual traces on the hardware (e.g. the magnetic poles on a hard disk). Formal materiality refers to the multiple perspectives that different types of software can offer on a specific digital file: while one type of software exposes one layer of the data object, another type of software could expose another. Kirschenbaum (2008) also distinguishes three common assumptions about electronic texts: these

assumptions have to do with the supposed ephemerality, fungibility, and fluidity of electronic texts.

If these common assumptions about electronic textuality are rejected and if the different levels of digital materiality are taken into account, new horizons for textual scholarship with regard to born-digital texts appear. The application in which the files are opened can thus expose different views on the text, which may also expose different layers. Questions concerning the 'originality' of the file are resolved in part by the fact that a file cannot be modified without leaving a trace. For the practice of textual scholarship, and particularly that of genetic criticism, the key is to understand that digital texts are more durable than initially thought.

All the observations discussed above are important to contextualise my current study, not in the least to establish its boundaries and limits. Given that this study does not employ digital forensic methods, it focuses solely on formal materiality – and, more specifically on the text rendered in the document displayed in the GUI of MS Word – while being aware that this may change when opened on a different device. My focus is nevertheless not on how the text is actually presented (e.g. how much text was visible on the screen on the moment the author made a revision). Although this is important contextual information for the genesis of a text, my study can address which characters were typed *and* in what order, as well as which other 'windows' were opened during the writing session. For the field of genetic criticism applied to *born-digital literature*, this in itself is a leap forward. While definitions of text that include material aspects are important, Shillingsburg's definition of a text as being the actual order of words and punctuation suffices for the context of my study. In fact, the *order* of the words and characters typed during the process are the focal point of investigation in my study.

For other contextual and material information (e.g. with eye tracking or actual screen recording), the research would have become too intrusive (see Chapter 2). The same would have been the case if the authors had been asked to apply forensic methods to their computers, thereby resulting in a different study, with less connection to the field of writing studies. In general, however, and for the purposes of future research, it is important to be reminded that all levels of materiality – from hardware to forensic materiality and, obviously, all forms of formal materiality – can potentially inform geneticists on how the born-digital work came into being. What the geneticist is able to uncover, however, depends on the content of the genetic dossier.

The authors' ways of working, their knowledge of technology, the technologies that they use and their interest in archiving their files all affect the contents of the digital genetic dossier. These digital documents can then be studied by solely examining their contents, thereby revealing the inter-documentary variation. Some intra-documentary variation could also be revealed by resorting to digital forensic methods – if the original data carrier is also available. If an author is willing to log the writing process with a

keystroke logger, the document can be 'opened up' even more, such that it will show how the content of each document was transformed into the content of the succeeding document. The use of the metaphor 'document' for to refer to the digital files of a word processor and its everyday use has also appeared to create confusion within the context of textual scholarship (Ries 2018). For the sake of clarity, I therefore emphasise that I maintain this metaphorical use of the document. More specifically, in this study I use the term 'document' to refer to *each* Word file in the material recorded with Inputlog.[65] As indicated, each writing session creates two documents (Word files with .docx extension) and one XML file containing the keystroke information (.idfx extension). These documents can then contain different versions of the text.

In the discussion about how to define a version, it becomes clear that a version is an editorial construct, which requires explicating the size of the textual unit to which the version refers (Van Hulle 2022). My approach is to take the session version as the initial textual unit for my analyses of the writing processes. The concept of the session version allows for the reconstruction of how the writing proceeded during a writing session. The unit of the session is a meaningful scale for analysis that makes it possible to zoom in on a particular smaller unit, as well as to zoom out to examine the writing process on a larger scale. As mentioned above, Van Hulle proposes combining static transcription with a dynamic visualisation of each sentence to generate a digital scholarly edition of a born-digital work. Van Hulle acknowledges that this is limited to the process of writing the sentence version: "One of the aspects it does not capture well is an author's sudden decision, for instance, to jump from the middle of one sentence back to the beginning of the story to change something in one of the first sentences" (2021, 238). In practice, this becomes possible when a writing session is used as a unit for delineating versions. How such a dynamic visualisation of the keystroke-logged writing process is achieved is discussed in the next chapter.

Based on these session versions, general phases in writing processes can be distinguished. While the phases in De Biasi's typology – the pre-compositional phase, the compositional phase and the pre-publication phase – can still be distinguished in the digital documents, the digital document also makes it possible to examine more 'personal' phases of the authors. For Bogaert, the writing of *Roosevelt* was a cyclical process, in which he repeated the phases from inserting notes to the final revisions for each section within each chapter. Posthuma and Troch's planning phases were visible in the Word documents. Although Van Pelt and Van Rijswijk also planned ahead, this was less obvious in the material itself. One of the affordances of the word processor is that it allows text to be written in the document, and deleted from it just as easily. It is interesting to note that both Troch and Posthuma used the space at the end of the text to 'store' their notes, text they had removed from the running text and other loose text

---

[65] This is the same usage of the term 'document' as in Bekius 2021a.

fragments. Further, small notes with a short 'lifespan' (e.g. 'address this or that') were being produced throughout all writing processes. This indicates that, while the text is produced in one MS Word document, the operational functions of the text can still differ: not all the text is written with the intention of being part of the novel or story. In summary, compositional phases, revision, research and text production alternate with each other. The next chapters focus on these dynamics.

CHAPTER 4. RECONSTRUCTING REWRITING

UNVEILING DIGITAL TEXT PRODUCTION AND REVISIONS

Dankzij revisie krijgen verhalen de gelegenheid hun diepere gronden te openbaren, die zij, net als mensen, vaak niet bij de eerste kennismaking blootgeven. Een eerste versie geeft maar één visie op de betekenis van het verhaal, namelijk die van de schrijver in zijn roes. Een oppervlakkige, want koortsachtige visie, die rustig moet worden <u>herbezien</u> op haar merites. Voilà: revisie als hulpmiddel om de betekenis te doorgronden. Wat aanvankelijk leek op een verhaal over verloren onschuld, blijkt later over zelfbedrog te gaan. (Dorrestein 2017, 71).[66]

Revision has been considered an important aspect of writing (Conijn et al. 2022; Sullivan 2013; Lindgren and Sullivan 2006b). As the Dutch author Renate Dorrestein points out, revision is literally a way to 're-vise', a way of looking *again* to a text to uncover its meaning (2017, 71). And the American novelist Richard Powers, for example, states that, as a writer, "[y]ou have to have a great appetite for revision" (Van Wing 2022). Indeed, Hannah Sullivan notices in *The Work of Revision* that contemporary writers "draw attention to their habit of revising" (2013, 1). And when they refer to their revision practices, Sullivan observes, writers tend to use durative adverbs such as 'constantly', 'everyday' and 'all the time' to emphasise the intensity of and their commitment to the (seemingly indefinite) process of revising (2). She regards these remarks of constant revision nonetheless as "an exercise in nostalgia", like "the current fetish for refitted manual typewriters" (269). Sullivan argues that "the association of revision and literary value is the legacy of high modernism and the print culture that nourished it" (2013, 2).

---

[66] Translation: "Thanks to revision, stories have the opportunity to reveal their deeper grounds, which, like people, they often do not reveal on first encounter. A first draft provides only one vision of the story's meaning, that of the writer in his daze. A superficial, because feverish vision, which should be *re*vised slowly on its merits. Voilà: revision as a tool for grasping meaning. What initially looked like a tale of lost innocence later turns out to be about self-deception".

She states that Modernist writers revised overtly, "not for stylistic tidying-up but to *make it new* through large-scale transformations of length, structure, perspective, and genre" (2). Her historicist, comparative approach leads Sullivan to argue that revision is – more than initial composition – affected by writing technology and medium, since it requires interaction with something that has already obtained a material form (8). She points out that authors in the digital age already "see their work in something closely resembling its final printed form from the first draft in Microsoft Word", this in contrast to the Modernist writers, who would see their work in different instances: from longhand, to typescript, to proof, to several editions (8). The focus of Sullivan lies on "laborious, belated, even otiose changes, made without reference to the linguistic 'felicity' or basic communicability of the original version" (15) by Modernist writers and states that writing on a computer makes this "belated, laborious revision less likely" (256). Her claim is that, while digital culture makes revision easier, it will also make it less likely and necessary, because "no text is ever really fixed" (257). Yet, in the same way, we could say, the word processor is also the stimulus for the 'constant' revision some writers refer to (see also Chapter 5.1).

Sullivan acknowledges that to prove or disprove such a hypothesis would require a comparison of born-digital and pre-digital archives and an investigation of the saving habits of authors (257). David Stephen Calonne stated in 2006 that since "the computer is still a relatively recent form of writing technology, we do not yet have a complete understanding of the ways it has altered creative writers' revision methods" (2006, 163). To date, attempts to address questions about digital revision have been hampered by the difficulty of accessing digital literary archives. Yet, Sullivan questions whether the materials in these archives "will prove as rewarding to study […] as the interlined manuscripts, typescripts, and proofs of twentieth-century writers" (258). For Sullivan, only the "graphic fossils of thinking in action" can give us information "above and beyond the change itself", as they also point towards *why* a revision was made (258). Even a Microsoft Word document with complete Track Changes history would not suffice (258).

Nonetheless, digital revision has received considerable attention for decades in the field of writing studies and this research has benefited from the advances of keystroke logging. Compared to these studies, Sullivan's focus on *belated* revision – though very common in textual criticism – is rather narrow, especially in the context of digital literary writing. In writing studies, one of the most cited definitions of 'revision' is that of Fitzgerald, who describes the activity of revision as "making any changes at any point in the writing process" (1987, 484). Genetic criticism therefore still seems to lack a thorough understanding of *digital* literary revision in its broadest sense. Using keystroke logging to record digital literary writing processes helps to fill this gap, since "keystroke logging allows revisions undertaken during a writing session to be viewed in the context

of their occurrence" (Lindgren and Sullivan 2006a, 158). Therefore, if we can reconstruct the entire composition process based on keystroke logging data, can we then also find revisions that supersede stylistic felicity? And perhaps even indications of the reasoning behind these revisions? As I illustrate throughout this study, important digital revisions are not necessarily belated but take place at every stage in the composition process, especially in what Sullivan terms the "continual textual present" (2013, 268). But what, then, occurs in this textual present that initiates revision? These questions are central to this chapter, as I focus on reconstructing the writing process based on the keystroke logging data and what we can infer from these reconstructions.

In this chapter, I first compare the approaches of the study of revision within writing studies on the one hand, and textual scholarship and genetic criticism on the other. Since the process-oriented properties introduced by writing studies will prove valuable for genetic criticism, they will also provide the basis for the encoding of keystroke log data discussed in the second section. Ultimately, revisions – as well as regular text production – need to be brought 'back to the screen' in order to study textual genesis in detail. The second section will thus describe the foundations of the TEI-XML encoding of the keystroke logging data within the textual context, to enable dynamic visualisations of the writing processes. Based on these visualisations, I will dedicate the last section to the question 'why?', specifically with regard to revision. What, in other words, can the keystroke logging data tell us about the author's intentions during the writing process; can we say something about the cause of the revision? The keystroke logging data gives us more information about the writing process on which we can base our hypotheses on the triggers of revisions.

4.1 DIGITAL REVISION

As Renate Dorrestein point out, stylistic revision is still only one aspect of the revision process: "revisie van stijl – hoe zeg ik het zo elegant, oorspronkelijk en efficiënt mogelijk – is maar één onderdeel. Revisie van structuur – hoe zeg ik het op een dramatisch effectieve en coherente manier – haalt al heel wat meer overhoop" (2017, 70).[67] But what is revision? Conijn et al. (2022) define a revision event "as an external revision, which starts from a deletion or insertion and ends with normal text production unrelated to the revision or a new external revision" (5). Text production is related to the revision when it is used to replace the deleted text. They further delineate the revision as follows: a revision starts when 1) the writer makes a deletion anywhere in the document, or 2) the writer relocates the cursor to produce new characters; and a revision ends when 1) the writer starts a new revision event, or 2) the writer continues

---

[67] Translation: "Revision of style-how do I say it in the most elegant, original and efficient way possible-is only one component. Revision of structure-how do I say it in a dramatically effective and coherent way-overturns a lot more".

with normal text production that is unrelated to the revision (Conijn et al. 2022, 5). Based on a set of properties, the revision events can be described more extensively.

The taxonomies for analysing online revisions, as developed within writing studies, can be divided into two classes: product-oriented taxonomies and process-oriented taxonomies. Product-oriented taxonomies analyse the effect of a revision on the text. The operation (deletion, insertion, substitution), the syntactic level (word or sentence) and the object (spelling, content, or text organisation) of the revision are used to classify the revision. Process-oriented taxonomies analyse the revisions as they occur in real time and categorise revisions "according to the time and place where they occurred within the writing process" (Lindgren and Sullivan 2006b, 41). Conijn et al. (2022) provide a comprehensive overview of features of digital revision, taking into account the categories of revision discussed in previously developed taxonomies. In their product-oriented *and* process-oriented tagset, Conijn et al. distinguish ten categories of properties: processing, trigger, orientation, evaluation, action, linguistic domain, spatial location, temporal location, duration, and sequencing (2022, 3).

The property 'processing' relates to the different modes in which revision can take place: internally and externally. Internal revisions take place mentally – before the text is transcribed on paper or with the keyboard – and external revisions are visible in the text produced so far (Conijn et al. 2022, 5). The internal revisions, in turn, can also be sub-divided into pre-linguistic and pre-text revisions. The type of pre-linguistic revisions entails mental revisions of non-linguistic representations, mostly conceptual. The pre-text revisions are revisions "where the writer 'rehearses' a formulation in the head before it is transcribed into text" (Lindgren and Sullivan 2006b, 38). For the processing property, Conijn et al. distinguish between these pre-linguistic and pre-textual internal revisions, and external revisions. In their study, Conijn et al. did not annotate the processing property, since they did not use think-aloud protocols that would allow for the annotation of internal revisions. In the same way, internal revisions can also not be studied in this current study of literary writing processes. This property will therefore also not be discussed in further detail.

Four of the other revision properties can be seen as mostly process-oriented properties – spatial location, temporal location, duration, and sequencing – and another four mostly as product-oriented properties – orientation, evaluation, action, and domain. The trigger property lies in the middle of this process-product continuum, and is used to label the cause of a revision (2022, 6). A revision, for example, could be triggered by reading or evaluating the text produced so far (Tillema et al. 2011). Several studies make the distinction between error-triggered and non-error-triggered revisions (Stevenson, Schoonen, and de Glopper 2006; Leijten, Van Waes, and Ransdell 2010; Quinlan et al. 2012). According to Stevenson et al., "[m]any language revisions are likely to be triggered by an error" (2006, 211). But they add that this is not always the case,

since "non-error-triggered language revisions may be prompted by considerations such as style, tone, and cohesion" (211). However, as Conijn et al. point out, triggers happen almost always internally (mentally; not externalised through typing) and cannot be observed in the keystroke logging data. The result of the revision could, however, indicate the trigger (Conijn et al. 2022). Since this entails interpretation, for which we thus need to study both the textual product and the writing process, the trigger property of revision will be addressed in a separate section at the end of this chapter.

The other four process-oriented properties and the four product-oriented properties of revision outlined by Conijn et al. are discussed below and, where possible, related to the way in which revisions are analysed as they occur in analogue drafts and manuscripts. As such, I examine how these two approaches – writing studies on the one hand, and textual scholarship and genetic criticism on the other – can complement each other in the analysis of revisions in keystroke logged literary writing processes. This provides the foundation for my reconstruction and analysis of the writing processes.

4.1.1 Process-oriented properties

Temporal location

The temporal location is the "point in time when the revision is made" (Conijn et al. 2022, 13). This can be within drafts as well as between drafts. For example, Monahan (1984) describes several temporal locations between drafts: "prewriting stage, during the first draft, between drafts, during the second draft, after the second draft" (Conijn et al. 2022, 13). Alternatively, Conijn et al. include one feature related to the temporal location in their tagset: "the time (in milliseconds) until the revision from the start of the writing process" (14). This can then be used to investigate when in the writing process this revision took place, for example, in the beginning, the middle or the end of the writing session. Whichever approach is chosen, the exact time of revision – or any other writing action – is always clear with keystroke logging.

In genetic criticism, describing the temporal location of revisions between drafts has always been one of the primary concerns of the methodology, since all the documents related to the genesis of the work need to be collected and ordered before one can begin a close analysis of the documents. As we have seen in the previous chapter, Pierre-Marc de Biasi distinguishes between four phases in genetic documents: the precompositional, compositional, prepublication and publication phase (De Biasi and Wassenaar 1996). By linking documents to these phases, the general temporal location of the revisions on these documents can also be determined. Thus, the revision can be linked to a particular part of the writing process: the beginning, the middle or the end of the process. Within documents, different writing tools or dates may indicate different stages of writing. However, establishing the temporal location of revisions within analogue documents is more difficult than within keystroke logging data. To

some extent, the temporal sequence of local changes may be recoverable, as Desmond Schmidt points out that "[a]uthors provide many cues such as the position of inserted text, deletion marks, sense, the carrying over of changes into revised blocks, etc. that make this possible" (2019, §65). Nevertheless, the exact order of revision may not be clear, at least not in the way that keystroke logging provides.

In writing on paper, the document pages bear witness of textual genesis through *intra*documentary variation (Schäuble and Gabler 2018, 165). Besides this stratification visible on the page, a large part of the textual development also happens *off the page*, for example in the rewriting of a text (165). This kind of *inter*documentary variation can only be made visible through collation. To account for these differences, Schäuble and Gabler proposed a distinction between textual *layers* and *levels*. Textual layers represent the intradocumentary variation – the revisions made to a single document – while textual levels describe the interdocumentary variation – the differences between two documents (169). For non-keystroke logged digital writing processes, interdocumentary comparison would, in most cases, be the only way to analyse revisions – although, as we have seen, digital forensic methods can help to reveal the revision history.

When using Inputlog, the author can continue writing in a single document throughout the writing process, while intermediate versions are saved as separate documents in the background. If these are logged consistently and without error, the keystroke logging data includes all intradocumentary variation, which in turn provides the information about the interdocumentary variation (the difference between the Word documents saved at the beginning and the end of a writing session). The keystroke logging data thus provides more information about interdocumentary revisions than we can extract from analogue material: all interdocumentary variation is captured as intradocumentary variation within the keystroke logging data. When a keystroke logger is used while writing, textual development can no longer occur *off the page* – or at least: when the author keeps their writing process within the confines of their Inputlog enabled computer, and when that process is logged without errors.[68]

Spatial location

Since revisions occur at different locations in the text, the spatial location indicates where in the text the revision was made (Conijn et al. 2022, 12). The point in the text where the writer is currently producing or deleting text is called "the point of inscription" (Lindgren et al. 2019, 346). A specific form of 'the point of inscription' is "the leading edge" (346). Both the point of inscription and the leading edge refer to "a

---

[68] This way in which keystroke logging eliminates the distinction between interdocumentary and intradocumentary variation also appeared in Bekius 2021a. See this same article for a more elaborate example of how keystroke logging data provides more information about interdocumentary revisions than we can extract from analogue material.

fluid location that moves around with writers' text production as writers produce text at the end of the text produced so far or move around to insert, replace or delete text" (346). However, as Lindgren et al. make clear, the difference between the point of inscription and the leading edge is that the leading edge "is restricted to the point in writing where new meaning is being created" – this can occur at the end of the text produced so far, as well as at the end of an insertion where new ideas are inserted – whereas the point of inscription refers to all actions performed in previously written text as well as at the end of the text produced so far (346).

As for the location of the revisions, Lindgren and Sullivan distinguish between 'pre-contextual' revisions – "revisions made before an externalised context is completed" – and 'contextual' revisions – "revisions made within a completed externalised context" (Lindgren and Sullivan 2006a, 159).[69] This distinction relates to two notions of variant that genetic criticism uses to describe two types of revisions: *la variante d'écriture* and *la variante de lecture* (Doquet 2014, 22). The first describes a revision that occurs during writing, the latter denotes a revision in text that has already been written (22).

According to Lindgren and Sullivan, pre-contextual revisions have three distinguishing features: 1) when the revision takes place, there is only externalised text before the place of revision, 2) when the revision takes place, there is no externalised text after the place of revision, and 3) when the revision takes place, the revised text is the last externalised text (2006a, 159). The revisions that occur at the end of the current text – at the end of the text produced so far – are also referred to as revisions "at the point of utterance" (Matsuhashi 1987). These revisions – pre-contextual revisions – occur "when the writer notices and decides that something that has just been or is in the process of being transcribed needs to be adjusted" (Lindgren and Sullivan 2006a, 161). Like revisions in general, pre-contextual revisions can entail altering both form and meaning of the text.

A difference between digital and analogue revision frequently mentioned is the increase of revisions at the point of inscription carried out during revising in a digital environment (Bowen and Van Waes 2020; Mahlow, Ulasik, and Tuggener 2022). Nonetheless, pre-contextual revisions occur quite often in analogue writing processes as well and are generally referred to as *currente calamo* or *Sofortkorrektur* revisions. The term *currente calamo* "literally means *with a running pen*, and signifies a revision made on the fly without much premeditation" (Van Mierlo 2013, 22, emphasis in original). As an example of a *currente calamo* revision, Schmidt discusses a revision in manuscript A89 of the Australian poet Charles Harpur (1813-1868): "[Harpur] started the sentence with "All his", then corrected this immediately to "The whole of his poetry". Later he started writing "his political effusions lack th..." then corrected this to "his political effusions

---

[69] This description of pre-contextual and contextual revisions also appears in Bekius 2021a.

are wanting…" then corrected the last two words to: "are greatly wanting" (2019, §61). Such immediate revisions are carried out at the baseline (Schmidt 2019, §60). This means that the 'corrected' words are written in the running text.

As Van Hulle rightly points out, "it does not seem appropriate to keep using this term from the quill age", yet he argues that "apart from the obsolescence of the name, the concept it denotes is still relevant" (2022, 198). And indeed, as the discussion of pre-contextual revisions makes clear, such revisions are quite common in digital writing. Of course, pre-contextual revisions in digital writing also always occur at the baseline, but the order in which the actions are carried out is not necessarily the same as in its analogue equivalent. As well as revising directly at the leading edge, the writer can also perform a pre-contextual revision after relocating the cursor to a location earlier in the sentence in production, to make a revision before finishing the sentence at the end of the leading edge.

Contextual revisions, as opposed to pre-contextual revisions, are those revisions that are made in the text that has already been written. The writer relocates the cursor "to insert new text or to delete, substitute or rearrange already written text" (Lindgren and Sullivan 2006a, 171). These revisions can take place at any point after writing the text in question, such as immediately after the production, or at a later stage.

In analogue documents, this type of revision can be found in liminal parts (flyleaves; first and last pages), the margins, the left-hand pages that some authors leave blank for revision, interlinear spaces, and patches (Pierazzo 2022). As an example of an analogue contextual revision, we may take an interlinear revision discussed by Wim Van Mierlo. In the autograph manuscript of 'Ode to a Nightingale' Keats made a change "in the seventh stanza from 'the *wide* windows opening on the foam' to 'the *magic* casements'" (Van Mierlo 2013, 21). The word 'wide' has been struck through, and the substitution 'magic' is written above.

Sequencing

Instead of focussing on single revisions, sequencing "describes the relation of a revision to a preceding revision in time" (Conijn et al. 2022, 14). When revisions form a unit or group, these groups are generally referred to as 'revision episodes'. In the case of revision episodes, the revisions are triggered by one another. Kollberg identified three types of revision episode: 1) episode with repetitive revisions at one cursor location, 2) episode with embedded revisions and, 3) episode with a sequence of revisions in previously written text (Kollberg 1998, 85–95). These revision episodes will be further discussed in the third section of this chapter. Conijn et al. include four types of sequencing that could be automatically extracted: repetitive revisions, embedded revisions, sequence forward revisions, and sequence backward revisions (2022, 15). They also added two manual features: 'overrides previous revision', for a repetitive

revision that changes "the same linguistic domain at the cursor location"; and 'continues on previous revision', for when "the previous revision caused the subsequent revision" (Conijn et al. 2022, 15).

In genetic criticism, the term 'revision campaign' is used when the traces of writing suggest that the revisions were part of the same work episode, like, for example, in this description by Van Hulle about a revision Milton made in the draft of his poem *Lycidas*:

> Milton first chose a gruesome image, describing the floating head as a gory scalp ('goary scalpe'). He deleted the passage and substituted it with a less grisly description ('divine head') in the right-hand margin. But then, in a third *revision campaign*, he changed it again on the facing verso page, in three stages: first 'divine visage', then 'gorie visage', and finally 'goarie visage' (Trinity College Library MS R. 3.4, folio 15v). (Van Hulle 2022, xvii, emphasis mine)

The TEI SIG for genetic editions refers to this as a 'stage', which is defined as "a reconstructable stage in the evolution of a text, represented by a document or by a revision campaign within one or more documents, possibly assigned to a specific point in time" (TEI Special Interest Group 2013 quoted in Van Hulle 2013, 24). As Van Hulle points out, this concept usually applies to "the entirety of the work" (2013, 25). For modern manuscripts, however, the sequence of revisions may be hard to discern, especially when the revisions are carried out in the same hand and ink-colour (Schmidt 2019, §3; Van Hulle 2013, 25). As an example, Van Hulle (2013) discusses a revised sentence in the manuscript of Samuel Beckett's *The Unnamable*, which is written and revised in blue-black ink:

> How can you ~~reflect~~ think and ~~speak~~ say something at the same time (HRC MS SB 5-9-3, page 3r., in Van Hulle 2013, 25)

Van Hulle argues that "it is likely that Beckett made the two substitutions during the same stage of revision ("*campagne de révision*")" (2013, 26). However, he points out that it can also not be excluded that "say something" was added some time *before* "think", or that the addition of "think" preceded the addition of "say something" (26). In both these cases the "two separate additions constitute two separate stages" (26). Nonetheless, even for analogue documents it is sometimes possible to discern a sequence of revisions within a single revision campaign.

Brett Barney, for example, experimented with encoding his "hypotheses regarding 'how the page was filled' for a single example manuscript" of Walt Whitman (2020, 17), and came to a plausible writing order. Sometimes the context of the revisions indicates the order of the revisions. Schmidt describes for instance seven changes in one stanza from Harpur's poem "Vestibulary Stanzas" in manuscript B78 written in 1855. For two of these seven changes the temporal order could be distinguished: "It is clear that the

transposition in line 2 preceded its replacement by line 1. Also the replacement of 'blaze' by 'fire' in line 5 preceded the replacement of 'The inward rolling fire' by 'How burns the fire within'" (Schmidt 2019, §22). The remaining changes, Schmidt argues, have to be seen as independent of one another. Schmidt regards a change as "*independent* if it can be carried out without affecting the sense, grammar or metre of the rest of the text" (2019, §24). In his view, then, "a cluster of dependent changes as some point in the text-stream" constitute a *revision site* (Schmidt 2019, §29). Revisions sites are therefore also the result of revision episodes, particularly that of revisions that continue on previous revisions.

The genetic scholar can only hypothesise about the order in which revisions were made and can by no means be completely certain about the order of the revision in relation to the rest of the text. If the writing process is logged, the sequence in which the text was written can be deduced from the keystroke logging data – effectively eliminating the need for this type of guesswork.[70]

Interdocumentary variation – between drafts – textual levels	
Intradocumentary variation– during a draft – textual layers	
Leading edge point in the text where the writer is creating *new meaning*, usually at, but not limited to, the end of the text produced so far.	Point of inscription point in the text where the writer is currently producing/deleting text.
Pre-contextual revisions la variante de l'écriture – currente calamo – Sofortkorrektur e.g., baseline revisions Immediate revisions made during the production of a sentence. ************************************** As I was writing this sentence, I deleted ~~a word~~ two words to be precise.	Contextual revisions la variante de lecture e.g., in margins/interlinear Revisions after the point of writing the sentence ******************** I already wrote a sentence, but I deleted ~~a word~~ two words to illustrate the term.
Revision episode – revision campaign – revision site	
episode with repetitive revisions at one cursor location	
episode with embedded revisions	
	episode with a sequence of revisions in previously written text

Table 4.1: Process-oriented properties of revision and related terms from genetic criticism and textual scholarship.

---

[70] This same observation has been made in Bekius 2021a.

Duration

The last process-oriented property discussed by Conijn et al. is the duration of the revision. Conijn et al. (2022) included two features related to the duration of the revision in their tagset: the duration of the revision itself – by measuring the time between the first key press of the revision to the last key press – and the pause time before the revision – by measuring the time between the last key press before the revision and the first key press of the revision (14). For analogue documents, the duration of the revision is impossible to determine. One may draw up hypotheses on the basis of the handwriting – was it "barely legible, seemingly hurried" (Van Mierlo 2020, 21) or quite regular, "formed in a measured, almost leisurely manner" (22) – but this is in most cases only applicable for fragments of text and not for individual revisions. However, for the analysis of revision within the perspective of genetic criticism, the duration of revisions is also not the primary concern, and I will therefore also disregard it in my further analyses of the revisions.

4.1.2 Product-oriented properties

Action

Generally, revisions are divided into several revision actions, or mechanical operations performed by the writer: "deletion, insertion, substitution, and reordering (also referred to as transpositions, reorganization, and permutations)" (Conijn et al. 2022, 10). Naturally, these adhere to the basic categories distinguished in genetic criticism, where writing events can always be described in terms of additions, deletions, replacements of one segment by another, and relocations of a text segment (Doquet 2014, 22). This thus means – as Sullivan explicates – that the revising writer at the simplest level has three choices: "a) to add material, so the final version is longer than the first draft, b) to delete, so it is shorter, and c) to substitute, producing a first and final draft of similar length" (Sullivan 2013, 15). With regard to the meaning-preserving changes, as distinguished by Faigley and Witte, additions "raise to the surface what can be inferred" and deletions force the reader "to infer what had been explicit" (1981, 403). Within a literary context, the same applies, as Sullivan points out: "If extension makes texts more difficult through comprehensiveness, proliferation, complexity of diction, and parataxis, excision makes texts hard by asking its readers to fill in the gaps" (2013, 109).

Orientation

Revisions are usually classified on the basis of their orientation (e.g. Faigley and Witte 1981; Lindgren and Sullivan 2006b). Faigley and Witte, for example, based their taxonomy on "whether new information is brought to the text or whether old information is removed in such a way that it cannot be recovered through drawing

inferences" (1981, 402). They therefore distinguished between surface changes and text-based (or meaning) changes, which are each again subdivided into several categories. Conijn et al. distinguish for the orientation property between surface and semantic revisions. The surface changes further divided into formal changes (like typos, spelling, grammar, capitalisation, punctuation etc.) and meaning-preserving changes. The semantic changes are further divided into microstructure changes – "adding or removing supporting information, changing emphasis, understatement, coherence, and cohesiveness" – and macrostructure changes – "altering the overall aim and adding or removing entire subtopics" (Conijn et al. 2022, 10).

Within genetic criticism, some of these orientation properties are also described, although they are not termed that way. For example, Kafka's substitution of '*ich*'s by '*K*.'s – thus changing the text from first person to third person narration – in the draft of *Das Schloss* (Cohn 1968) can also be discussed as being a macrochange. And Sullivan's description of revisions to David Foster Wallace's story "Backbone" reveals that she regards these changes as surface changes, meaning-preserving changes in particular: "the changes are, once again, mostly simple substitutions – small nudges toward stylistic felicity. They resemble the kind of edits that a sensitive reader or teacher would prompt" (2013, 262). Or take, for instance, Van Hulle's description of the interdocumentary variant discussed above:

> The word 'viduity' is an *interdocument* variant by means of which Beckett creates a sort of hole in Krapp's memory, a lexical gap in his vocabulary. This gap was not present in the first two drafts (the manuscript and the first typescript), where he simply used the word 'widowhood'. It is thanks to a simple inter-document variant – changing 'widowhood' into 'viduity' – between the first and second typescript (version 2 and 3) that Beckett creates the opportunity for himself to build the whole scene with the dictionary and elaborate on the gap in Krapp's vocabulary, a lexical emptiness or 'viduity' as it were. (2022, 69–70)

From this explication we can ascertain that the revision is seen as a microstructure change that allowed for the addition of supporting information about Krapp.

However, the orientation categories distinguished by Conijn et al. (2022) do not seem meaningful or specific enough in themselves for a genetic analysis, as they do not focus on the characteristics of a literary work. Therefore, my colleague Floor Buschenhenke is developing a narratological taxonomy to classify revisions (Buschenhenke forthcoming). Her tagset includes the following narratological categories: plot, character, setting, time, and point of view. In my current study, I will not classify the individual revisions into specific groups in terms of orientation, but rather analyse how the text (and its meaning) was constructed gradually or focus on the trigger of revisions based on the context or a sequence of revisions. In Chapter 6, this

will entail a narratological approach, as I examine how and when details of aspects of the narrative text – such as the setting, time, and consciousness representation – are integrated into the text and how they change during the writing process.

Linguistic domain

Revisions can also be classified according to the linguistic domain, which focuses on the linguistic properties of the revision. These can focus on large or small domains, such as theme, discourse, paragraph, sentence, clause, phrase, word, subword, morpheme, grapheme, subgrapheme and punctuation (Conijn et al. 2022, 11). In genetic criticism, the revisions are usually not described solely with regard to their linguistic domain. However, the linguistic domain can be a starting point for describing and understanding the revisions or the revision practices of a particular author. Sullivan, for example, states that Auden, "rarely changed more than a single line (most of the revisions are to a word or phrase)" (Sullivan 2013, 244). For a macrogenetic analysis, it may be useful to have an overview of the number of revisions in relation to a particular linguistic domain. However, as I am primarily interested in the development of the text rather than individual revisions, this will not be part of this study.

Evaluation

With regard to the evaluation of revisions only formal revisions have been analysed, since they are necessary to adhere to linguistic requirements (Conijn et al. 2022, 10). In their tagset, Conijn et al. only evaluated revisions of typography, spelling, or grammar since these can easily be evaluated as correct or incorrect. In genetic criticism revisions are usually 'evaluated' with regard to the context of the text. For example, the change that Keats made in the seventh stanza in 'Ode to a Nightingale' from "the *wide* windows [sic] opening on the foam" to "the *magic* casements" add, according to Van Mierlo, subtle depth to the poem: "The windows, after all, are not real but imagined windows, recollected from an auditory vision of the nightingale's song in 'fairy lands forlorn' (Keats 1978: 371)" (2013, 21). Van Mierlo states that it is therefore a reasoned change, which he contrasts to a *currente calamo* change of 'keelless' into 'perilous' in the final line of the stanza: "This *currente calamo* revision happened almost without interruption, for 'keelless' seems hardly a fitting adjective for the sea. 'Perilous' is very much *le mot juste*" (22). In Van Mierlo's discussion of the revisions, we can clearly spot some reasoned judgements: the first revision is a "reasoned change" because it does not refer to real windows, and for the second revision Van Mierlo argues that 'keelless' is not the right adjective for sea, while perilous – according to him – is.

From the discussion of both the process- and the product-oriented properties of revisions, I conclude that for the analysis of revisions of the keystroke-logged literary

writing processes, the approaches of revision in writing studies and genetic criticism can best be combined. The approach stemming from writing studies introduces a focus on the process aspects of revisions. In genetic criticism, this process aspect has also been of primary concern, but the analogue material – in which the aspect of time is not a given – has withheld a more structured analysis of these temporal revision properties. However, the two fields share a similar approach to the spatial location property of the revisions: pre-contextual revisions relate to *currente calamo* revisions, and contextual are primarily known as the revision in the margin or in-between the lines.

For the analysis of revisions, the tagsets from writing studies are not sufficient for a proper textual genetic analysis since the categories used by themselves do not fully cover the needs of a literary critical interpretation. This suggests that we need the textual context of the revisions in order to arrive at a meaningful genetic analysis of the keystroke-logged revision. Both in time and space. This is the focus of the next section: visualising the revisions captured in the keystroke logging data in context. As a result, we can imagine gaining a better understanding of how ideas emerge, develop, and change during the writing process.

4.2 REVISIONS BACK ON THE SCREEN

Studying the texts' geneses in such detail keystroke logging allows for is only possible when the revisions can be brought 'back on the screen'.[71] Take, for example, the elaboration of the following paragraph in the first chapter in *Roosevelt* narrated by the character Faraaz. At the beginning of the 196th writing session, this paragraph reads as follows:

> Een kelner moet de dingen in de gaten houden. altijd het overzicht houden. En de controle, beheersing. Hij moet beslissingen nemen, ongemerkt. Help me herinneren. Lange groene linnen schort
> Gustav keuken. Gustav noemen.[72]

The sentences in this paragraph are written in a somewhat telegraphic style and seem to have the function of structuring the text. The sentences indicate what should be addressed in this paragraph: the behaviour of the waiter, who should always have an overview of the situation and must make decisions. We also find a reference to the collection of short stories *Help me herinneren* (2012) by Bernlef. The Atoma notebook

---

[71] This section contains revised parts of the following publication: L. Bekius, "The Reconstruction of the Author's Movement Through the Text, or How to Encode Keystroke Logged Writing Processes in TEI-XML", *Variants* 15-16 (2021).

[72] Translation: "A waiter must keep an eye on things. always keep the overview. And the control, the grip. He has to make decisions, unnoticed. Help me remember. Long green linen apron. Gustav kitchen. Mention Gustav."

contains a photocopy of the first pages of Bernlef's short story "Parlez-moi d'amour", which was published in *Help me herinneren*. The story describes the behaviour and appearance of the 'perfect' waiter. During Session 196, in a first attempt of textualisation, these notes and the reference to *Help me herinneren* were worked out into more or less proper sentences – although the text still contains typographical errors. At the end of the session, the text reads:

> Een kelner moet de dingen in de gaten houden. Hij moet altijd het overzicht hebben. En de controle en de zelfbeheersing. Hij moet beslissingen nemen, ongemerkt. Zijn stem is helder en duidelijk maar hij praat *en sourdine*. Zijn handen en zijn vingernagels zijn brandschoon, zoals de witte overhemd dat hij altijd draagt en de lange groene schort die hij heeft voorgebonden en waarop een Portugese naaister in rode letters GUSTAV heeft genaaid. Hij is anoniem en onzichtbaar. Hij beweegt geruisloos. wanneer hij is voorbijgelopen ben je hem al vergeten. Je kan ook niets aan hem merken. Hij kan onzindig [sic] veel kleien [sic] geheimen verbergen. Of één heel groot.[73]

Just by reading the text we can already observe that the second and the third sentence have been slightly revised, that the notes "Help me herinneren", "Gustav keuken" and "Gustav noemen" were deleted, that the note "Lange groene linnen schort" has been textualised into a detailed description of the appearance of the waiter, that is loosely based on Bernlef's short story, and that Bogaert added six new sentences. The variation between the start document and the end document of Session 196 can also be represented by means of a visualisation of the TEI-P5 output of a collation made by CollateX (Figure 4.1). Contextual information allows us to interpret every text segment that appeared in the start document but not in the end document of the session, as deleted text. This can therefore be visualised as strikethrough text in red. On the contrary, every text segment that appeared in the end document but not in the start document can be regarded as added text and can subsequently be visualised as green underlined text.

However, this static visualisation of the collation output does not give any information about how these changes were made; it does not reveal much about the writing process. To be able to analyse all the changes and how they were implemented during Session 196, we should take the keystroke logging data into account.

---

[73] Translation: "A waiter must keep an eye on things. He must always have the overview. And the control and self-control. He has to make decisions, unnoticed. His voice is clear and distinct but he talks *en sourdine*. His hands and his fingernails are spotless, like the white shirt he always wears and the long green apron he has tied and on which a Portuguese seamstress has sewn GUSTAV in red letters. He is anonymous and invisible. He moves silently. by the time he has passed you have forgotten him. You can't notice anything about him either. He can hide an endless number of small secrets. Or one very big one."

Een kelner moet de dingen in de gaten houden. ~~Hij moet~~altijd het overzicht~~houden~~ hebben. En de controle~~, beheersing~~ en de zelfbeheersing. Hij moet beslissingen nemen, ongemerkt.~~Help me herinneren.~~ Zijn stem is helder en duidelijk maar hij praat en sourdine. Zijn handen en zijn vingernagels zijn brandschoon, zoals de witte overhemd dat hij altijd draagt en de ~~Lange groene~~ lange groene ~~linnen~~ schort~~Gustav keuken.~~ die hij heeft voorgebonden en waarop een Portugese naaister in rode letters ~~Gustav~~ GUSTAV ~~noemen.~~ heeft genaaid. Hij is anoniem en onzichtbaar. Hij beweegt geruisloos. wanneer hij is voorbijgelopen ben je hem al vergeten. Je kan ook niets aan hem merken. Hij kan onzindig veel kleien geheimen verbergen. Of één heel groot.

Figure 4.1: Static visualisation of CollateX TEI-P5 output

Although Inputlog is developed for textual and cognitive study of writing, the data output from the writing processes of Bogaert, Posthuma, Van Rijswijk, Troch and Van Pelt, generated with Inputlog, is not immediately suitable for literary textual research. While Inputlog provides a video replay of the recorded writing session, some issues emerge when replaying the writing process. Short writing sessions comprising linear text production are replayed accurately, but as soon as larger segments of text are relocated or deleted, when the writing is characterised with non-linearity, or when the logged session is of considerable length, the replay mode is affected and represents the revisions and text production at the wrong location in the text. Moreover, relying solely on a video replay of the writing session for text genetic analysis also seems undesirable, as one would need to watch a writing session of, say, two hours in its entirety, while constantly pausing to analyse the effect of the revisions. A static reconstruction of the writing session – whether or not in combination with a video replay, as in Dirk Van Hulle's proposal for a "Dynamic Facsimile" (2021) – is favoured to ensure adequate analysis. Hence, in order to be able to study the revisions (contained in the keystroke logging data) in their textual context, the twofold output of Inputlog – the Word document and the keystroke logging data – requires some reassembly.

Within cognitive writing process research, a method has been developed to study revisions in context, including their dynamics: the S-notation (Kollberg 1998). This represents the changes in the text at their location and provides information about the range, order and structure of the revisions (Kollberg and Eklundh 2002, 91). This computer-based notation can be generated using the keystroke logging data from Inputlog and is provided within the Analyse-feature of the software. However, the S-notation was initially developed to visualise revisions of short writing processes in experimental settings. As such, it appeared to be unsuitable for the study of longitudinal literary writing processes logged in their natural setting. Literary writing processes may take up several years and hundreds of writing sessions, with the production of an extensive number of words. As a result, the S-notation could not be generated using the keystroke logging data gathered from the writing processes of Bogaert, Posthuma, Troch, Van Rijswijk and Van Pelt. More generally, the S-notation does not allow for

further annotation and processing. Another problem concerns the representation of deleted text. Since Inputlog logs the position of the event according to its position on the x- and y-axes of the MS Word document, the deleted text is not always presented correctly (the only information the keyboard provides about a deletion is usages of the delete or backspace key). This hinders an automatically generated visualisation of the revisions in their textual context.

The disadvantage of keystroke logging data such as those recorded by Inputlog is that the processes are recorded in such detail that the logs become almost incomprehensible to the untrained eye. To make these logs more accessible to researchers, they need to be presented in a way that captures only the relevant information and conveys the researcher's interpretation of the data in a format that is easy to read and preferably familiar to their peers.

This is exactly the strength of the Text Encoding Initiative (TEI), whose guidelines recommend the use of XML tags to both transcribe the text as it is recorded on the document, and to encode the researcher's interpretation of that record in a human and computer readable format. Since TEI-conformant XML is widely used to create a digital form of humanities data – texts, manuscripts, archival documents and so on – I opted to encode the keystroke logging data in TEI-XML to visualise and analyse revisions in their textual context (Burnard 2014). These transcriptions function as a tool to gain more insight into the textual genesis: to provide transcriptions of the writing processes that could be used to analyse (the sequence of) the revisions and text production in each logged writing session in their location in the text.

`<add>`	`@hand="#GB"`	`@rend="#blueInk"` `@rend="#blackInk"`	`@type="#alternative"`	`@place="#marginleft"` `@place="#supralinear"` `@place="#inline"`
`<del>`	`@hand="#GB"`	`@rend="#blackInk"`	`@type="#crossedOut"` `@type="#overwritten"` `@type="#instant Correction"` `@type="#underlined"`	

Table 4.2: Specifications for additions and deletions in analogue material.

`<add>`	`@seq="201308230815"`	`@type="context"` `@type="pre-context"` `@type="typo"` `@type="nt"` `@type="translocation"`	`@n="15"`	`@evidence="6514-6556"`
`<del>`	`@seq="201308230830"`	`@type="context"` `@type="pre-context"` `@type="typo"` `@type="translocation"`	`@n="16"`	`@evidence="6557-6567"`
`<mod>`	`@seq="201308230840"`	`@type="continue"`	`@n="17"`	`@evidence="6576-6599"`

Table 4.3: Specifications for additions and deletions in digital material.

4.2.1 Encoding Keystroke Logging Data Instead of the Document's Layout

When the text is composed in a word processor, revisions cannot be specified with the attributes used in encoding analogue material. Instead of indicating the specific writing tool (which may be encoded in `@rend`) or the location in the document (which may be encoded in `@place`), digital revisions (specifically: `<add>`s and `<del>`s) can be further specified using their location in the text. These diverging behaviours in the writing of analogue versus digital documents forces us to completely rethink the ontology we use for encoding relative location in our transcriptions. In the following, I therefore propose a list of revision types that first of all focuses on the spatial location of the revisions. With this as a starting point, we can encode the relative location of additions and deletions in the attributes of `<add>` and `<del>` tags in a way that is more relevant to born-digital writing processes (see Table 4.2 and Table 4.3).

With regard to the spatial location of the revisions, we have already seen that a distinction can be made between "pre-contextual" revisions ("revisions made before an externalised context is completed") and "contextual" revisions ("revisions made within a completed externalised context") (Lindgren and Sullivan 2006a, 159). These location-based revision types best resemble the use of attributes like `@rend` or `@place`, as they are based on the relationship between the keystroke logging data and the place in the document, rather than purely on the scholarly editor's interpretation.

In the encoding of the keystroke logging data, these revision types may be applied to the textual unit of a sentence. In the proposed encoding scheme, each sentence is therefore encoded with a `<seg>` tag. Building upon the taxonomy by Lindgren and Sullivan, the attribute `@type="context"` can be used to indicate that the revision is a contextual one, that is: a revision made in a previously written sentence. The attribute `@type="pre-context"`, by contrast, may then be used to indicate revisions made before a sentence is completed and so concerns the author's most recently typed characters. Diverging from Lindgren and Sullivan's definition of pre-contextual revisions, pre-contextual deletions can take place at a point in the text with externalised text after the deleted text.[74]

A large number of revisions in digital writing occur as a result of typographical errors. Within the scope of genetic criticism, such "typos" are not the most meaningful entities because they do not immediately affect the meaning of the text. Within cognitive

---

[74] In the taxonomy by Lindgren and Sullivan, one feature of a pre-contextual revision is that there is no externalized text after the place of revision (Lindgren and Sullivan 2006a, 159). As literary writing is often a non-linear process, new context can be created in other places than at the end of the text. In order to be able to distinguish revision within a sentence before this sentence is completed, I also regard these revisions as "pre-contextual".

writing process research, typos are regarded as a revision type that "often blur[s] the picture of the writing session" (Kollberg 1998, 68). Typographical errors are "low-level, and hence less important, types of revision", and filtering them out would therefore allow for a more nuanced analysis of revision (Conijn et al. 2019, 71). But the revision of typographical errors can also break the flow in writing and therefore influence the writing process (72). For this reason, I propose to encode this type of revisions with a separate `@type` attribute: `@type="typo"`. This allows such errors to be filtered out in visualisations where they are irrelevant, while still allowing us to evaluate their effect on the writing process.[75]

The use of a keystroke logger allows for an exact reconstruction of the textual development. This includes the moment a new sentence is produced. Therefore, the production of new sentences can also be incorporated in the encoding (`@type="nt"`; "new text"), to be able to differentiate between "new" and "old" sentences. Some text segments may also be deleted and added again in a revised form; thereby maintaining a semantic relationship with the previously deleted text. This is not necessarily "new" text and may therefore be given another attribute: `@type="rt"` ('revised text'). However, the encoding of such revised text adds a new level of interpretation to the transcription. Whereas "new text" is not a very subjective interpretation – as the text is typed into the document for the first time – the classification "revised text" relies on the scholarly editor's interpretation and their understanding of the text.

As pointed out, writing is not always a linear process and sentences are not always finished before modifications are performed elsewhere in the text. The author could, for example, move away from the point in the writing where new meaning is produced: the so-called "leading edge" (Lindgren et al. 2019). In the definition formulated by Lindgren et al., the leading edge is located "typically at the end of the text produced so far, but can also occur at the end of insertions within previously written text where a writer inserts new ideas (not only revises form)" (Lindgren et al. 2019, 347). Unlike the point of inscription, which comprises "all writers' actions in previously written text as well as at the end of the text produced so far", the leading edge is restricted to the creation of new meaning (347). During the production of a sentence the author can

---

[75] Typing errors can be hard to distinguish from spelling errors. In the encoding of the typing errors, I therefore used a list of criteria (developed by Stevenson, Schoonen, and Glopper 2006) for distinguishing *typing* revisions from *spelling* revisions. According to the checklist developed by Stevenson et al., a revision can be identified as a typing revision, if one or more of the following applies: "a. the pre-revision form does not conform to the orthographic rules of the language; b. the pre-revision form involves a letter string which does not conform to a likely pronunciation of the word; c. the semantic context indicates that the pre-revision form could not have been intended; d. the same word is written correctly at an earlier point in the text; e. a letter is replaced by an adjacent letter on the keyboard" (Stevenson, Schoonen, and Glopper 2006, 232).

WRITING ACTION	TEXT AFTER THE WRITING ACTION
adds *Soms beiden*	Soms beiden
adds *krijgt h*	Soms **krijgt h**beiden
adds *kan hij meer*	Soms **kan hij meer** krijgt hbeiden
adds *en dan wat hij voor zon' kunstwerkje vraagt*	Soms kan hij meer krijgen dan wat hij voor zon' kunstwerkje vraagtt hbeiden
deletes *n*	Soms kan hij meer krijgen dan wat hij voor zo' kunstwerkje vraagtt hbeiden
adds *n*	Soms kan hij meer krijgen dan wat hij voor zo'**n** kunstwerkje vraagtt hbeiden
deletes *t hbeiden*	Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt
adds *, maar dat il hij niet*	Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt**, maar dat il hij niet**
adds *w*	Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt, maar dat **w**il hij niet
adds the full stop	Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt, maar dat wil hij niet**.**
adds *ooit*	Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt, maar dat wil hij n**ooit**iet.
deletes *iet*	Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt, maar dat wil hij nooit.

Table 4.4: Writing actions in the composition of the sentence "Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt, maar dat wil hij nooit."

decide to leave the sentence produced so far to make a revision elsewhere – in the same sentence or at another segment of the text – after which they return to the end of the sentence they were writing. This would not be an addition, because the sentence is not yet completed. However, the fact that the author moved away from the leading edge is meaningful for the interpretation of the writing process, as it provides information about the steps that were taken to write the sentence. To be able to identify this return to the leading edge, the text can be encoded using <mod>. According to the TEI P5 Guidelines, the <mod> element may be used to represent "any kind of modification identified within a single document" (The TEI Consortium 2020, sec. 11.3.41). For the purpose of analysing digital writing processes, it may also be used for the "modification" of unfinished sentences – the continuation of writing the sentence – using the attribute: @type="continue". A transcription with the inclusion of <mod> tries to model the flow of writing. To ensure a visualisation that can 'replay' the writing session, each separate writing action has to be encoded. This means, for example, that when the production of new text is interrupted with an immediate pre-contextual deletion or typo – in which the cursor does not change position – the add element of the new text production is closed directly after the closing tag of the deletion. When the author

continues with the new text, this is encoded in a new add element, indicating that it is a continuation of the action with `@type="nt|continue"` – this addition of `"|continue"` can be used with all other possible actions as well (such as `@type="context|continue"`).

Most of these revision types can be illustrated using the steps taken by Bogaert when he wrote the example sentence from Table 4.4 which he wrote in Session 30: "Soms kan hij meer krijgen dan wat hij voor zo'n kunstwerkje vraagt, maar dat wil hij nooit".[76] He started by writing "Soms beiden", then moved his cursor between the two words using the left arrow key. There he wrote "krijgt h". This is a pre-contextual addition, because it takes place before the sentence is finished.

Bogaert then continued writing with another pre-contextual addition between "Soms" and "krijgt": "kan hij meer". After inserting this fragment, he relocates the cursor between the letter "g" and the letter "t" in the word "krijgt" and continues writing. Bogaert left the leading edge (the point where he created new meaning) to make the pre-contextual additions, but after these insertions a new leading edge is created as he continues writing the sentence between the letter "g" and the letter "t". At the new leading edge, he writes: "en dan wat hij voor zon' kunstwerkje vraagt"; the screen would now have displayed the sentence as:

Soms kan hij meer krijgen dan wat hij voor zon' kunstwerkje vraagtt hbeiden.

The new leading edge was not positioned at the end of the unfinished sentence, but after the letter g in the word "krijgt". It was thus followed by "t hbeiden". Bogaert then corrects the typo made in the production "zon'" (the apostrophe was incorrectly positioned) after which he eventually deletes the bulk of unused characters at the end of the sentence. These are all pre-contextual deletions; the sentence is still not finished. Now the leading edge is positioned at the end of the sentence, where Bogaert continues writing: " , maar dat il hij niet". After correcting the typo with an addition – he missed the letter "w" in writing "wil" – he types the full stop. This marks the moment the writing of the sentence is finished. Somewhat later in the session, Bogaert returns to the sentence to make a contextual revision. He substitutes "niet" with "nooit" by adding "ooit" and deleting "iet". The writing process of this sentence illustrates the complexity of digital writing, but also demonstrates that the proposed encoding succeeds in capturing every step in the process. Still, this encoding misses an important aspect of the writing process: time.

---

[76] Translation: "Sometimes he can get more than he asks for such a work of art, but he never wants that"

```
<add type="nt">Soms <add type="pre-context">kan hij
meer </add><add type="pre-context">krijg<mod type="
continue">en dan wat hij voor zo<type="typo">n</del>'
<add type="typo">n</add> kunstwerkje vraagt</mod><del
type="pre-context">t h</del></add><del type="pre-con
text">beiden</del><mod type="continue">, maar dat <add
type="typo">w</add>il hij n<add type="context">ooit
</add><del ="context">iet</del></mod><mod type="cont
inue">.</mod></add>
```

4.2.2 Specific encoding of time

Inputlog logs every keystroke and mouse movement in combination with a timestamp. Unlike analogue writing processes, the keystroke logging data allows us to incorporate the specific time of writing into the encoding. Through this temporal aspect, the writer can – so to speak – be followed through the text. Lindgren and Sullivan mention this aspect of keystroke logging too when they argue that the location of revisions

> shows how the writers move their points of focus during text composition; this can be viewed as the route writers take through their texts. The actions writers perform during composition can, for example, hint at the writers' developing ideas and associated shifts in text focus. (Lindgren and Sullivan 2006b, 39)

Incorporating the recorded time of the revision into the encoding thus offers a unique opportunity to study the text's genesis on a microscopic level – what I refer to as its "nanogenesis" (see Chapter 5.2).

The timestamp enables the genetic scholar to investigate the location at which the author was working before they made a revision at another place in the text, when (and how quickly) revisions were made, and if there were certain revision campaigns. To analyse this, the scholarly editor may encode the timestamp for each addition and deletion and every other event worth mentioning, by using the @seq attribute (e.g. @seq="yyyymmddhhmmss"). The scholarly editor can choose to incorporate the dates of the writing sessions, so as to visualise the chronology of the writing process, not only within a single session but also across several (or all) sessions. The hours, minutes and seconds indicate the time after the start of the session added to the time the session is started.[77] As such, this notation provides the exact time the textual input

---

[77] The time is derived from the "StartClock" in the "General Analysis" of Inputlog, which is added to the start time of the session in question. In order to be able to retrieve the event in keystroke logging data, the unique ID of each event in the keystroke logging data needs to be included in @evidence. As the time of every keystroke is given, the editor needs to make a decision as to which time to incorporate in the encoding. For a genetic analysis, the time of an event's first keystroke may be the most fitting option; for example, when the first key is pressed to start production of a new sentence.

took place. The TEI P5 Guidelines propose the attribute @seq (sequence) for assigning "a sequence number related to the order in which the encoded features carrying this attribute are believed to have occurred" (The TEI Consortium 2020, sec 11.3.1.4). In the case of the logged writing processes, the @seq attribute can be very specific as the data provides information about the time the deletions and additions were being made.

The timestamp given in @seq can subsequently be used to number all the changes in @n. Using an XSLT script, the events can be listed chronologically in <listChange> and allocated a number.[78] The @n includes the number of appearances of all insertions and deletions, as well as all returns to the leading edge. From a computational perspective, the number in @n provides the same information as is given in @seq: the chronology of the modifications. The benefit, however, is for the (human) reader. In the eventual transcription, the numbers will offer the reader the possibility to see the sequence of the revision in one glance. This is one step into making the complexity of the (digital) writing process more easily analysable for the reader. The example below shows the encoding of the same sentence discussed above, with the inclusion of the time and the order of appearance, starting from 27.

```
<add seq="20130826151155" type="nt" evidence="1342"
n="27">Soms <add seq="20130826151216" type="pre-
context" evidence="1400-1411" n="29">kan hij meer
</add><add seq="20130826151210" type="pre-context"
evidence="1374-1383" n="28">krijg<mod seq="20130826
151221" type="continue" evidence="1423" n="30">en dan
wat hij voor zo<del seq="20130826151232" type="typo"
evidence="1508-1509" n="31">n</del>'<add seq="20130826
151233" type="typo" evidence="1513" n="32">n</add>
kunstwerkje vraagt</mod><del seq="20130826151236"
type="pre-context" evidence="1552-1557" n="33">t
h</del></add><del seq="20130826151237" type="pre-
context" evidence="1559-1581" n="34">beiden</del><mod
seq="20130826151241" type="continue" evidence="1595-
1643" n="35">, maar dat <add seq="20130826151248"
type="typo" evidence="1643" n="36">w</add>il hij n<add
seq="20130826154211" type="context" evidence="10622-
10625" n="170">ooit</add><del seq="20130826154212"
type="context" evidence="10626-10633" n="171">
iet</del></mod><mod seq="20130826151254" type="cont
inue" evidence="1666" n="37">.</mod></add>
```

The number gives the exact order in which the modifications were carried out while keeping editorial interference to a minimum. This contrasts with analogue sources, where the complexity of documentary evidence turns the numbering of revisions into a

---

[78] I would like to thank Vincent Neyt for writing the XSLT script for this purpose.

highly interpretative act resulting only in speculative readings (Dillen 2015, 90). By comparison, the keystroke logging data allows for a detailed reconstruction not only of the revisions made at sentence level, but those at complete-text level as well. If the author first made a revision to a sentence in the middle of the text and then another in a sentence at the top of the text, this movement through the text can be reconstructed, and also – crucially – referenced in analyses of the writing process.

4.2.3 Peculiarities of digital writing

In the encoding of the keystroke logging data, the way the text is typed is taken into account. This way, we can distinguish between different typing styles. In this respect, at least two characteristics in digital writing become apparent: 1) the recycling of words and characters, and 2) the different ways of performing a deletion. These characteristics may guide scholarly editors in the decisions they make in the encoding of born-digital writing processes.

Recycling

Although the act of deleting is effectively free of cost in a word processor (Sullivan 2013, 256), authors might recycle words and characters in rewriting their texts. This characteristic is also noted by Kollberg in her study of digital revisions (1998). She discusses how a writer in her corpus keeps the "t" in substituting "there" for "it":

> Probably in order to minimize the effort to make this change, the writer keeps the *t* in *it*, and uses it in the new word *there*. Two elementary character level revisions performed at different positions are the result, but the effect of both revisions is at the word level (and the words are at the same position). Many writers would have deleted the whole word in this situation. (Kollberg 1998, 78, emphasis in original)

Kollberg concludes that people develop personal habits in their use of the word processor; each writer has their "own personal set of organization of operations" and is used to performing "certain actions in certain ways" (1998, 78). Not unlike handwriting, typing styles contain a "fingerprint" of the writer (Lindgren, Knopse, and Sullivan 2019, 5).[79] Because genetic criticism is interested in the author's way of working, the way in which they make use of the word processor also needs to be apparent in the transcription.

---

[79] This also depends on the computer or laptop used; the keyboards used in a desktop PC set-up are more likely to incite use of the delete key, which is not available as a single button on many laptop keyboard layouts.

As for Bogaert's way of typing, his recycling of words is very prominent. Indeed, it is already present in the sample sentence discussed above. Here, Bogaert added the clause "en dan wat hij voor zon' kunstwerkje vraagt" between "krijg" and the letter "t" of the word "krijgt". As such, he recycles the word part "krijg", re-using it in the word "krijgen". This re-use can be detected at the word level as well, as Bogaert kept the letter "n" in the substitution of "niet" with "nooit". This is quite characteristic of Bogaert; as Kollberg remarked in a similar situation quoted above, many others would have deleted the entire word.

This recycling of words and characters makes transcription of the writing process a complex matter, as it makes the concise representation of the flow of writing more challenging. This characteristic therefore highlights the importance of encoding the returns to the leading edge with a different tag. In the genetic transcription, it is possible to reconstruct that "en dat wat hij voor zo'n kunstwerkje vraagt" (n30) was written between "krijg" (n28/1) and "t h" (n28/2) while considering that this is not a regular addition, but the writing of the sentence itself. In this visualisation, the process of the writing is emphasised and the singularity of Bogaert's writing accentuated. Hence, it is important to encode the separate steps in the writing process to be able to reconstruct the flow of writing.

Deletions

Bushell emphasises the uniqueness of acts of deletion in a writing process (Bushell 2009, 187). She states that "although it is a functional act, deletion is surprisingly individualized, so that the way of writing and the nature of any writer's body of compositional material is strongly defined by the modes of revision used" (Bushell 2009, 187). It is equally surprising that this also still applies to digital writing processes, as another way in which typing styles become apparent is the usage of the keyboard in making a deletion. As Kollberg notes, a delete operation can be performed in two directions: forwards and backwards (1998, 29). A forward deletion removes characters to the right of the cursor, a backward deletion those to its left (Kollberg 1998, 29). A forward deletion may be carried out by using the delete key or by selecting the characters to the right of the cursor and pressing the backspace key. Using only the backspace key performs a backward deletion. The way the writer uses the keyboard in performing revision affects the encoding of the revision.

When a writer uses the backspace key to delete characters in a substitution – a backward deletion – the deleted word will usually appear in front of the inserted word. If the author uses the backspace key to delete words during the production of a sentence, the cursor is continuously positioned at the end of the leading edge. During

production of the clause "de sheerne khoran kan worden gebracht",[80] Bogaert changed the simple past tense verb "kon" into the simple present tense "kan". After writing "kon", he deleted "on" and then continued writing by typing "an":

```
<seg>[...]de sheerne khoran k<del seq="20140717142431"
type="pre-context" evidence="6689-6690">on</del>an
worden volbracht [...]</seg>
```

This pre-contextual deletion can be considered as the digital equivalent of *currente calamo* deletions in analogue material, which "usually characterize writing produced by an author in the throes of composition, with corrections or revisions made immediately rather than later" (Beal 2011, 104). The linearity of the pre-contextual deletions and the production of the sentence facilitate the readability of the encoding.

The author may also use the delete key to remove a part of the text – a forward deletion. Bogaert prefers this technique: when he makes a substitution, he writes the addition prior to the deletion so that the new word appears to the left of the older one – in the substitution of "niet" with "nooit", for instance, the writing of "ooit" preceded the deletion of "iet". The addition therefore appears before the deletion in the encoding. In the encoding of analogue material, however, Elli Bleeker notes that a deletion is normally located

> before [i.e. "to the left of"] an addition in a transcription (regardless of the actual positioning of these elements [on the document]), simply because – in the western world – we read a transcription from left to right and we usually assume that a word is first deleted and then replaced. (2017, 87)

This choice is usually guided by the goal of the transcription of analogue material: to render the text more readable (Bleeker 2017, 87). In the transcription of a digital writing process, the goal is also to reconstruct that process – as this is not visible in the document – and to capture the author's way of working. The additions and deletions are therefore best placed in the position at which they occurred: the way the deletions are performed dictates the decisions made in the encoding.

4.2.4 Interpretation, selection and argumentation

The proposed encoding produces a transcription of the keystroke logging data in order to provide data output suitable for a genetic analysis. Specifically, it allows for the examination of revisions and new text production, their sequences, and their effect on the text. This transcription alone is not sufficient to create a digital genetic edition, but it provides a sound basis for the visualisation of the writing process. Moreover, the act

---

[80] Translation: "the sheerne khoran can be brought".

of encoding the keystroke logging data does coincide with the encoding practice for analogue material in that, here too, "relatively simple text encoding forces us to make editorial decisions" (Bleeker 2015, 112). In the case of keystroke logged writing processes, the need for abstraction (and therefore interpretation) of the recorded material only increases, because there is so much additional information available to the scholarly editor. As the examples above demonstrate, simply converting the writing actions that are recorded in the logging data to their editorial representations already implies making a series of editorial choices, such as selecting the data and deciding where the encoded insertions and deletions should be located in the transcription. While the keystroke logging data serves to make more objective observations about the sequence of the writing, it also forces the scholarly editor to make their interpretation of the material even more explicit.

The transcription of the keystroke logging data tries to be as objective as possible. In its proposal to complement a text-oriented approach with a document-oriented one, the Workgroup on Genetic Editions, part of the TEI Manuscripts Special Interest Group (TEI Ms SIG), refers to the opposition in German editorial theory – as coined by Hans Zeller – between the "Befund" and the "Deutung". Respectively, these refer to "what is there in the source document, the record" (Befund), and "the interpretation of this phenomenon" (Deutung) (Workgroup on Genetic Editions 2010, sec. 1.1). The Workgroup notes that one cannot talk about the record without any interpretation (especially not in the realm of genetic criticism) but does make a distinction between different levels of interpretation:

> there is an obvious difference between the interpretation that some trace of ink is indeed a specific letter and the assumption that a change in one line of a manuscript must have been made at the same time as a change in another line because their effects are textually related. (Workgroup on Genetic Editions 2010, sec. 1.1)

The Workgroup therefore proposes making a distinction between the interpretation of "what's there" (document/fact) and "how does it relate" (text/interpretation) (Workgroup on Genetic Editions 2010, sec. 1.1). A similar distinction is made in research into cognitive writing processes, when it differentiates between *elementary revisions* and *interpreted revisions*. According to Kollberg, an elementary revision is a single deletion or insertion, and the analysis of such elementary revisions is therefore based only on "the writer's overt action in manipulating the text, with a minimum of interpretation of how revisions may be related according to the writer's intentions" (Kollberg 1998, 16). Interpreted revisions, on the other hand, are revisions which are analysed at a higher level. For example, if "two or more elementary revisions that are

seemingly united by the same goal may be combined and interpreted by the researcher as a unit" (Kollberg 1998, 17).

Following that logic, the proposed encoding therefore focuses solely on the elementary revisions. For example, the distinction between contextual and pre-contextual revisions rests only on the author's actions as they are recorded in the keystroke logging data. When a revision is made during the production of a sentence (before the author presses the full stop), it is marked as a pre-contextual revision. When the revision is made within a completed sentence (after the full stop is typed), it is marked as a contextual revision. In addition, revisions are encoded according to the way they are performed and the replacement of one word with another is not encoded as a substitution.[81] Nevertheless, this can only allow for a certain degree of objectivity, since the selection of the material already involves interpretation (Dillen 2018, 38).

Selection plays a pivotal part in the encoding of the keystroke logging data. Although the transcription sets out to represent the author's movement through the text, many indications of movement have been left out of the encoding. The encoding marks only the textual output, as generated by the keyboard, consisting of characters and punctuation marks. As such, it omits the keyboard events "UP", "DOWN", "LEFT" and "RIGHT". The same applies to the mouse movements. In focusing on the textual output, the encoding also ignores data provided about pauses, their locations and the timing of each action. Moreover, as the only time that is encoded in the transcription is the start time of any modification, the time between the end of one revision and the start of another is omitted as well. This means that the time between two subsequent revisions cannot be deduced. This does not imply, for instance, that long pauses or other time indications cannot be encoded in the transcription, but rather that such an encoding does not lie within the scope of this particular transcription. The main aim of this transcription is to help the scholar/reader follow the sequence of text production and revisions with a focus on the text and its meaning. By reducing the presentation of other kinds of information, the scholar/reader is less distracted from analysing the text. Still, even with a focus on the textual output, interpretation remains a key factor in the encoding as "the idea of presenting a text in an objective way is problematic and arguably impossible" (Bleeker 2015, 114). This seems even more true when encoding keystroke logging data, as the scholarly editor is setting out to reconstruct a state of the text which has never existed in full. However, providing a transcription in which the pauses are encoded might still be useful, since "pauses, moments of physical inactivity during writing, offer observable clues to the covert

---

[81] For example, when "niet" is changed to "nooit" by adding "ooit" and deleting "iet", only "ooit" and "iet" are marked with elements.

cognition processes which contribute to discourse production" (Matsuhashi 1981, 114).[82]

When all the deletions and insertions within a given session are encoded, we arrive at a state of the text that has never actually appeared on the author's computer screen as such, and has therefore never interacted with. This might present us with some issues, such as the question where to encode insertions and deletions when several revisions are located at virtually the same position in the logs (Kollberg 1998, 34). The scholarly editor's interpretation is necessary in these instances, especially when the author makes an insertion next to a place in the document where there was a previous deletion. That is the case because when the author inserts text, the text is inserted at the cursor location. But since any previously deleted text remains visible in the encoding, there is no straightforward place to locate the insertion in relation to the previously deleted text (Kollberg 1998, 34). A protocol for such cases could be that when it pertains a single insertion, the insertion should be encoded to the right of the previous deletion – in line with Bleeker's argument for encoding deletions and additions in analogue witnesses. And when it comes to substitutions, the way in which the insertion and deletion were performed could help make the most accurate decision. For example, when new text was inserted first, and the old text forward deleted afterwards, we could transcribe the insertion first (i.e. to the left) and the deletion second (i.e. to the right).

The scholarly editor's interpretation also comes to the fore in the transformation of the TEI-XML encoding. Joris van Zundert and Tara Andrews argue that the interface of the digital edition functions as an argument: "Our first observation is that a digital edition's interface is an argument – not just an argument about the text, but also an argument about the "attitude" of the editor, a window into his or her take on methodology and the digital edition itself" (2018, 7). The interface of a digital scholarly edition "is always closely linked to the data model of the underlying data and the editorial principles expressed in this data model", so they function as "an interpretation of knowledge and provide users with a more or less "guided" tour through the data and its general presentational setting" (Bleier and Klug 2018, VII). While different transcriptions given below cannot be considered as a fully developed interface, they already function as "an integral part of rhetorical form" since they foreground the textual development (Andrews and Zundert 2018, 8). To address how the transcriptions

---

[82] I have conducted an experiment with encoding pauses longer than 1000 ms. in the transcription. Although the result of this experiment was promising and pauses are an important point of references for investigating for example revision and discovery in writing, I have chosen not to include them in further transcriptions of the keystroke logging data due to time constrains, and more importantly, not every pause will be relevant for the analysis of the writing process. Instead, my analysis will start with analysing the production of the text, when examining a particular writing action in detail, I will then also consult the pauses within this event to account for it in my interpretation.

assist in guiding attention towards the dynamics and non-linearity of the writing process – in this case within a single writing session – we have to take the eventual visualisations into account.

4.2.5 Visualisation

The resulting transcriptions of each writing sessions provide detailed information about the writing process and the locations of the modifications in the text, but a visualisation of the transcription is essential to facilitate the analysis. Therefore, I will now turn to the HTML visualisations of the transcriptions in TEI-XML that visualise all the types of modifications that were made in the paragraph at their location in the text. The visualisations can be found on trackchanges-cmg.uantwerpen.be. As an example, I will use the aforementioned paragraph from *Roosevelt* in the first chapter narrated by Faraaz (Session version 10, Session 196; see Figure 4.2).

Some digital editions, such as the Shelley Godwin Archive (SG-A) and the Samuel Beckett Digital Manuscript Project (BDMP), offer the possibility to visualise certain features while obscuring others. Since Mary and Percy Shelley worked together on the novel *Frankenstein*, the SG-A allows the reader/user to solely visualise Mary's writings, or, conversely, only Percy's revisions. And in the BDMP "the user has the possibility to manipulate the visualisation of the transcribed text to some extent, by selecting a tool in the navigation bar" (Dillen 2018, 46). As such the reader/user can visualise information about the location of additions with the 'place indications' tool, indications of different writing tools with 'writing tools', and solely visualise the addition and removing deletions with 'top layer' (46).

To direct the gaze of the reader/user to the aspect of textual development, we chose the state of the text at the start of the writing session as the default representation in the visualisation of the transcription. This default representation makes the reader/user aware of the state of the 'ground layer' on which the author started to work. Then, by means of *toggles*, one can visualise all the textual operations by clicking 'all modifications'. This visualisation gives a reconstruction of all the textual operations within their textual context and promotes reading the text with all the modifications made during this session (see Figure 4.3). The different types of modification are visualised in different colours, which indicate that the writing of this paragraph proceeded in different steps. With the toggle 'symbols', symbols are added to the visualisation to provide an option for people with colour vision deficiency to still be able to distinguish the different modifications. Adding these symbols also enhances readability at the borders of the insertions and deletions. At a glance, we can see the dynamics that underlie the writing process. In the elaboration of the note "Lange groene linnen schort", for example, some words of the note were recycled. Bogaert did not write a new sentence to address the information given in the note, but he continued

finishing this note and transformed it into a syntactically correct sentence. In several steps, he added new words to the note, gradually textualising it. In doing so, he also experimented with the colour of the apron: changing it from green, to blue, to green again. Because he eventually chose for green, this decision-making process could not be traced in the Word documents. This visualisation of 'all modifications' therefore also highlights the amount of information that was captured in the keystroke logging data and could not be reconstructed based on the Word documents. Another example of this is the translocation of the sentences "Hij is anoniem en onzichtbaar. Hij beweegt geruisloos".

The recycling of the note "Lange groene linnen schort" can also be observed by removing all the additions with the toggle 'additions'. Here, the deletions are visualised in dark red, which allows us to see which text has been removed during the session. When the additions are visualised again, but deletions obscured by clicking the toggle 'deletions', we can analyse the text at the end of the session – the 'top' layer – and observe how this text was written. For example, the adverb "heel" has been added before the sentence was finished by typing the period.

Within the scope of genetic criticism, 'typos' are not the most meaningful entities because they do not immediately affect the meaning of the text. Therefore, the deletions which were made to correct typographical errors can be filtered out of the visualisation by means of the toggle 'typos'. Only the deletions are removed, because obscuring the additions would make for a representation of the text that is also not typographically correct. The discussed visualisations of the transcriptions in TEI-XML thus far allow us to analyse how the text was modified during the writing session. The different types of modifications that are represented in different colours indicate aspects of the writing process; it shows how sentences were produced. These visualisations – displaying all the modifications or highlighting certain types of revisions while obscuring others (e.g. removing all the deletions from the visualisation) – coincide to a large extent with traditional visualisations of transcriptions of analogue manuscripts, since they have until now neglected unique features of the keystroke logging data: time and movement.

The visualisation of all the modifications allows for a microgenetic analysis, however, since the time is also encoded in the transcriptions, the nanogenesis can easily be made visible as well. One way of doing this is through visualising the numbers by means of the toggle 'numbers' (the numbers can be enlarged by the toggle with the magnifying glass). As such, we can – so to speak – follow the author through the text.

The transcriptions of the keystroke logging data and its visualisation represent a text that has never actually appeared on Bogaerts screen as such, and with which he therefore never interacted. Needless to say, deletions disappear from the screen when writing in a word processor and additions do not all occur at the same time. Therefore, another visualisation feature can reconstruct the nanogenesis as closely as possible
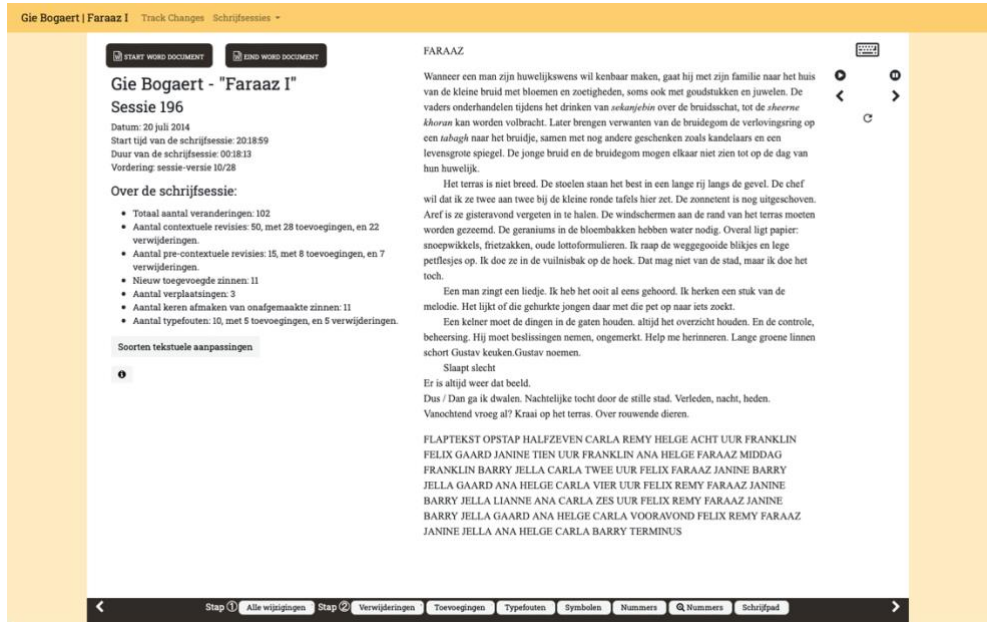
Figure 4.2: Screenshot of the visualisation, showing the default representation of the text.
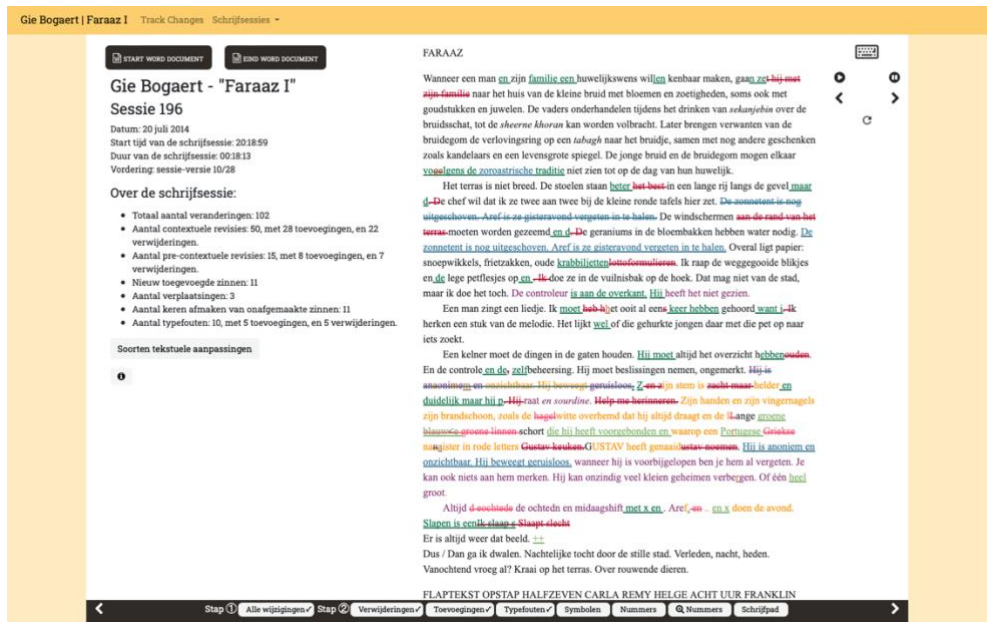

Figure 4.3: Screenshot of the visualisation with all the modifications shown.

through replaying the encoded writing operations step for step: the 'replay' function. By clicking on the 'play button', the additions are visualised in their order and deletions are removed from the visualisation at the moment they are deleted. Before the deletions are removed from the visualisation, the deleted text is briefly formatted in bold to help the reader/user detect the deleted text. The reader/user can also 'walk' through the session at their pace by means of the arrows. In this way, the nanogenesis can be studied even more conveniently and we can reconstruct what the text looked like after any number of modifications.

In order to be able to identify sessions in which Bogaert made revisions at several locations in the text – within and across chapter boundaries, and which may be an indication of a revision episode – I implemented a visualisation that serves to give an overall impression of the nanogenesis. This visualisation provides an overview of what, in this specific context, might be called the 'writing path': the route the author took through the text during the writing session. This idea of the 'writing path' is predicated on the idea of a "genetic path" or, in D'Iorio's words, 'parcours génétiques' (2010, 53). The idea of the genetic path is further described in the pearl/pearl-diver model of HyperNietzsche (see Bartscherer 2003), which is summarised clearly by Bleeker and Kelly:

> In this model, the editor is a metaphorical pearl diver who collects beautiful gems from the bottom of the sea. Such a gem – a *pearl* – is a special element or unit in the archive. Pearls range in size and could consist of textual fragments, editorial commentary, or complete notebooks among others. For instance, if we want to study the development of a certain word, we could search for all pearls that contain this word, string them together in a chronological order, and thus generate a new sequence called a genetic path. (Bleeker and Kelly 2018, 202)

If we apply such a metaphorical description to the idea of the 'writing path', I suggest we call it the 'spider's web' approach. As the author modifies the text during the writing session, the writing operations are, so to speak, linked by a thread of silk. At the end of each writing session, we can follow the threads and see the 'spider's web', visualising the author's movement through the text.[83] Whereas the genetic path links textual units across the boundaries of the document on the basis of semantic similarity, the writing path links textual units within the boundaries of the document (or rather: a writing session) and links them on the basis of the author's behaviour during the writing process.

---

[83] See Buschenhenke, Conijn, and Van Waes (2023), for a quantitative approach for studying the author's movement through the text, or rather the non-linearity of the writing processes, based on keystroke logging data.

The visualised spider web allows us to see in one glance whether the writing was characterised by linearity, or not (Figure 4.4). The orange lines indicate that the next modification took place in the same line or below the previous modification. Conversely, in case of the blue lines, the next modification took place above the previous modification.[84] In this way, the 'writing path' can be a first investigation into the writing episodes, as it indicates how the author moved through the text.

This visualisation can also be used to compare writing sessions with one another from the perspective of the nanogenesis (see Figure 4.4). As we have already observed in the discussion of the writing of the specific paragraph in Session 196, Bogaert revised and inserted new text in this chapter of Faraaz mainly in a linear way – except for a couple of movements within the chapter and one movement to the second chapter of Faraaz.[85] A comparison with the writing path of two other sessions may clarify the usability of this visualisation. A clear example of a linear writing session was Session 202. During this session Bogaert worked, with a few exceptions, from the start of the first chapter narrated by Faraaz towards its end. We get an opposite impression of Session 243, in which the crisscrossed lines indicate that Bogaert made a lot of jumps in modifying the second chapter narrated by Faraaz and that he also made adjustments in other chapters.

In Session 202, which lasted 33 minutes and 12 seconds, Bogaert primarily made contextual revisions. Thus, since most of the revisions took place in already existing text and the revisions were made from top to bottom, this enables us to characterise this session as a 'revision session'. The text was not yet finished, but he re-read the text he had already written to make revisions when necessary and to align his text with his intentions of that moment. Session 243 was considerably longer (2 hours, 13 minutes, and 23 seconds). Bogaert added a lot of new text and made a great number of revisions. During this productive session he was iteratively going back and forwards. Thus, he was constantly revising what he had written, also during the production of new text, which may indicate that he was struggling to find the right words, was experimenting with formulations or that he was distracted by 'errors' elsewhere in the text while formulating

---

[84] I am aware that the layout in the visualisation differs from the layout in Microsoft Word (also because the deletions are all given in the text in this particular visualisation), this may cause that the distance of the revisions varies sometimes slightly from the actual distance on the moment the text was modified in the word processor. This is due to our choice to capture mainly the textual features, since the material features – such as the position of the text in the digital document – changes with each modification and is therefore almost impossible to reconstruct. However, we believe that the representation is accurate enough to give a *general* idea of the author's movement through the text.

[85] Mathias Fürer analysed the movements of Swiss television journalist through their texts to distinguish different writing phases, which he defines as: "changes in dominant revision behavior, for example by a shift from moving forth to moving back or from linear to nonlinear, fragmented writing" (Fürer 2017, 21).

new text. The writing of the paragraph in Session 196, on the other hand, might have been easier to write because he could fall back on Bernlef's story for this specific section. With the visualisations, we now have a convenient way to study the *process* and the *product* of writing.
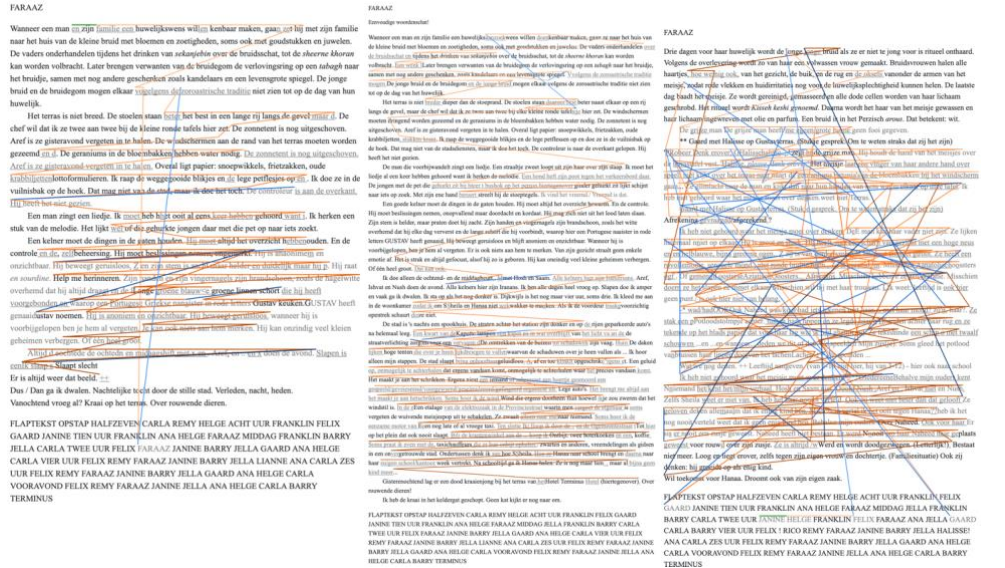
Figure 4.4: The 'writing path' visualisation of the writing of respectively the first chapter narrated by Faraaz (Sessions 196 and 202) and the second chapter narrated by Faraaz (Session 243). The orange lines indicate a revision on the same line or below the previous writing operation (down-ward movement) and the blue lines a revision above the previous writing operation (upward-movement).

4.3 THE QUESTION 'WHY?'

Now that the genesis of the text can be studied in great detail, thanks to proper reconstructions and visualisations of the writing process, including all the revisions, we can examine the final property of revision that we have not yet touched upon: the trigger (the cause) of the revision. The study of triggers is an interpretive act, and therefore fits within textual scholarship and genetic criticism, where interpretation plays a prominent role. John Bryant defines revision as "the visible sign of altered intentions" (2002, 12). In order to understand this 'altered intention', the revisions need to be interpreted by the geneticist. Take, for example, Van Mierlo's analysis of another of Keats' revisions in the autograph manuscript of 'Ode to a Nightingale' (Van Mierlo 2013). I will quote here in full, since it shows how Van Mierlo comes to develop his hypothesis on the reasoning behind the revision based on his awareness of rhyme schemes:

> What is fairly certain is that the change from 'painful' to 'drowsy', although doubtless an improvement, comes as a result of altering 'falls' to 'pains'. But

> how did this come to pass? Are we to imagine that Keats held a poem in his mind that began with an irregular *abcb* rhyme scheme – an irregularity that he removed by adding 'pains' so that it pairs with 'drains'? Somehow this seems unlikely. So is it that he revised the first line *currente calamo* before composing the next? In this case, 'pains' existed before 'drains', which implies that the third line was an invention on the fly. Whatever line the poet first had in mind it must have rhymed with 'falls'. Or is it rather that he invented the third line on the fly, discarding whatever originally he had in mind, which then necessitated correcting the opening rhyme word? The second scenario seems the more likely one, because the syntax joining up lines 1 and 2 works properly with 'pains | My sense', but not with 'falls | My sense' and because the placement of the final word, 'drunk', dips slightly below the base of the line, indicating that it was written after 'pains' above. (Van Mierlo 2013, 22)

This particular example shows that handwritten manuscripts can lead to well-founded arguments about the trigger of revisions, especially when analysed within the context of the text. With keystroke logging, however, we may be able to come even closer to the writer's decision-making process or 'altered intentions', since we have information at hand about the order of the revisions.

Before we go into more detail about the order of revisions, consider the second pre-contextual revision (n97) by Posthuma in the following sentence:

> [94]De vriend ~~is ha~~[95] [96]en haar man bespreken nu de … van kernenergie en de vriendin is opgestaan om ~~een nieuwe fles wijn te halen~~[97] [98]de afwasmachine in te laden. (Session 6)

My interpretation of the pre-contextual revision (n97) is that it has an effect on the meaning of the text, as it adds a reference to the traditional gender roles. In this sentence, Posthuma describes how the first-person narrator's husband and the male friend at whose house they are staying are discussing a serious topic (something about nuclear energy), after which the female friend gets up to fetch a new bottle of wine. In the course of writing this description, Posthuma deletes the reference to fetching a new bottle of wine, replacing it with the reference to loading the dishwasher. Now the description, in the eyes of the first-person narrator, emphasises the differences between men and women; men can speak confidently about any subject, while women remain silent and take care of the household. Through the pre-contextual revision, the sentence thus gains additional meaning. What then could be the reason for this revision, could the trigger be a slight change in the author's intention? From the intention of describing a scene at a dinner where there are heated discussions and the wine flows freely, to the intention of highlighting the gender patterns on such occasions?

While Lindgren and Sullivan emphasise that the exact reason behind a pre-contextual revision (or any revision) cannot be known from the keystroke logging data

alone, they do state that pre-contextual revisions indicate that "some form of decision process is occurring that relates to how the composition is to proceed" (2006a, 166).[86] This is what genetic criticism is interested in. In fact, Dirk Van Hulle, states that genetic dossiers make us aware of the mutability of the authors' intentions: "The writer who cancels a word is already different from the one who wrote it" (2022, 11). In writing studies, the word 'intention' often pops up in discussions of the trigger of revisions. In "What Triggers Revision", John R. Hayes explores the theoretical and empirical developments with regard to the question "What is it that leads writers to change what they have written or what they had *intended* to write?" (2004, 9, emphasis mine). Even more, some researchers (e.g. Sommers 1980) have suggested that "the process of revision is triggered when writers notice a dissonance between what they intended to write and the text that they have actually written", which Hayes terms the dissonance models of revision (2004, 11). In Hayes discussion, and critique, of these models, we can see that this idea of the mutability of intentions is also recognised in (cognitive) writing studies, as he states: "Even when the text on the page is exactly what the writer intended at the time, he or she may decide later, looking at the text from another point of view, that it needs to be changed" (11). This thus makes clear that, when giving a possible interpretation of revisions (which, in itself, remains rare in the field), researchers in writing studies do not skimp on using the word 'intention', although they are, of course, cautious about drawing conclusions: "it is not possible to draw any conclusions about the writer's intention on the basis of a keystroke logfile without complementary data" (Lindgren and Sullivan 2006a, 185).

In literary studies, however, critics started to avoid the term 'intention' in the twentieth century, because it became such a taboo (Van Hulle 2022, 11). This is the result of the literary-critical debate around the 'intentional fallacy', a term introduced by Wimsatt and Beardsley in their influential essay of the same title. In "The Intentional Fallacy", they critique literary criticism in which the 'authorial intention' is seen as decisive in the interpretation of the literary work; that the authorial intentions lead to *the* meaning of the text.[87] Despite this traditional literary-critical debate, Van Hulle states that the notion of intention cannot be avoided in scholarly editing. Therefore, textual scholars defined the term intention more clearly. Peter Shillingsburg, for example, distinguished between two concepts of intention: the "intention to mean" and "the intention to do" (1996, 33). He clarifies the distinction as follows:

---

[86] The first pre-contextual revision in this sentence (n95) indeed stresses the difficulty of interpreting pre-contextual revisions. Posthuma deletes "is ha", but it is impossible to know what she intended to write in the first place.

[87] See Bushell (2009; 2005) for a discussion of the debate on the intentional fallacy "in terms of its consequences for the study of compositional material" (2005, 63).

> In the context of creative acts we might see the author – as we no doubt all have experience in our own writing – intend to mean something, followed by the intention to express that meaning in verbal form, followed by the intention to do something (i.e., write down the signs for the verbal expression of the meaning intended). In the act of writing or later upon reflection, the author may discover a new meaning to mean, which is followed by a new intention to couch the meaning in the verbal expression, followed by the intention to do something (i.e., write down the signs for the verbal expression of the meaning intended). (Shillingsburg 1996, 34)

For Shillingsburg, the "intention to *do*" is more directly deducible from the written signs, and the task of the scholarly editor is to be accurate about the "process of doing and discovering new things to do" (35). It is from "this level of intention critics attempt to construct or reconstruct meaning, about which the editor's best hope is to be plausible" (35). Yet, he emphasises that there are various intentions in compositions and that "'intentions to mean' are irrecoverable" (35).

Sally Bushell (2005; 2009) developed her own model of authorial intention in the compositional process, as part of her Anglo-American genetic or compositional criticism with intention as its basis. She acknowledges that there are wrong uses of 'authorial intention', such as when it is "assumed to be the primary purpose of interpretation", it is seen as holding the absolute meaning of the text, and when it is seen as being something static (Bushell 2009, 50). Yet, she outlines situations when it is "perfectly acceptable draw on authorial intention", such as when:

1. It is understood by the creator as a motivating force for the coming-into-being of a literary work (without necessarily being fully understood by consciousness).
2. It informs our understanding and interpretation of the manuscript page as one dimension of it, along with other nonintentional meaning contexts.
3. It is understood to be dynamic and contradictory (forming part of a constantly changing stream of meaning).
4. It is understood to be capable of reconstruction primarily in terms of acts on the page.
5. It is based on an understanding of language as both expressive and experimental (performative). (Bushell 2009, 50)

Taking this into consideration, I focus in this section on the trigger of revision and authorial intentions: the reasons for doing something during the writing process. For me, the trigger of a revision belongs to the "intention to do". Of course – and I will continue to stress this – my interpretations of the writing actions remain hypothetical. Yet, Bryant states with regard to intention, that "textual fluidity offers a more focused

perspective on intentionality that allows us to sharpen the dimensions of speculation" (2002, 9). My aim is to explore how we can indeed 'sharpen' our hypotheses on triggers and intention now the material of the writing process is even more precise, as I focus on how writing technology can trigger revision, how revisions can trigger other revisions, and how revisions are triggered by re-reading. In all these cases, the order of text production and revision will seem to be a grounded starting point.

4.3.1 Triggers and the affordances of the word processor

The act of writing is always situated in the material world, as the implements of writing (e.g. keyboard, screen, pen, paper) have mass (Haas 1996, 4). But can these implements – or the writing technology – also function as a trigger for revision? As noted above, Sullivan argues that the writing technology influences revision practices: "being able to see texts *fixed* in many visually different forms seems to promote textual fluidity" (Sullivan 2013, 8). This question of motivation underlying revisions also returns in Sullivan's discussion of Doug Reside's work on Jonathan Larson's Word files: "we have seen that writers are often driven to revise when they see their work in new formats, and particularly when those formats contain space for earlier novels. This screenshot [of Larson's revision data in a hex editor] *shows* revision, but it would never motivate it. The evidence is no longer the means" (Sullivan 2013, 259). However, although not *fixed*, authors working in a digital environment are able to see their work in a different form, for instance by changing the font. Joanne Harris, for example, changes the font to spot mistakes: "I write in Word, using the Perpetua font. When I edit, I always do it on screen (although at that point I change the font to Times Roman – changing fonts helps me spot things I've missed). For some reason I'm very sensitive to fonts" (Harris in Johncock 2012b). And, of course, they can also make printouts of their work. As Julie Meyerson points out, the printouts may help to see the text in a different way: "when I print my work out and re-read it on the page, I'm always dismayed at how much less good it is – so I have to re-write it" (Meyerson in Johncock 2011b). A combination of how printouts and the digital medium initiate revision can be found in an account of Jon Gregor on his writing process:

> The work of writing is in the revision and the refining. I find that this is more successfully done when working with hard copy. I also find that by artificially creating more "process" (retyping the whole draft from scratch, rather than simply dipping in and making selected corrections), I can force myself to consider the text more deeply. (If I can't be bothered to type it, I probably shouldn't expect anyone to bother reading it.) I mean, turn off the wi-fi. But apart from that, it's all just different coloured pens. (Gregor in Johncock 2012c)

As Calonne points out "there is a natural movement from one writing technology to another during the process of receiving the initial idea, drafting, revising and publication" (2006, 163). Authors write initially on the computer and then change to paper for the final editing. From the authors studied here, Ellen Van Pelt, indeed also made printouts of her story for revision. She explained that seeing her text on paper, also enabled her to spot mistakes more easily. In all these cases, revision still may lead only to what Sullivan calls 'stylistic felicity', but it does indicate that digitally working authors are also able to find ways to see their work 'anew'. Thus, seeing their work in a different form can still be a trigger for revisions.

Making visible the affordances of the word processor, especially the ease of revision, can also affect our understanding of a writing action as a revision. The encoding discussed above, models the writing process at the scale of the writing session. The writing session (and thus the session version) is the starting point of choice because it represents – in most cases – the author's work method, like working in long or short sessions. However, this scale can also obfuscate some aspects of revision, such as the distinction between new text production and revision. Since keyboard-based word processing allows for easy text production and revision, these writing processes are often characterised by a high degree of nonlinearity (Leijten, Van Waes, and Ransdell 2010). This means that new text production and revision can be interspersed at various points in the text produced so far. For instance, sometimes, a newly produced sentence is new text production at the level of the writing session, but this can at the same time also be revision of, say, a paragraph.

This blurred line between new text production and revision can be illustrated with an example from Troch's writing method. In Session 17, Troch added the following new paragraph to his story, but is this all 'regular' new text production, or are some parts better seen as revisions?

[141]S~~ot~~[142][143]toppen. [144]Nee. [13]Nu moet ik o[28]~~O~~[27]pstaan, ik moet ~~w~~[1415]maar weer eens opstaan. [148]De benen strekken. [149]In beweging blijven. [150]Hoeveel moeite me oo~~i~~[151][152]k kost. [153]Eerst nog even ~~a~~[154] [155]wat zout van mijn huid likken. [159]Dat geeft me kracht. [174]De waterfles zoeken. [175]De waterfles niet meteen [178]vinden. [179]V~~o~~[180]eel godverdommes v[200][181]loeken. [182]De waterfles dan toch onder een kast [186]vinden. [183]Vloeken h~~l~~[184][185]elpt altijd. [176]De dop van de waterfles schroeven. [160]~~D~~[161]Met de kan d[162]e waterfles vullen. [177]De dop weer op de waterfles schroeven. [206]Men ~~d~~[207][208]en~~dat ik~~ [213] beneden [214]de ouderwetse wine~~k~~[209][210]kelbel~~be~~[211]te [215][212]hoo~~o~~[216]ren[217]. [163]~~De~~ [164][165]E~~Zn~~ [166] [167]n dan met de fe~~l~~[188][187]les in de hand [189]maar eens voetje voor voetje [218]de trap af~~dalen,~~ [219].[220] [190]~~naar~~[222] In [221]~~B~~[169]b[170]eneden. [171][191]~~, de gelijkvloerse verdieping,~~ [223]de voormalige skilattenverhuur~~, daar is mijn opslagplaats.~~[192] waarvan ik [193]~~D~~[194]d[195]e etalage zo goed en zo kwaad mogelijk [201]heb ~~ik~~ [196]dichtgespijkerd

met het hout van lattenbodems uit Hotel Istoria, tweemaal met de ogen knipperen[224]. (Session 17)

Examining the order of the text production can help in investigating the new text production in more detail. At a certain point in Session 17, Troch has written: "Stoppen. Nee. Nu moet ik opstaan, ik moet maar weer eens opstaan".[88] He then copies some text he has already written and pastes it after these sentences. The replaced text fragment describes the ground floor of the building where the narrator now lives: a former ski rental shop. The narrator is still upstairs, but Troch now wants the narrator to go downstairs: he must get up. In order not to have a sudden transition in the story between the narrator being upstairs and then abruptly downstairs, Troch has to describe the narrator actually going downstairs. We can therefore say that his next writing goal is to describe this action. Let us now concentrate on how this proceeded.

In short, staccato sentences, Troch writes that the narrator says/thinks: "De benen strekken" (n148),[89] "In beweging blijven" (n149),[90] "Hoeveel moeite me ook kost" (n150-152),[91] "Eerst nog even wat zout van mijn huid likken" (n153-155).[92] Then Troch leaves the Word document and goes to check his mail, briefly goes to YouTube where he is listening to a live concert of The Jon Spencer Blues Explosion and visits a Social Media page. Then he continues writing: "Dat geeft me kracht" (n159),[93] "De waterfles vullen" (n160).[94] And he revises the latter by explicating that the narrator is filling the water bottle with a jug: "Met de kan de waterfles vullen" (n161-162).[95] After some pre-contextual revisions, and correction of typing errors, Troch writes the connecting sentence: "En dan maar eens de trap af naar" (n167),[96] which requires replacing the capital B with a regular one (n169-170). And he types a full stop to divide the sentence into two parts.

We could now say that he achieved his goal of writing the transition of going from the upstairs apartment to the shop downstairs. At this moment, he can continue by revising the already written paragraphs about being in the shop. However, Troch stops writing and visits social media and checks his mail. After about two minutes, he returns to the Word document. Instead of shifting his attention to the description of the ski rental shop, he starts to make changes to the text that describe the narrators' intentions of getting into action. He adds new, short sentences, like those that he had already

---

[88] Translation: "Stop. No. Now I must get up, I have to get up again".
[89] Translation: "Stretching the legs".
[90] Translation: "Keep moving".
[91] Translation: "No matter how much effort it takes".
[92] Translation: "First, I have to lick some salt from my skin".
[93] Translation: "That gives me strength".
[94] Translation: "Filling the water bottle".
[95] Translation: "Using the jug to fill the water bottle".
[96] Translation: "And then down the stairs".

written. He adds: "De waterfles zoeken" (n174),[97] "De waterfles vinden" (n175),[98] "De dop van de waterfles schroeven" (n176).[99] The narrator fills it with the jar as already described, then Troch adds "De dop weer op de waterfles schroeven" (n177).[100] And as described, the narrator goes downstairs. Then he changes the sentence in which the narrator finds the bottle to that the narrator cannot immediately find the bottle, by adding "niet meteen" (n178): "De waterfles niet meteen vinden".[101] He continues by adding more new sentences. Because the narrator can't find the bottle immediately, he swears: "Vloeken" (n179).[102] Then he finds the bottle anyway, swearing always helps: "De waterfles dan toch vinden. Vloeken helpt altijd" (n182-185).[103] Finally, Troch makes revisions in the sentences that were already written. The bottle lies underneath a closet and with the bottle in his hand, the narrator goes downstairs. In short, Troch expanded the list of tiny actions and thoughts by the narrator, which reveals the state-of-mind of the narrator: as he has been alone for years, and low on energy, such tiny events weight heavy in his daily life and he is very aware of every action he must carry out.

From the point of view of the whole writing session, the sentences in which the narrator searches for the water bottle, does not find it, swears, then finally finds it, because swearing always helps, can be seen as the production of new sentences. After all, he is adding text that did not exist elsewhere before he started this session, and so new text has been added. But if we look more closely at how this paragraph was written, we can also classify these sentences as revision. As we have seen, the sentences added at the very beginning of this paragraph were written with the intention of writing the transition between the narrator being upstairs and downstairs. Troch had to write new text to connect the fragments. Once he had done this, we could say that he had produced a first version of this text fragment, which was a well-written transition, although it was still a rather flat description. Therefore, we could assume that the sentences he wrote after his break – during which he checked social media and his mail – were dedicated to taking the text to the next level. To make it more literary. The aim of writing, and perhaps even the intention-to-do, changed. If we take this and the aspect of time, the sequentiality of writing, into account, then the new sentences become a revision of the text produced so far.

The way in which this paragraph was written was actually made possible by the affordances of the writing technology used: the word processor *allows* these new

---

[97] Translation: "Searching the water bottle".
[98] Translation: "Finding the water bottle".
[99] Translation: "Turning the cap of the bottle"
[100] Translation: "Turning the cap on the water bottle again".
[101] Translation: "Not finding the water bottle right away".
[102] Translation: "Swearing".
[103] Translation: "Finding the water bottle then anyway. Swearing always helps".

sentences to be inserted between the sentences already written. Troch was able to add these sentences at a later stage, gradually building up his ideas for this paragraph, precisely because the word processor would not make a mess of it. We can imagine that if Troch had used analogue writing technologies, he would have added these sentences later in the writing process, when writing another version. Or he would have planned this paragraph ahead, enabling him to immediately write down these sentences.

If we compare the addition of the new sentences to analogue revisions, as analysed within traditional textual scholarship and genetic criticism, these newly produced sentences can be regarded as revision by extension. Specifically, this strategy can be compared to the strategy that characterised that of James Joyce, which Hannah Sullivan describes in *The Work of Revision*. Sullivan, for example, describes the additive movement in the second half of the "Nausicaa" episode of *Ulysses*, as Joyce added sentences to Blooms thoughts of Gerty's limp. Across two sets of placard proofs, the thought is expanded – the additions are given in bold:

> Thought something was wrong by the cut of her jib. Jilter beauty. Glad I didn't know it when she was on show. Hot little devil all the same. Near her monthlies, I expect, makes them feel ticklish.

> Thought something was wrong by the cut of her jib. Jilted beauty. **A defect is ten times worse in a woman. But makes them polite.** Glad I didn't know it when she was on show. Hot little devil all the same**. I wouldn't mind. Curiosity like a nun or a negress or a girl with glasses. That squinty one is delicate.** Near her monthlies, I expect, makes them feel ticklish. (13.772-377) (Sullivan 2013, 161)

Somewhat further in her chapter on additions as revision, Sullivan states that "contemporary writers who write long, and who stuff in everything including the kitchen sink, necessarily do so as a result of additive revision" (185) and that "recent writers seem able to achieve excess from the first draft" (185). Further on, she analyses an artificial version of David Foster Wallace story "Backbone" – which is a section within *The Pale King* – that compares a version published in the *New Yorker* to a transcript of Wallace reading the story in 2000. Here she concludes that even though Wallace added three passages, he never added "substantial material within a sentence or even within a paragraph" (263), by which she means that "he never performs the Joycean trick of turning 'Jilted beauty. Glad I didn't know it when she was on show' into 'Jilted beauty. A defect is ten times worse in a woman. But makes them polite. Glad I didn't know it when she was on show'" (263). But does this mean that authors who write in a digital environment do not revise by extension at all?

Although it is only one example, Troch's writing method does show that the additive revision strategy is still practiced. Troch added sentences to an already existing

paragraph, like Joyce did in the "Nausicaa" section. It is not the kind of belated revision that is the focus of Sullivan's study of Modernist revision, it is, in fact, very immediate revision (within one writing session). Yet, it proves that digital revision is not only revision towards stylistic felicity; the same revision strategies, like adding substantial material within a sentence or paragraph, can still be found in digital writing processes, but the affordances of the word processor only make it easier to make such revisions earlier in the writing process – at moments at which the complete text is still in construction.

In sum, digital writing technologies can still be a trigger for revision. Writers can change the visual aspects of their word-processed text to make it look 'new', making it easier to identify sections that need revision. The word processor itself can also trigger revision, and not just through the built-in spelling and grammar checkers. Because the word processor allows you to start typing anywhere in the text, texts can be built up in a non-linear way, from the inside out, blurring the line between creating new text and revising it. This affordance of easy revision functions therefore also as a trigger for creativity in revision.

4.3.2 Triggers and revision episodes

In the study of the possible intention behind revisions, revision episodes (see Chapter 4.1) provide a sound basis for the interpretation thereof, since a revision can be a trigger for another revision. In fact, the word trigger is often used in discussing revision episodes (see Lindgren and Sullivan 2006a). As discussed above, Kollberg identifies three types of revision episodes: 1) episode with repetitive revisions at one cursor location, 2) episode with embedded revisions and, 3) episode with a sequence of revisions in previously written text (Kollberg 1998, 85–86). The keystroke logging data now also allows for distinguishing these revision episodes in the reconstructions of the literary writing processes, which in turn can reveal something about the author's 'intention to do'.

Episodes with repetitive revisions occur when the writer makes two or more immediate revisions at one cursor location. Most of the time these revisions are very low level, like the correction of typing errors, but they may also occur when writers are trying out different formulations (Kollberg 1998, 85). In this case, the writer's intention for the beginning of a sentence, for example, may change as they type, or "the writer started the episode with an intention that changed during the course of the revision episode" (Lindgren and Sullivan 2006a, 185). Take, for instance, the way in which Van Pelt came to write the sentence "De oudste van de bende, een kleine, pezige man kiest ons als prooi":[104]

---

[104] Translation: "The eldest of the gang, a small, wiry man chooses us as his prey".

¹⁸¹~~Dan houden~~¹⁸² ¹⁸³~~Ik h~~¹⁸⁴ ¹⁸⁵~~Fi~~¹⁸⁶ ¹⁸⁷~~Wann~~¹⁸⁸ ¹⁸⁹~~Een kleine~~ ~~man~~¹⁹⁰ ¹⁹¹~~, oudere~~¹⁹² ¹⁹³De oudste van de bende, een kleine, ~~grijze~~¹⁹⁴ ¹⁹⁵pezige man ~~met~~¹⁹⁶ ¹⁹⁷~~komt op ons af~~¹⁹⁸ ¹⁹⁹~~naar~~²⁰⁰ ²⁰¹ kiest ons als ~~por~~²⁰² ²⁰³rooi. (Session 2)

Van Pelt wrote this sentence at the leading edge – the place where she produced new content (there was externalized text following this paragraph) – and makes revisions at the same cursor location, which are all pre-contextual revisions. Van Pelt progressively revised the text as she wrote it. She started with "Dan houden" (n181), and immediately deleted this (n182) to write "Ik h" (n183), which is also revised into "Fi" (n184-185; the first two letters name of the son of the narrator, Finn), in turn she changed this into "Wann" (n186-187; she probably intended to write "wanneer" ["when"]), and this is replaced by "Een kleine man" (n188-189).[105] She then deleted "man" (n190) to add another adjective "oudere" (n191). After she had deleted everything again (n192), she wrote: "De oudste van de bende, een kleine, grijze" (n193).[106] She now found the 'correct' start of the sentence. After deleting "grijze" (n194), she continued writing the sentence with "pezige man met" (n195).[107] Then she deleted "met" (n196) and instead wrote "komt op ons af" (n197)[108], "op ons af" (n198) is subsequently changed into "naar" (n199), which is together with "komt" deleted and replaced by "kiest ons als por" (n200-201). The latter is a typo, and when this is corrected with "rooi" to produce "prooi", the sentence is finished. So, what happens here? All these revisions can be grouped in two categories. Up until Van Pelt wrote "Een kleine man", she was merely searching for 'what to write next'. When she decided on describing a pirate approaching the I-narrator and her son, the revisions became oriented towards 'how to formulate' this. The concept of the revision episodes with repetitive revisions at one cursor location, thus proves to be a useful tool for genetic criticism to investigate how ideas are initially transcribed.

Kollberg notes that episodes with repetitive revisions at one cursor location also seem "to depend on the writer's individual writing style with the computer; some writers experiment more than others directly on the screen" (90). This also applies to the writers discussed here (see also Chapter 5.2): Van Rijswijk and especially Van Pelt are merely thinking with their fingers, while Bogaert, Posthuma and Troch, in general, tend to revise after they have written a first version of a sentence.

Episodes with embedded revisions, then, take place when the writer makes an insertion which they modify before it is finished, which makes that the revisions made to the insertion are embedded (Kollberg 1998, 85). Kollberg observes that embedding often occurs when the writer corrects her writing (e.g. typing errors), or tries different

---

[105] Translation: "Een kleine man".
[106] Translation: "The oldest of the gang, a small, grey-haired".
[107] Translation: "sinewy man with".
[108] Translation: "is coming towards us".

formulations, ideas, spelling, etc." (Kollberg 1998, 90). Therefore, episodes with embedding do not always "reflect any writing concerns at higher levels – they are simply due to a number of typing errors, or orthographic issues" (90). An example of this type of revision episode can be found in the writing process of Bogaert as he revises the following sentence:

> Misschien vermoed[25]~~onderstel~~[2623]~~geloof~~[2124]~~denk~~[22] je dat ik een oude fantast ben, zo'n fabulant, die probeert de verveling te verdrijven door goedgelovige luisteraars vreemde verhalen op te dissen. (Session 33)

In previous sessions, Bogaert had already replaced the word "denk" ["think"] to "vrees" ["fear"], which was substituted by "geloof" ["believe"], which was then again changed back to "denk" ["think"]. During Session 33, we see that Bogaert again tests different words. Because these substitutions take place in immediate succession, they can be regarded as *currente calamo* revisions – he seems to 'taste' which word fits best. He starts by adding "geloof" ["believe"] and deleting "denk" ["think"], then he replaces "geloof" ["believe"] with "veronderstel" ["suppose"]. He then turns "veronderstel" ["suppose"] into "vermoed" ["suspect"] by adding 'moet' after 'ver' and removing 'onderstel'. In this substitution, Bogaert recycles part of the word, which occurs more often in Bogaert's way of typing. This also makes that the revision is embedded, since "vermoed" builds upon the previous addition of veronderstel". One interpretation of this episode is that Bogaert's intention was to find a fitting word to describe the thoughts of the addressee in this sentence. Bogaert's choice of "vermoed" ["suspect"] fits well with the context of the sentence – 'suspect' implies 'suspicion' – and the 'you' in the sentence appears not to trust the 'I' immediately. In Session 34, however, it is changed back to 'think'. These revision episodes seem to indicate hesitations, which are an important focal point in genetic criticism (Van Hulle 2004, 158).

In case text production is interrupted to make revisions in the text produced so far, the revisions constitute an episode with a sequence of revisions in previous written text (Kollberg 1998, 86). Kollberg discusses several writing actions during which these episodes may occur, such as re-reading the whole text, or the last paragraph or sentence that makes the writer notice things to revise, or the revisions are triggered by that what the writer has just written. The episodes may vary in length, level, and duration (Kollberg 1998, 100). An example of an episode with a sequence of revisions is present in the seventh writing session of Posthuma. She has just written a paragraph about the 'near-suffocation-incident' of the son of the main character: the son nearly chokes on a piece of his ice lolly. The sentences Posthuma has just written are: "Hun zoon laat een harde boer, en nog een, en dan ligt er een stuk feloranje ijs op de stoep. Ze troost haar

zoon en zegt tegen Tom: ik kon niets doen".[109] Then she relocates her cursor to the third sentence of that paragraph to add "het is sinaasappelsmaak, zegt hij en" (n203).[110] This turns the sentence into: "Haar zoon gaat op in zijn ijsje, het is sinaasappelsmaak, zegt hij en hij knijpt in het karton om het ijs naar boven te duwen en zuigt dan hard".[111] This revision is at least the result of rereading the newly produced paragraph (more on rereading as trigger for revision can be found in the next section), but we can also hypothesise that it was triggered by one of the last written sentences, in which Posthuma described the colour of the ice: bright orange. We can argue that by writing "dan ligt er een stuk feloranje ijs op de stoep", Posthuma gained the intention-to-do to explain the adjective 'bright orange'. To achieve this, Posthuma lets the main character's son mention that his ice lolly has the taste of orange. Then she continues revising the paragraph by adding a new sentence: "Tom! roept ze" (n204).[112] And in the last sentence "Ze troost haar zoon en zegt tegen Tom: ik kon niets doen", she first deletes "tegen Tom" (n205), after which she immediately adds it again with the addition of "en de anderen" (n206).[113] This in turn triggers revisions in the sentence "Tom wandelt naar buiten".[114] This sentence now also has to mention 'the others' and is therefore changed into: "Tom en de vrienden wandelen naar buiten" (n207-209).[115] Then she continues writing by making notes for the story in general. These revision episodes thus shows that the text produced so far can stimulate new ideas (Hayes and Nash 1996, 42).

Kollberg based her typology of revision episodes on Williamson and Pence's (1989) analysis of revision episodes. Williamson and Pence distinguished four types of revision episodes: 1) several repetitive changes on the same cursor location, 2) embedded changes, 3) changes related to a global change and, 4) changes inspired by another change (Kollberg 1998, 86). The first two types correspond with the first two types distinguished by Kollberg, although Kollberg's goal differs. Kollberg wants to "formalise the analysis using the S-notation", while Williamson and Pence's classification is "mainly informal and interpretive" (86). Therefore, Kollberg combined Williamson and Pence's third and fourth type – which are equivalent to each other with regard to the text editing actions undertaken by the writer – into one as an "episode with a sequence of revisions in previously written text" (86). However, for an analysis of revision episodes from a text genetic perspective the interpretive distinction between

---

[109] Translation: "Their son lets out a loud burp, and another, and then there is a piece of bright orange ice on the pavement. She comforts her son and tells Tom: I couldn't do anything".
[110] Translation: "it's orange-flavoured, he says and".
[111] Translation: "Her son is absorbed in his ice cream, it's orange-flavoured, he says, and squeezes the carton to push the ice cream upwards and then sucking hard".
[112] Translation: "Tom! she screams".
[113] Translation: "and the others".
[114] Translation: "Tom walks out".
[115] Translation: "Tom and the friends walk out".

changes related to a global change and changes inspired by another change seems useful, since we want to interpret the (order of the) revision to gain an understanding of how the text came to be and how the revisions had an effect on the *meaning* and the *direction* of the text. Let us therefore take a closer look at this distinction. According to Williamson and Pence, a global change may require individual changes throughout the text, for example when the tense in the text is changed (Williamson and Pence 1989 in Kollberg 1998, 83).

To imagine how the analysis of this revision episode may prove beneficial for genetic criticism, a small thought experiment might be useful, supposing Franz Kafka would have written *Das Schloss* on a computer.[116] At the draft stage, Kafka introduces a shift from a first-person to a third-person narration. Dorrit Cohn examined the shift in the novel's manuscript: "all *ich*s [are] subsequently deleted and replaced by *K.*s" (Cohn 1968, 28). Cohn observed that "[t]he *K.* initial was for the first time directly placed into the manuscript toward the beginning of Chapter III of the published text, in the course of the first dialogue with Frieda in the bar of the Herrenhof" (29). She therefore argues that (the difficulty of) the writing of this scene with Frieda may have invoked the shift in person:

> several elements in the scene of the first meeting with Frieda might have, singly or together, drawn Kafka's attention to the difficulties of writing his novel from the vantage point of a first-person narrator: difficulties concerning the time dimension of his narrative, brought out by the image of the future in Frieda's eyes; the corollary exclusion of the narrator's death from that future; and the problem of self-presentation of an opaque psyche, highlighted by Frieda's knowing comment and by the conjectural *wie ermüdet* that accompanies the first *K.* placed directly into the text. (Cohn 1968, 41)

Cohn clarifies that this is, naturally, not entirely certain: "it may well be that Kafka's motivation for transposing the manuscript worked by slow accumulation rather than sudden insight" (41).

But if we, hypothetically, imagine Kafka writing *Das Schloss* on a computer and logging his writing process along the way, we may then also imagine gaining a better understanding of the temporal aspect of this *K.* revision. For example, we would be able to investigate the time that elapsed before writing of the first *K.* in the text – did Kafka pause before writing the *K.* or did it happen in one flow? We would be able to examine whether the first directly placed *K.* was written prior to the substitutions of the *ich*s by

---

[116] This thought experiment has also appeared in the conclusion of L. Bekius & D. Van Hulle, "The literary draft in the 21st century: from paper notebooks to keystroke logging", *Toward a Comparative History of the Literary Draft in Europe. Comparative History of Literatures in European Languages Series* (forthcoming). This chapter is co-authored, but the analysis and the writing for this section was done by Bekius, supported by comments from Van Hulle.

*K.*s in the manuscript or whether he started the first-person to the third-person revision before continuing writing the third-person narrative directly.

Cohn also tried to discover which other revisions in the relevant part of the manuscript could be connected to the *K.* revision. Only for three out of ten revisions that Cohn listed as possibly connected with the *K.* revision was she able to tell with certainty that they were made during or after the *K.* revision. Trying to uncover this based on keystroke logging data would not cause any uncertainties since the order of the revisions would be extracted from the data. By studying the nanogenesis, we can therefore formulate answers to those questions about the sequentiality of revisions, and potentially also about the reasoning behind these revisions. The latter still remains hypothetical, because it goes without saying that this is only the order of the *actions* of the revisions and not an indication of the moment the *idea* for the revision came to mind.

Just as Kafka in writing *Das Schloss,* Posthuma introduced a shift in narration, but in this case from the third person to the first person. What, then, are our possibilities in analysing this shift – a revision episode in which a global change introduced other changes – now that we can base our analysis on the keystroke logging data? An analysis of this change is given in Chapter 6.3, but within this context of the revision episodes it is already important to note that in the session that Posthuma changed the narration, there were also revisions that are not directly related to the shifts in narration, but that were performed while making this change – these revisions were thus triggered by rereading the text. During Session 4, Posthuma alternated between new text production, revision thereof and the global revision of the change in person. This indicates that, at least in digital writing, revisions and other writing actions that are not necessarily carried out with the same intention or writing goal can intersperse each other during the writing session. For analogue writing processes, this teaches us that related revisions do not have to be made within the same sequence.

Next are the 'changes inspired by another change', which are exemplified by Williamson and Pence with an insertion of a focussing sentence made by a student in the beginning of a paragraph after adding details to the end of that paragraph (Williamson and Pence 1989 in Kollberg 1998, 83). This type has therefore the "greatest degree of inference", but the researcher is allowed to deduce this kind of relatedness through the proximity in time and meaning (Williamson and Pence 1989 in Kollberg 1998, 83). We can again turn to an example from a genetic analysis to exemplify this revision episode and how we can use keystroke logging data to go beyond what is possible with analogue material. In *Kladbewaarders* (2007), Dirk Van Hulle describes how Willem Elsschot gradually introduced a theatrical metaphor in the manuscript of *Achter de Schermen.* Van Hulle argues that most references to the theatre were added at a later stage, but the idea of the theatrical metaphor must have arisen during the writing of the

manuscript, albeit not right at the beginning (2007, 225). The first direct reference to the theatre appears in the form of 'the director' in the addition on page 6A: 'Walgelijk genoeg. Maar ~~ik~~de regisseur vind [sic] het ~~niet~~ ongewoon op een huiselijke tafel.' However, this is an addition (a substitution of the crossed-out pronoun 'I') in an addition (the added sheet 6A). Van Hulle, therefore, argues that it is possible that the idea of the theatrical metaphor originated further on in the text. On page 11, Elsschot wrote: "~~Zitten wij niet samen~~ 't Is hier ten slotte toch maar een schouwburg." And on page 12, the succession of deletions and additions illustrates how Elsschot gradually introduces a new aspect of the theatrical metaphor (make-up): "~~Ik kan Aandikken~~ ~~Aankleeden Opsmuk, Laarmans~~ Nu nog wat schmink" (Van Hulle 2007, 226). The latter sentence illustrates a revision episode with repetitive revision on one location and the overall gradual insertion of the theatrical metaphor is a good example of how one change (the introduction of the metaphor) leads to another change (the substitution of 'I' with 'the director' on the inserted page. Although this order of the revisions is plausible, it can only be a hypothesis. However, if we now imagine that Elsschot's writing process took place on a computer and that he also used a keystroke logger, we could deduce this revision episode and its exact order from the keystroke logging data.

For a simple example of a change inspired by another change we may turn to Bogaert's writing process. In Session 197, Bogaert is working on a sentence about a dead crow in the first chapter of *Faraaz*:

> G[109]5[108]~~Eerg~~[87][110]g[86]s[88]ister~~en~~[111]en[112][89]ochted[90]nd[91] [92] lag er een doo[93]de kraaib[105]~~n~~[106]ij het t~~g~~[107]erras van HotelT[94][95]~~t~~[96]erminus ([96]hiertegenover)[9785]
> (Session 197)

Then he scrolls up in the text to Janine's chapter – who talks about the same bird – where he replaces "merel" ["blackbird"] with "kraai" ["crow"] in the sentence: "De dode merel ligt nog altijd in het keldergat".[117] Then, in Session 198, Bogaert first adds the word "jong" ["young"] between "dode" ["dead"] and "kraai" ["crow"] in Faraaz's chapter – "een dode jonge kraai"[118] – then he scrolls up to Janine's chapter to type "enjong" after "dode kraai": the 'dead crow's young'. Thus, the keystroke logging data allows us to reconstruct that the changes in the chapter of Faraaz immediately lead to changes in the chapter of Janine – in this case, we can state that he made those revisions with the intention to avoid anomaly. Schmidt points out that "[c]hanges that are local to one revision-site are less likely to be connected with changes elsewhere in the same manuscript the further apart they are. A change at the start of a document is very probably (although not certainly) unrelated to a change at its end" (Schmidt 2019, §30).

---

[117] Translation: "The dead blackbird lies still in the cellar hole".
[118] Translation: "A dead crow's young".

However, this example illustrates that it is in fact possible that revisions that are made in very different locations of the text are directly related to one another and are also even carried out in close proximity in time. That we are able to establish this connection in time is the unique possibility that keystroke logging brings us.

By discussing examples of different revision episodes, we have seen how using the concept of revision episodes can guide our interpretations of revisions in a literary context. Because we can now use keystroke logging to examine the context of revisions in terms of the process – for example, what writing action occurred before or after the revision – we can have a more thorough understanding of what might have triggered the revision, or even: what might have been the intention of the revision. The three types of revision episodes identified by Kollberg (1998) – episode with repetitive revisions at one cursor location, episode with embedded revisions and, episode with a sequence of revisions in previously written text – proved useful for a genetic analysis of the authors evolving (micro-)intentions. The revision episodes at one cursor location are useful for interpreting the small shifts in intentions-to-do in the initial composition of a sentence. Episodes with embedded revision, tend to accentuate the author's intention-to-do, such as finding the most fitting word, and mark hesitations during writing. An episode with a sequence of revisions in previously written text address how intentions-to-do can be invoked by the text that is currently produced. In Williamson and Pence's typology (1989), these episodes with a sequence of revisions in previously written text are sub-divided into 'changes related to a global change' and 'changes inspired by another change'. This distinction is valuable for genetic criticism because it takes into account the impact of revisions on the text. By comparing these revision episodes in the keystroke logging data with analyses of analogue writing processes that focus on similar revisions, we see how keystroke logging can enrich hypotheses in genetic criticism.

4.3.3 Triggers and re-reading

With regard to revision and intention, we can hardly ignore Sally Bushell's micro-reading of Wordsworth's *Prelude*: the "micro-*Prelude*" as she calls it. In this analysis, she breaks the compositional process down into its smallest parts, which would "allow for a refinement of our understanding of the nature of composition as a process that is *both* mental and physical by approaching it in terms of a sequence of intentional acts" (Bushell 2009, 90). She uses John Searle's account of intention as outlined in his book *Intentionality* (1983) for this purpose, where he "looks closely at the relationship between intention as a mental state and the action that results from it and argues for an "intention in action" that forms part of every act, however small" (Bushell 2009, 91).

With her study, Bushell "envisages a movement between willed, conscious intention and a giving up of the self to language (in the state of creative composition)",

while it "assumes a return to conscious intention and a reclaiming of perceived ownership of language by the writer through acts of revision (which might also lead to a further giving up on the self, and so on)" (2009, 91). To exemplify this, she looks at the following revised sentence in Wordsworth *Prelude*:

<div align="center">
ridge cliff alone
While on the perilous ~~edge~~ I hung (DC MS 19 (MS JJ), X^v)
</div>

Bushell reconstructs the composition process as follows:

> we could assume that Wordsworth begins by rereading the original line and finds himself dissatisfied with the word "edge," so he crosses it out and replaces it with one, or two, alternatives. In its smallest component parts, the process thus runs: prior intention (I intend to cross the word out); intention-in-action (I am about to cross the word out); bodily movement (pick up pen/place pen on paper); action (physically mark a line through the word). In this particular example the sequence is either immediately followed by, or perhaps coterminous with, a second sequence: prior intention (I intend to replace an unsuitable word with another); bodily movement (place pen on paper); action (physically enter the word "cliff" on the page). (92)

In literary-critical terms, Bushell argues, this sequence of revisions results in changes of meaning. Wordsworth most likely deemed the word "edge" too specific and narrow and therefore changed it into the word "cliff", which creates "a more topographical sense of the context and scale" (Bushell 2009, 92). In other words, the specificity of the word "edge" was the trigger for the revision. This then prompted – triggered – another "intentional sequence" that led to the addition of another option to replace this word with: "ridge", that enlarges "the context and sense of place" even more (92).

Bushell states that it would be impossible to undertake this level of microanalysis at any great length, but she argues that it is "helpful to be *aware* of it and to see that the creative process for written composition is capable of being broken down to this extent" (92). According to her, such a reconstruction can be done with "reasonable accuracy", although she acknowledges that "we still cannot be sure of the amount of time occurring between any of the changes (apart from the evidence provided by changes of ink or hand), and we cannot access the all-important initial point of composition" (92).

What Bushell's reconstruction also points to is the importance of rereading within this revision process. Reading is important in the writing process: "writers move back and forth between writing and reading as they produce text; they reread to detect coherence problems and to proofread" (Haas 1996, 55). Logically, reading is also part of models of the writing process (Flower et al. 1987; Leijten et al. 2014). Reading can involve processing the text on many levels "from decoding words and identifying grammatical structures to identifying gist and inferring the writer's intentions" (Hayes

2004, 12). The revision model by Flower et al. (1987) focused on revision that occurs while the writer is reading the text produced so far and therefore distinguished three modes of reading: "*reading to comprehend*, *reading to evaluate*, and *reading to define problems*" (Hayes 2004, 12). Quinlan et al. point out that "experienced writers are in constant visual contact with their own text", the text produced so far can namely function as a basis "for building internal coherence, generating ideas, and/or visualizing possible revisions" (2012, 350). In short, the text produced so far can function as "a catalyst for invention" (Leijten et al. 2014, 325).

For the keystroke-logged writing process, we do in fact have access to the initial point of composition, and in my analysis below, I link this to the importance of rereading as a trigger for both the production of new text and revision. Take, for instance, the writing process of the sentence "Mijn geschuifel langs de heiligenbeelden heeft veel weg van het getrippel van een muis"[119] in Session 18 of Troch's writing process. At a certain point in this session, Troch is revising a paragraph about a memory the narrator has of his father, who told him that his mother could sleep through every noise, while he already wakes up to the pitter-patter of a mouse, or as he says in Dutch: "Het getrippel van een muis krijgt me zelfs wakker".[120] Then Troch writes "Trippelen". Since we can see that Troch wrote "Trippelen" after he had been working on this paragraph – although he did visit YouTube and social media in the meantime – we can argue that re-reading this sentence of the father triggers the idea of a new sentence. He has the intention to do something with the word "Trippelen", but what? Ten seconds later he has an idea. In front of the word, he adds: "Mijn heeft veel weg van het ge" (n99).[121] And at the end of the sentence, he adds "van een muis" (n102),[122] after which he corrects a typo. A word is still missing in the sentence, what is exactly very much like the pitter-patter of a mouse?

> M~~I~~[104]i[105]jn heeft veel weg van het ge[9998]~~T~~[100]rippel~~en~~[101] [102] van een muis. (Session 18)

He opts for "geschuifel" (n106) – shuffling' – a word he used before to describe the walking movement of the narrator. Then Troch visits social media and checks his mail again. When he returns, he fixes a typo and then adds "langs de heiligenbeelden" (n110),[123] to remind the reader that the narrator is in a church.

---

[119] Translation: "My shuffling past the statues of saints is very much like the pitter-patter of a mouse".

[120] Translation: "The pitter-patter of a mouse even gets me awake".

[121] Translation: "My is very much like the".

[122] Translation: "of a mouse".

[123] Translation: "past the statues of saints".

M~~I~~[104]i[105]jn geschuifel [106]langs de heiligenbeelden [110]heeft veel weg van het ge[9998]~~T~~[100]t[109]rippelen[101] [102] van een muis. (Session 18)

This moment of initial composition of the sentence already shows the engagement with the text produced so far, since the re-reading of the paragraph triggered the writing of the new sentence in the first place.

The sentence remains the same for a couple of sessions. Then, in Session 23, he adds another new sentence to this paragraph which builds upon the previous sentence: "En ik die dacht dat ik een kat was" (n399).[124] The narrator compared himself earlier in the story with a cat, since he licks the sweat of his skin to avoid dehydration. And perhaps not even that coincidentally, Troch revised a sentence in which the narrator makes this comparison at an earlier point in this writing session. With this cat comparison fresh in mind, he thus later arrives at the mouse comparison sentence. This may have made him realise that it is perhaps somewhat contradictory to compare the narrator to both a cat and a mouse, those two traditional opponents. This, we may argue, could well be the trigger for the addition of the sentence "And I who thought I was a cat".

By elaborating on the word "trippelen", Troch created coherence at the local level of the text, at the level of the paragraph, but by now repeating the comparison with the cat, the coherence of the text as a whole is also maintained. Instead of keeping the comparison with both a cat and a mouse in the text, the comparison with the cat is now refuted. The narrator thought he was like a cat, but in fact, he is not. Yet the comparison with the cat earlier in the text is very apt. We may regret that it is omitted at this point in the text.

If we continue this train of thought, we might be able to understand the revisions in the next session. In this session, Troch substitutes "veel" ["very much"] with "niets" ["nothing"] (n313-314). This small substitution changes the meaning of this sentence completely, his shuffling along the statues of saints is now nothing like the pitter-patter of a mouse. In its current form, the sentence still builds upon the preceding one, but the comparison with the mouse is disproved instead of confirmed. This makes it possible to reiterate the comparison with the cat. The narrator calls himself a cat, a purebred tomcat. Troch briefly checks the definition of "kater" online. Then he continues revising this particular sentence to make it more specific. Instead of being a cat, the narrator now compares his shuffling with that of a cat. However, the word shuffling is perhaps not the most fitting to describe the pace of a cat, this might be the reason why, six seconds later, he substitutes the two instances of "schuifel" ["shuffling"] with "sluip" ["creep"] (n328-329).

---

[124] Translation: "And I who thought I was a cat".

The analysis of the writing process of this sentence illustrates the constant interaction with the text produced so far. While the writing of this sentence itself was triggered by rereading the text, the revision of it was also influenced by the process of rereading the text in general. This may seem to suggest that, while producing new text, Troch was not always aware of what he had written before. Only after rereading the text he could make connections between different paragraphs, connections which he had perhaps not thought of before. Nancy Sommers argues that experienced writers "seek to discover (to create) meaning in the engagement with their writing, in revision" (Sommers 1980, 386). This is also an important aspect of literary writing, in which the text emerges through a constant engagement with what has already been written. Or as Bryant puts it: "Revision is the writer-as-reader's rewriting" (Bryant 2002, 98). All the writing actions that eventually led to the sentence reveal different writing goals, different intentions for the text, from first creating local coherence by building upon preceding sentences, to elaborating this to coherence in the text at the larger level by repeating the cat comparison.

4.4 CONCLUDING REMARKS

Through a comparison of the properties of digital revision distinguished in writing studies with how revisions have been discussed with regard to analogue writing processes, the process-oriented properties seemed to be most valuable for this current study, and for enriching the theory of genetic criticism in general. Genetic criticism has always been concerned with the process of writing, but the analogue nature of the material has made an in-depth study of time in writing unfeasible. This is where keystroke logging comes in, yet the keystroke logging data appeared not to be immediately suitable for a text genetic analysis. The methods from digital scholarly editing offered the solution for capturing only the relevant information from the keystroke logging data and making it easy to read and familiar to peers. The use of these methods, in particular the use of TEI-XML, in combination with a focus on the process-oriented properties is then the option of choice, as it leaves room for later interpretation.

The spatial location therefore served as a starting point for the encoding of the keystroke logging data in TEI-XML, since the writing actions they represent are also known in textual scholarship (e.g. pre-contextual revisions relate to *currente calamo* revisions). The encoding models the writing process at the scale of the writing session and the combination of the Word documents (the session versions) with the keystroke logging data (the process) serves to uncover a text that had become invisible during the writing process because of the overwriting nature of word processors. The visualisations demonstrate that it is possible to achieve a genetic transcription of keystroke-logged writing processes. This transcription also makes it possible to represent all the different actions that took place while the text was being written. By adding the time to the

revision, the sequentiality of all the revisions can also be reconstructed, which enables a dynamic visualisation for a detailed analysis of the way the author moved through the text and how sentences were produced. According to Elena Pierazzo, such a scholarly consideration of time plays a pivotal role in the case of modern autograph drafts and working manuscripts because: "the stratification of corrections, deletions and additions can give insights into an author's way of working, into the work itself, the evolution of the author's *Weltanschauung*, the meaning/interpretation of the text" (2009, 171). Compared with analogue text genetic material, the keystroke logging data encompasses detailed information about the process underlying text production. Logging writing processes with a keystroke logger enables an analysis of the textual genesis at a finer granularity.

This, in turn, provides a sound basis for investigating triggers of revision. The cause of revision, the trigger, can be related to the author's "intention to do" (Shillingsburg 1996). In writing studies, the word intention is often used in descriptions of the cause of revision, and whilst intention tends to be avoided in literary studies, textual scholars can hardly escape the use of the term. Using the term to focus on the reasons for doing something during the writing process, as intentions to do, provides us with the terminology to hypothesise about the reasons behind certain revisions. It becomes clear that digital writing technologies can still be a trigger for revision. By focusing on how Troch revised a paragraph by interweaving it with new sentences, it became visible that contemporary writers still revise by extension: only not after a substantial period of time, but immediately in the composition process. This revision was probably driven by a shift in intention to do. The data from the keystroke logger makes Troch's revision strategy visible again. However, it also demonstrated that this revision strategy was only possible by using a word processor. The different categories of revision episodes proved to provide insights into the (changing) intention-to-do. From shifting intentions during initial production (e.g. from what to write to how to write it), to narrowing down the author's intentions-to-do, to how such intentions are initially evoked. Another example from Troch's writing process showed how initial text production and revision can both be triggered by the act of rereading the text produced so far. This relates to the discovery of new elements to implement through rereading the text and through writing.

John R. Hayes points out that "revision may be initiated by the discovery of opportunities as well as by the detection of problems" (2004, 13). Yet, he emphasises that we need to gain a better understanding in the role of discovery in initiating revision. He states that research has primarily looked at "revisions that fix problems", but he deems "revisions that are stimulated by the discovery of new connections, new ideas, or new arguments" more interesting, as they can be associated more with the substance of the text and mark the moments "the writer learns something through the act of writing" (Hayes 2004, 20). Since, research on discovery through writing has only

advanced (e.g. through the work done by Galbraith and Baaijen 2018; Baaijen and Galbraith 2018). In my opinion, the text-genetic approach, due to its text-focused nature, could be applicable on these questions around discovery in writing as well. In the following chapter, I focus on how genetic criticism can be enriched with a detailed temporal dimension. Now that ambiguity is no longer omnipresent in an analysis of the order of production of the text, I discuss the possibilities of analysing this aspect of discovery in literary writing processes. For if there is one thing that the closer examination of revisions with a focus on intentions emphasises, it is the importance of the temporal dimension of writing.

CHAPTER 5. THE TEMPORAL DIMENSION OF WRITING

DISCOVERING NEW ASPECTS OF TEXTUAL GENESIS

Ik tik zo langzaam en zo amateuristisch dat ik dus ook niet mezelf voorbij kan rennen. Ik kan nooit sneller tikken dan mijn hersenen gaan. Dat kunnen de meeste mensen wel. Ik maak dus van de nood dat ik niet snel kan tikken de deugd dat ik kleine verhaaltjes eruit haal. Uit die machine. (A.L. Snijders in Conijn 2012, 00:18:22-00:18:44)

Applying genetic criticism to keystroke logging data immediately allows for, or almost demands, an interdisciplinary view on the writing process – the word processor thus serves as a mediator between the two fields. In this study, the methodology of genetic criticism is complemented by theories developed in cognitive writing process research. In this way, a better understanding can be gained of how keystroke logging may change, or even enrich, the theory of genetic criticism. The main point of investigation, in this chapter, is the detailed temporal dimension of writing. In the quote above, the writer of (very) short stories, A.L. Snijders, refers to the pace of writing. He says that while other writers may type so fast that their brains can barely keep up, he himself is a slow typist, which allows him to think before he writes. Such statements can easily be verified by examining the temporal dimension captured by keystroke logging.

As Christophe Leblay and Gilles Caporossi point out, pen and paper and screen-keyboard are both technologies that highlight the phenomena of writing; both, in their own way, reveal certain characteristics of writing that are not mutually exclusive (Leblay and Caporossi 2014, 7). However, they add that the combination keyboard-screen as a writing technology – in cases in which the writing process can be replayed, such as with keystroke logging – makes it possible to highlight important characteristics of writing, such as time (7). The ability to analyse time, they argue, allows scholars to really study *production*. For their definition of textual production, Leblay and Caporossi rely on the

definition proposed by Ganasica, Fenoglio and Lebrave (2004): "le résultat d'un processus de production qui, comme tel, est inscrit dans une temporalité" (Ganascia, Fenoglio, and Lebrave 2004, 92).[125]

Leblay and Caporossi point out, quoting Hay (2009), that just as with pen and paper, our observation of the production of writing can only begin 'on the moment when the pen touches the paper' (13). This remains true for keystroke logging, as we are still only recording actual keystrokes on the keyboard or mouse movements, and are therefore still confronted with the second, inner black box mentioned above (Chapter 2). Nevertheless, the enhanced investigation of the temporal dimension of writing made possible by keystroke logging gives us more 'evidence' with which to formulate hypotheses about the writing process. In this chapter, therefore, I examine in detail what a study of this temporal dimension can bring to the field of genetic criticism.

The first aspect of textual genesis addressed in this chapter is how the keystroke logging data can shed light on some of the authors' writing habits. How much time was spent at the computer for writing the text? Over how many sessions was the writing process divided? And on which days did these sessions take place? How many words were written per session, or even per minute? Such questions could not (always) be answered with handwritten texts, but now they can be answered on the basis of the keystroke logging data. Handwritten manuscripts could reveal revision habits, but now we can examine these habits throughout the writing process: for example, do the types of revision change as the writing process proceeds? The second aspect is the new type of text genetic research made possible by keystroke logging, which I termed the nanogenesis. The nanogenesis examines the movements through the text by means of the sequentiality of the writing actions. I show how this provides insights into the connections between revisions, allowing new interpretations of the text and the writing process. Through the nanogenesis, we arrive at the third aspect fortified by the temporal dimension, which is discovery in writing. By studying the nanogenesis, it may be possible to find moments in the writing process that relate to discovery. The fourth aspect in which the temporal dimension can enrich genetic criticism concerns the exogenesis. The keystroke logging data contains all the websites and other 'windows' the author has opened while writing, which can be used to determine what external sources have been used. I use a tree as a metaphor to discuss how the keystroke logging data offers new possibilities for studying the use of sources in the writing process. For example, what are the roots of the tree (the sources used for the text)? How many roots are there? And how can we follow their transformation?

---

[125] Translation: "the result of a production process which, as such, is inscribed in a temporality"

5.1 CHRONOLOGY AND GENERAL WRITING HABITS

Genetic criticism, as Van Hulle points out, "incorporates the temporal dimension in literary studies", which makes chronology an important aspect of the genetic dossier (Van Hulle 2022, 60). Sometimes the manuscripts are dated by the author – or in some cases the author has even dated each writing session – which makes it feasible to reconstruct an 'absolute' chronology (60). However, this is usually not the case, so that only a 'relative' chronology can be derived by comparing versions (60). In the case of a digital genetic dossier, the chronology of the files can be reconstructed more easily, but even then, caution is required. For example, the 'last modified' date will change if the file is saved again after a minor change, such as deleting a space, or can be affected simply by opening the file, or the computer's clock can be set incorrectly (Kirschenbaum 2016, 229). Even the timestamps of keystrokes in keystroke logs can contain strange errors. However, this does not detract from the level of detail with which the chronology can be studied; with few exceptions, the order of writing sessions, revisions, and individual keystrokes can be determined without much effort. Based on the metadata (e.g. the start time and duration of the sessions) and the revisions, we can gain new insights into the working methods and writing habits of the authors.

Siegfried Scheibe describes work methods as "the general and specific form of an author's creative process of production, that is, the concrete way and means by which an author – whether writer, composer, scholar, or whatever – labors on a work, prepares it, conceives it, writes it down, and revises it" (1995, 171). These work methods will reveal specific recurring habits, among other things: "the use of different kinds of paper at different stages of production of a work", "the different ways of using this paper", "the way in which the author covers this paper with different writing instruments", "the manner of corrections" or "the way the corrections are related to the existing text" (171). Such work methods, or writing behaviour, play an important role in the model of writing developed by Zimmerman and Risemberg (1997), which focuses on self-regulation in writing. By self-regulation, Zimmerman and Risemberg refer to "self-initiated thoughts, feelings, and actions that writers use to attain various literary goals, including improving their writing skills as well as enhancing the quality of the text they create" (1997, 76). These processes can, according to Zimmerman and Risemberg, be grouped in three categories of self-regulatory influence:

1. environmental processes: "writers' self-regulation of the physical or social setting in which they write" (76);
2. behavioural processes: "writers' self-regulation of overt motoric activities associated with writing" (76);
3. personal processes: "writers' self-regulation of cognitive beliefs and affective states associated with writing" (77).

Corresponding to these triadic influences, they also distinguish three classes of self-regulation: environmental self-regulation ("the adaptive use of an context-related strategy"), behavioural self-regulation ("the adaptive use of a motoric performance strategy"), and covert (personal) self-regulation ("the adaptive use of cognitive or affective strategies") (Zimmerman and Risemberg 1997, 77). As they use these strategies, writers reflect on their use of them, and decide whether adjustments are favoured to improve their writing.

Most forms of self-regulatory behaviour can only be studied on the basis of writers' personal accounts of their writing behaviour and preferences, of which there are many: for example, the interviews with present-day Dutch poets in *Een mogelijk begin van veel* (2021), by Van Hasselt and Sistermans. As in-depth interviews were not part of the Track Changes project, we do not have such detailed accounts of the writers to be able to study their behavioural strategies. However, I was able to outline some habits based on personal conversations I had with some of the authors. During an interview at the Track Changes symposium (2022), for example, Roos van Rijswijk stated that she regularly reads her text out loud, and Jente Posthuma explained that she usually writes (or tries to write) around three hundred words per writing day. And Gie Bogaert, as mentioned above, strictly adheres to the distinction between the 'creative process' (in the Atoma notebook) and the 'linguistic creative process' (in MS Word). Furthermore, the temporal aspect of the writing sessions allows us to examine whether general writing habits can be distinguished regarding the duration of the writing sessions, the time of writing, the writing days (all planning and management), and the revision habits (cognitive strategies) – specific, of course, to the writing processes they logged with Inputlog. But whether the writers also literally immersed themselves physically in the topic of the text they were working on, as Miek Zwamborn did, remains for our imagination:

> Door fysieke inspanning komt de taal los. Ik stuur mijn lichaam als een drone vooruit om er woorden mee op te sporen. Voor het schrijven van mijn laatste roman, *De duimsprong*, verdiepte ik me in geologie en leerde ik bergbeklimmen. Voor *Vallend hout* liep ik een jaar lang met een boomchirurg mee. Mede door het vertalen van *Honingprotocollen* van Monika Rinck hebben mijn vriend en ik nu een bijenvolk. (Zwamborn in Hasselt and Sistermans 2021, 84)[126]

---

[126] Translation: "Physical exercise frees language. I steer my body ahead like a drone to track words with it. Before writing my last novel, *De duimsprong*, I immersed myself in geology and learned mountaineering. For *Vallend hout*, I accompanied a tree surgeon for a year. Partially thanks to translating *Honigprotokolle* by Monika Rinck, my boyfriend and I now have a bee hive".

5.1.1 Duration of the writing process and sessions

The first temporal dimension that can be deduced from the metadata of the keystroke logging files, is the duration of the entire writing process and that of the individual writing sessions. Gie Bogaert wrote *Roosevelt* in 259 days, for 277 hours, 14 minutes and 24 seconds, distributed over 422 writing sessions. At least, this is the time that was logged with Inputlog. Bogaert also clarified that he occasionally wrote in Evernote on his tablet when he was unable to write on his desktop PC. The text he produced on other devices was later pasted into the Word document, but the composition process of these fragments was therefore not logged. As a result, the writing process was longer than the data above would suggest – this is one of the examples that proves that we should always be critical of the metadata. Keystroke logging provides a lot of data, but much is left 'unlogged'.[127]

	BOGAERT	POSTHUMA	VAN RIJSWIJK	TROCH	VAN PELT
Kind of text	Novel	Short story	Short story	Short story	Short story
Number of writing sessions	422	37	7	20	18
Number of days	259	29	5	8	13
Duration of the writing process	277:14:24	64:41:55	11:48:01	27:48:05	15:19:26
Average duration of sessions	00:39:25	01:44:55	01:41:09	01:23:24	00:51:05
Number of words in the final text	34105	3489	3449	4115	3017
Average word production per session	132	357	1101	440	342
Average word production per minute	3,454	4,223	10,675	5,492	6,315

Table 5.1: Overview of some specifics of the writing processes.

---

[127] I opted not to remove long pauses from the data, partly because in some cases they may well have been part of the writing process, and partly because they can (but only very limited) vouch for the large portions of the writing process that were not logged. After all, the writing process is not limited to the moment the author sits down behind the computer.

The duration of the logged writing process of each of the short stories is, of course, considerably shorter. The longest writing process of the short stories was that of "En daarom haten ze zichzelf", which took Jente Posthuma 29 days, and 38 sessions, to write. The total duration of the writing process was 64 hours, 41 minutes, and 55 seconds. The writing process of Roos van Rijswijk's "Zorgvlied" was the shortest of all the writing processes discussed in this study. Van Rijswijk wrote the story in 7 sessions across five days. The total duration of the writing process was 11 hours, 48 minutes and 1 second. David Troch took three days longer to write "Mondini", but during those eight days he logged considerably more sessions than Van Rijswijk. The story was written in 20 sessions, totalling 27 hours, 48 minutes, and 24 seconds. Ellen Van Pelt's logged writing process was shorter – 15 hours, 19 minutes, and 26 seconds – but she wrote "Dauphin" in more days: 18 sessions were logged over thirteen days.

Then there is the average duration of the sessions, which gives an idea of how the writer approached writing. For Bogaert, the average duration of all the sessions was 39 minutes and 25 seconds; the shortest session lasted only 28 seconds, while the longest lasted for 4 hours, 34 minutes and 50 seconds.[128] So Bogaert wrote mostly in short writing sessions. More than half of his writing sessions (248 sessions) lasted less than half an hour: 149 sessions lasted only between 0 and 15 minutes, and 99 sessions lasted between 15 and 30 minutes. The most common duration of the writing sessions for all authors is between 30 and 45 minutes. For Van Pelt, the largest percentage of sessions (33%) fell into this category, although the average duration of her writing sessions is 51 minutes and 5 seconds. Van Pelt's shortest session was 17 minutes and 53 seconds, and her longest was 1 hour, 19 minutes and 22 seconds. Another common duration for these authors is between one hour and one hour and fifteen minutes. 40% of Troch's writing sessions where of this duration, but his average was slightly longer: 1 hour, 23 minutes and 24 seconds. The duration of the sessions ranged from 12 minutes and 2 seconds to 2 hours, 43 minutes, and 10 seconds. The average durations of Van Rijswijk and Posthuma are more or less the same: respectively 1 hour, 41 minutes and 9 seconds and 1 hour, 44 minutes and 55 seconds respectively. Van Rijswijk's 7 sessions all varied in length, ranging from 31 minutes and 47 seconds, to 3 hours, 21 minutes, and 40 seconds. Posthuma's 37 sessions varied from 7 minutes and 50 seconds to 4 hours, 5 minutes, and 47 seconds. Most of Posthuma's sessions (19%) had a duration between 2 hours, and 2 hours and 15 minutes.

One thing that becomes clear is that none of these authors chose to write in very long sessions. Only 4 sessions – one by Posthuma and three by Bogaert – were longer than four hours. This does not necessarily imply that they only wrote for a short time

---

[128] One session in the genetic dossier, Session 275, even gives 00:00:00 as recorded time. I considered not including it in the genetic dossier, but some events were indeed logged (replacements and insertions), so I decided not to ignore it.

DURATION	BOGAERT	POSTHUMA	VAN RIJSWIJK	TROCH	VAN PELT
00:00–00:15	149 (35%)	2 (5%)		1 (5%)	
00:15–00:30	99 (23%)	2 (5%)			2 (11%)
00:30–00:45	57 (14%)	6 (16%)	1 (14%)	3 (15%)	6 (33%)
00:45–01:00	36 (9%)	2 (5%)	1 (14%)		4 (22%)
01:00–01:15	17 (4%)	3 (8%)	1 (14%)	8 (40%)	4 (22%)
01:15–01:30	16 (4%)	1 (3%)	1 (14%)	1 (5%)	2 (11%)
01:30–01:45	12 (3%)		1 (14%)	1 (5%)	
01:45–02:00	8 (2%)	1 (3%)		2 (10%)	
02:00–02:15	8 (2%)	7 (19%)		1 (5%)	
02:15–02:30	4 (1%)	5 (3%)		1 (5%)	
02:30–02:45	3 (1%)	1 (3%)		2 (10%)	
02:45–03:00	2	3 (8%)	1 (14%)		
03:00–03:15	3 (1%)	2 (5%)			
03:15–03:30	1		(14%)		
03:30–03:45	3 (1%)	1 (3%)			
03:45–04:00	1				
04:00>	3 (1%)	1 (3%)			

Table 5.2: Duration of the writing sessions.

in a day; the writing time in a day was just distributed over several writing sessions. This could well be an effect of logging the process with Inputlog. Logging their process may have made the authors more inclined to stop recording when they stopped writing to do something else for a moment. Only Posthuma left her laptop for a significant period of time while a writing session was still in progress; on these occasions, she activated the lock screen. This happened twenty times, for a total period of 9 hours, 48 minutes, and 50 seconds.

In addition to the duration of the writing sessions, the time at which the sessions were started also provides information about the writing habits during these specific writing processes. Bogaert started most of his sessions (108 sessions) between 3 p.m. and 6 p.m., followed by 92 sessions that started between 6 p.m. and 9 p.m. When the duration of these sessions is also considered, it becomes clear that the sessions he started between 9 p.m. and midnight were relatively short: these sessions had an average duration of 23 minutes and 32 seconds. However, when he started a session between noon and 3 p.m., these sessions were relatively long, with an average duration of 57 minutes and 25 seconds. All 3 sessions lasting more than four hours started between noon and 3 p.m.
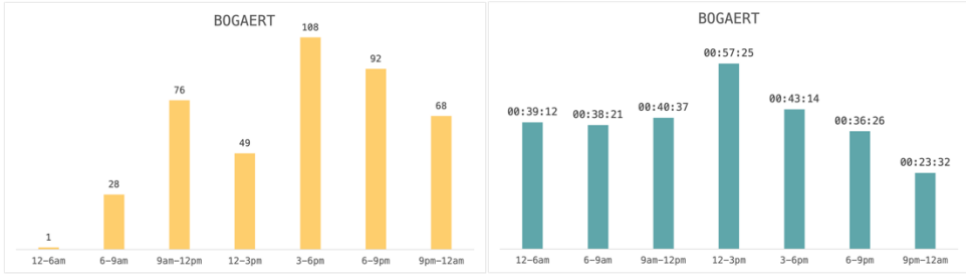
Figure 5.1: Distribution of writing sessions of Bogaert, regarding the start time of the session and the average writing time in these categories.

Figure 5.2: Distribution of writing sessions of Posthuma, regarding the start time of the session and the average writing time in these categories.

Figure 5.3: Distribution of writing sessions of Troch, regarding the start time of the session and the average writing time in these categories.

Figure 5.4: Distribution of writing sessions of Van Pelt, regarding the start time of the session and the average writing time in these categories.
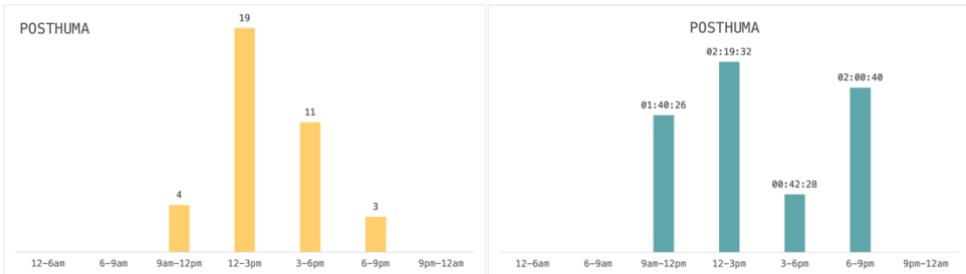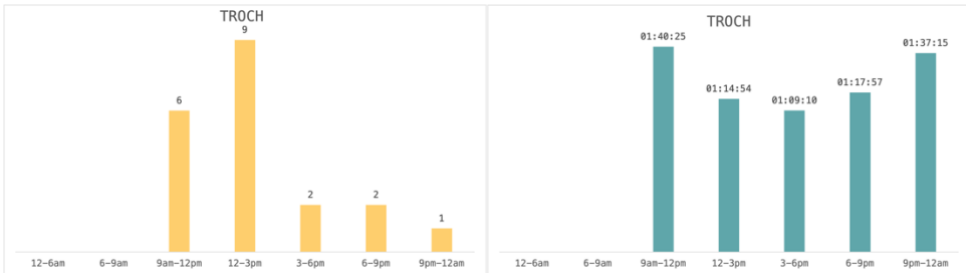
Figure 5.5: Distribution of writing sessions of Van Rijswijk, regarding the start time of the session and the average writing time in these categories.

Posthuma started the most sessions (19 sessions) between noon and 3 p.m., followed by 11 sessions that started between 3 p.m. and 6 p.m. She also started 4 sessions between 9 a.m. and noon, and three between 6 p.m. and 9 p.m. The average duration of these sessions, however, indicates that the sessions that were started between 3 p.m. and 6 p.m. were relatively short. Troch and Van Pelt also started most of their sessions between noon and 3 p.m. Van Rijswijk's writing sessions are divided into periods starting between 9 a.m. and 6 p.m., with one session starting at 8 p.m. This session is, however, the longest of all her sessions. This makes it clear that Posthuma, Troch, Van Rijswijk and Van Pelt wrote mostly during 'regular' working hours, and not early in the morning or late in the evening.

5.1.2 Writing days and dates

In addition to time preferences, the temporal dimension also makes it possible to examine preferred writing days, and periods. The most extensive analysis can be done for Bogaert, as his writing process stretched over several years, from July 2013 to December 2015. Three periods can be distinguished in which Bogaert wrote quite regularly: from September 2013 until February 2014, June, July, and August 2014 and from March 2015 up to and including October 2015. In both March and May 2014, Bogaert only wrote on two days, although the sessions in May were slightly longer. April 2014 was somewhat more productive with 3 hours, 26 minutes and 57 seconds of writing time recorded.

The month in which Bogaert spent the most time writing was August 2014 – Inputlog was activated for 36 hours, 12 minutes and 52 seconds (39 sessions; 20 days) – closely followed by July 2015 (34 sessions; 20 days; 33 hours, 30 minutes, and 23 seconds) and October 2015 (43 sessions; 21 days; 29 hours, 43 minutes, and 29 seconds). The summer months were therefore the most productive for Bogaert. The productive summer of 2014 was followed by a period in which Bogaert only wrote very little – only 4 writing sessions in five months.

The above reflects Bogaert's personal life. Bogaert taught at a secondary school, and the holidays, especially the summer holidays, were particularly productive. And the
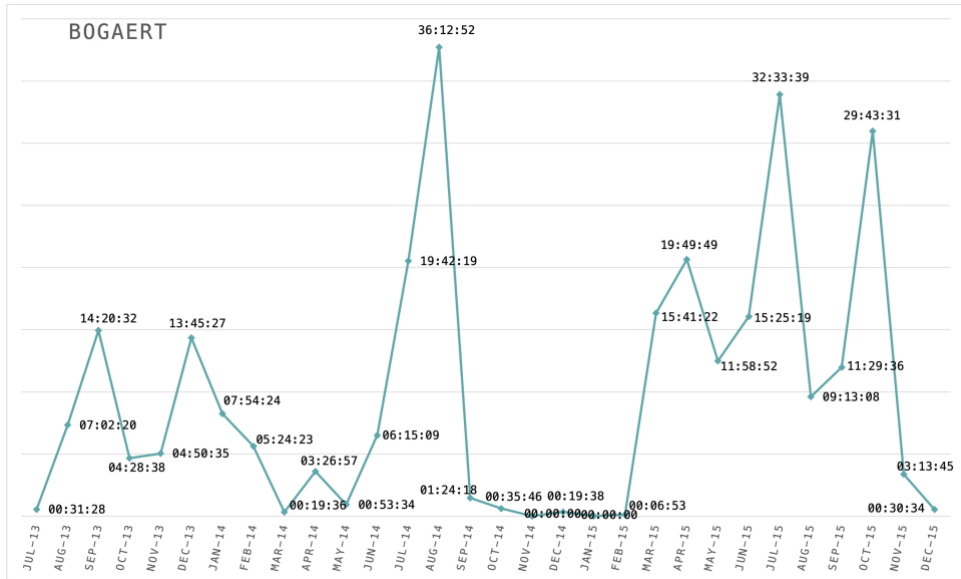
Figure 5.6: Total writing hours per month for Bogaert.

months in which Bogaert wrote little or almost nothing were the periods in which he moved house. This shows how the writing is influenced by the daily activities of the writer. Of the short story writers, only Posthuma and Van Pelt's writing processes took place over several months. Posthuma wrote four days (4 sessions) in August 2020, sixteen days (19 sessions) in September, and nine days (14 sessions) in September. Van Pelt wrote three days (4 sessions) in November 2020, eight days (10 sessions) in December, and also two days (4 sessions) in January 2021, in which she mostly revised her text. Troch wrote "Mondini" mostly in August 2020, but 2 sessions took place in September during which he mostly revised the story. Van Rijswijk wrote "Zorgvlied" exclusively in December 2020.

When all the writing days are arranged according to the days of the week, we can see that Bogaert had a slight preference for writing on Mondays throughout the whole writing process: out of the 259 writing days he wrote 57 times on a Monday (22%).. Furthermore, the writing is more or less evenly distributed over the working days, with 13% on Tuesdays, 15% on Wednesdays, 15% on Thursdays, and 11% on Fridays. 24% of the writing days took place on a weekend day (9% on Saturdays and 15% on Sundays). This roughly corresponds to his teaching schedule: he was free on Monday afternoons and Wednesdays. Despite his work as a teacher during the week, we can see that this did not cause him to work more on the novel at weekends.
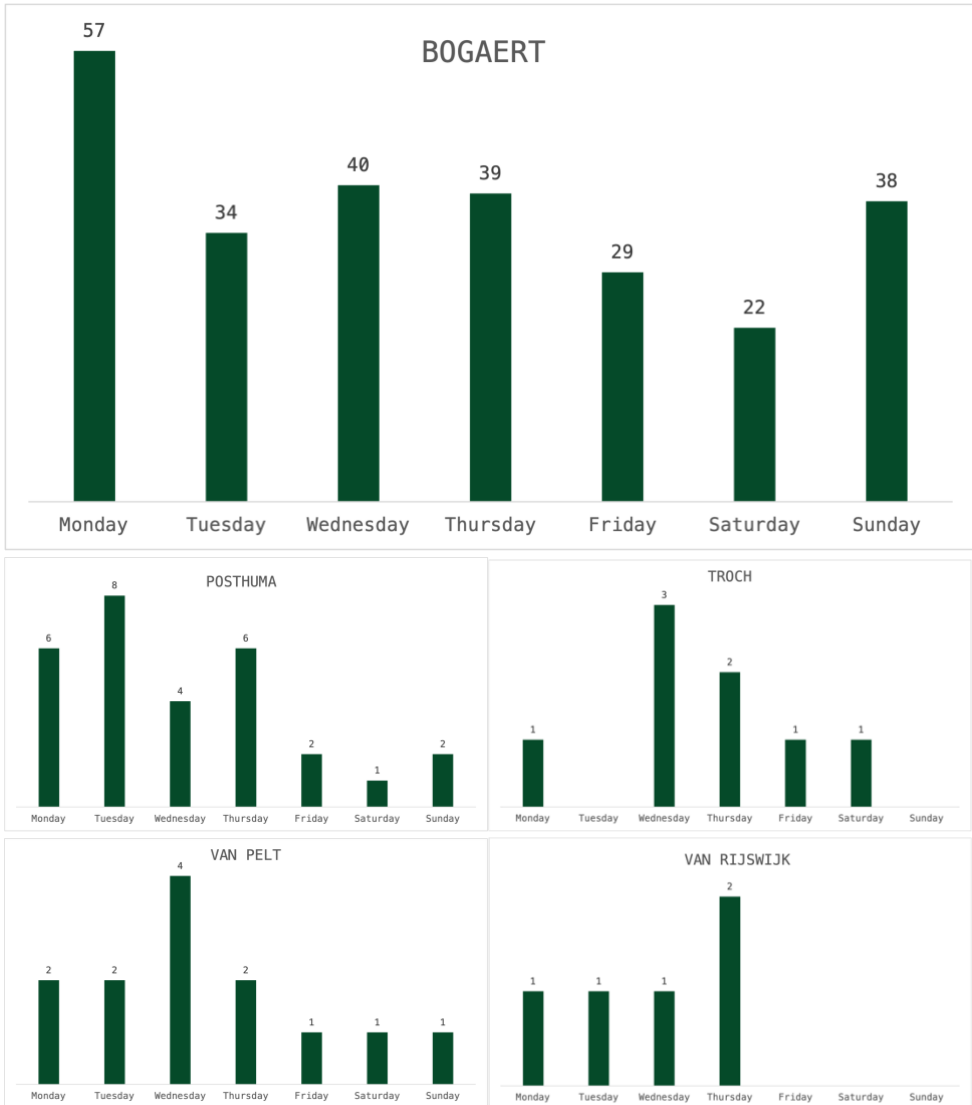
Figure 5.7: Number of writing sessions per day for Bogaert, Posthuma, Troch, Van Pelt & Van Rijswijk.

Posthuma wrote on every day of the week, but mostly on Tuesdays and Thursdays, respectively 19 hours, 14 minutes and 14 seconds, and 15 hours, 20 minutes, and 48 seconds. Van Pelt also wrote on every weekday, but had a slight preference for working on Wednesdays, with a total writing time of 4 hours, 53 minutes, and 24 seconds. Wednesday was also Troch's favourite day to write, with a total writing time of 10 hours, 3 minutes and 26 seconds. He did not work on "Mondini" on Tuesdays and Sundays. Van Rijswijk only wrote her story in one week, from Thursday to Thursday, but she did

not write on Friday, Saturday, and Sunday. The longest time that Van Rijswijk worked on her story was on Wednesday. On that day, she worked on the story for 6 hours, 17 minutes, and 23 seconds on the story in 2 sessions.

If productivity is measured only by the number of words produced during the writing sessions, it is also possible to extract the most productive sessions. Bogaert wrote an average of 132 words per session, yet in some sessions he produced many more words. In Session 283, on 22 April 2015, he produced 970 words, and in Session 405, on 6 October 2015, he wrote 908 words. He also produced more than seven hundred words in Sessions 150, 265 and 264. However, these sessions were also among the longest sessions, lasting between two and a half hours and four and a half hours.

Posthuma's three most productive sessions were on Tuesdays, which therefore coincides with her preferred day to write. In Session 17, on 22 September, she wrote 856 words; in Session 12, on 15 September, she wrote 880 words; and in Session 29, on 6 October, she wrote 883 words. These numbers of words are considerably more than the average number of words she produced per session, which was 357 words – this is in line with her general target of words per session/writing day, although the text did not grow as much as she also tended to delete text.

Van Rijswijk wrote an average of 1101 words per session. Her most productive session was the first session (Session 0), during which she wrote 2094 words. The other two productive sessions took place on the Wednesday, which was also the longest session. In these sessions, she produced 1425 and 1634 words respectively.

Troch's most productive session was Session 6, on Wednesday 12 August, during which he produced 1164 words. This session is followed by three other productive sessions. On Friday 21 August (Session 16) he produced 898 words, on Wednesday 19 August (Session 15) 897 words, and on Monday 10 August (Session 5) 849 words. These figures were also higher than the average of 440 words that Troch produced during each session.

The average number of words that Van Pelt produced per session was 342 words. The most productive session was Wednesday 25 November (Session 2), when she produced 960 words. She also produced over five hundred words in Session 3 (25 November 2020), 4 (30 November) and 16 (23 December).

However, this view on productivity does not take the aspect of time into account. The highest number of words is often produced in the longest sessions. The number of words produced per minute can give a different perspective on productivity. Bogaert produced an average of 3,454 words per minute. In Session 65, on 24 October 2013, he produced the highest number of words per minute: 12,4 words. In this session, he produced 73 words, but only in five minutes and 53 seconds. Ten days earlier, in Session 59, he also produced more than twelve words per minute (12,291). In Session 270, on 2 April 2015, Bogaert wrote 10,71 words per minute. On 20 October 2015 (Session

421), he wrote 10,416 words per minute, and on 4 September 2013 (Session 379), he wrote 10,101 words per minute. Three of these five most 'productive' sessions took place in the evening, Session 65 at 10:33:54 p.m., Session 270 at 10:08:40 p.m., and Session 355 at 8:56:43 p.m. Session 421, then, started at 08:31:01 a.m. and Session 59 started at 5:59:31 p.m. These all seem to be times when Bogaert wanted to add something to the Word document but did not have much time.

Posthuma produced an average of 4,223 words per minute. She produced the highest number of words in the first writing session (Session 0), on 3 August 2020 – a session in which she mainly wrote down notes on what to include in the story. In this session, she produced 20,914 words per minute. In Session 11, on 13 September, she produced 15,319 words per minute, and on 15 October (Session 32) she produced 7,854 words per minute. What these 'most productive' sessions have in common is that they were all started between 4 p.m. and 6 p.m., and they also belong to the shortest sessions.

The average number of words produced per minute by Van Rijswijk is 10,675 words. Like Posthuma, Van Rijswijk also produced the highest number of words per minute in the first writing session (Session 8), on 3 December 2020, in which she produced no less than 24,51 words per minute. In the session in which Van Rijswijk produced the second-highest number of words, on 8 December (Session 3), this number drops to 13,655 words per minute. And in Session 1, on 7 December, she wrote 10,169 words per minute.

Troch wrote an average of 5,492 words per minute. He produced the highest number of words per minute in Session 2 (10 August 2020), which was also the first writing session, in which he wrote 10,043 words per minute. In Session 14, 19 August 2020, Troch produced 9,059 words per minute. In Session 12, 12 August, 7,958 words per minute. The first writing session took place on a Monday, but the following three sessions with the highest number of words written per minute took place on Wednesdays.

Van Pelt wrote an average of 6,315 words per minute. In Session 3, on 25 November 2020, Van Pelt produced 14,549 words per minute. One session later, on 30 November (Session 4), she wrote 13,149 words per minute. And again, on 25 November, in Session 2, she wrote 12,237 words per minute. For the short stories it is noteworthy that almost all the sessions in which the authors produced the highest number of words per minute belong to the first couple of writing sessions. This could be related to the functions these sessions had in relation to the writing phases. In the first sessions, the focus is mainly on producing notes and writing down the first ideas. When revision becomes an important element in the writing process, the number of words produced in the session tends to decrease. This becomes clear when we examine the author's revision habits: in general, the production of new text decreases as the writing process progresses.

5.1.3 Habits of revision

Before moving on to the actual writing of the text based on the visualisations of the reconstructions of the writing processes (as discussed in Chapter 4.2), first a quantitative perspective on the author's revision behaviour. For all the authors, the most common type of revision is the contextual revision: revisions made in an already existing text (see Chapter 4.1). These revisions, after all, can be made throughout the entire writing process, whereas pre-contextual revisions can only be made during the production of a sentence. Lindgren and Sullivan point out that the use of pre-contextual revisions depends on the type of writer: "Depending on the type of writer, the revision process can be more or less interactive" (2006a, 171). They build on Galbraith's (1999) distinction between high self-monitors, who prefer to plan before writing, and low self-monitors, who generate content at the moment of typing and therefore work "more interactively with the text" (Lindgren and Sullivan 2006a, 171). According to Lindgren and Sullivan, "the interactive writers would probably revise extensively pre-contextually, as this is the location in writing at which content and ideas are most easily elaborated upon" (171). Among the authors studied here, we can also observe a distinction between authors who revise almost exclusively contextually and those who also revise pre-contextually.
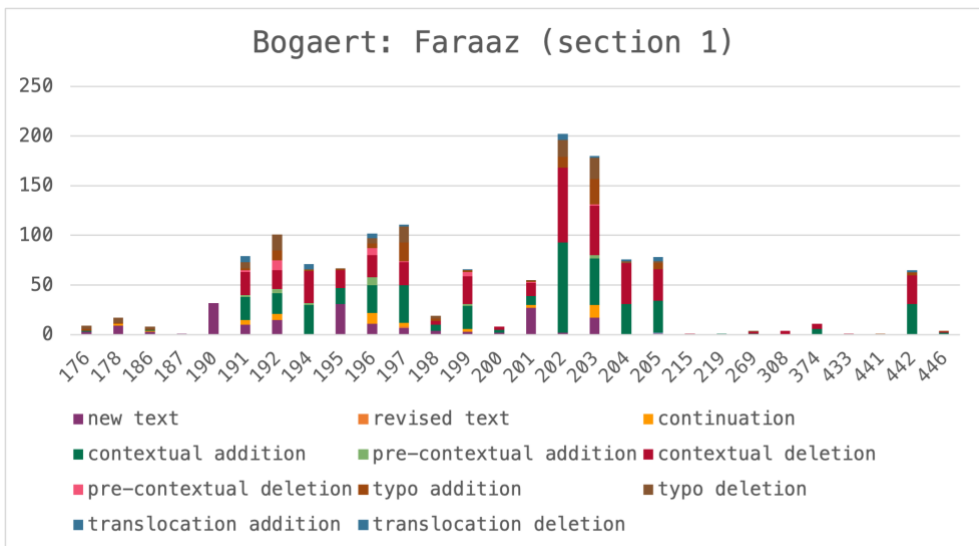
Figure 5.8: Number of the different types of modifications to the text per session for Bogaert (writing the first section of Faraaz).

Figure 5.9: Number of the different types of modifications to the text per session for Posthuma.

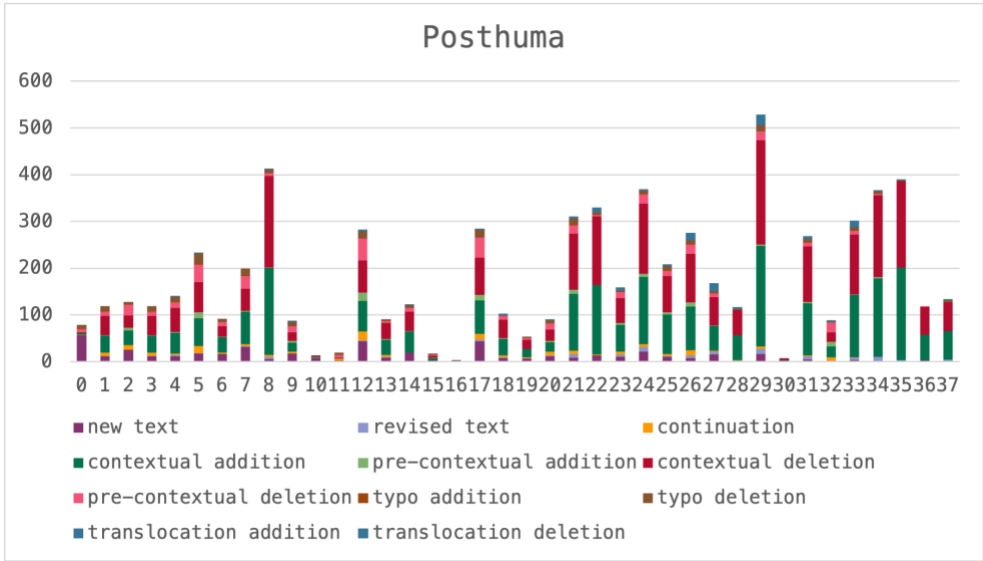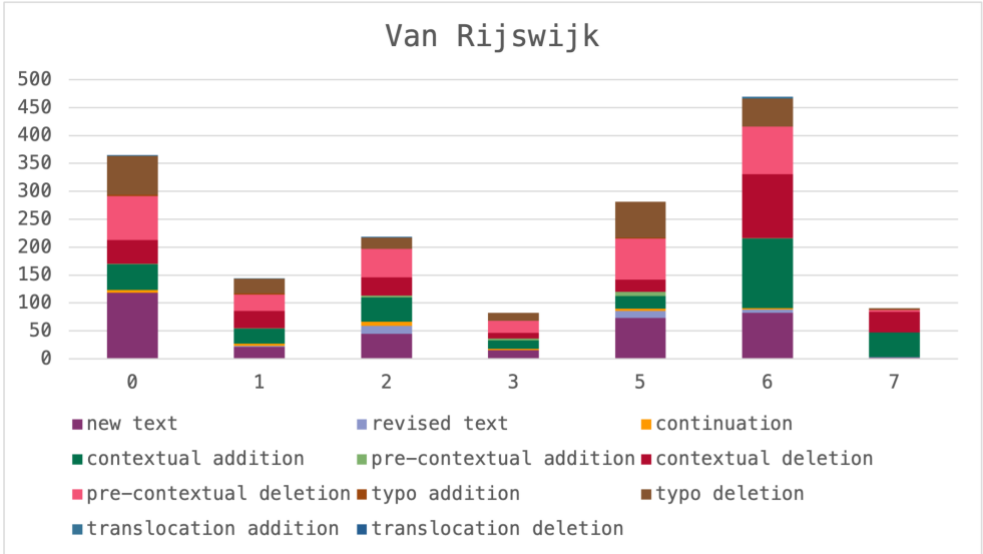*Figure 5.10: Number of the different types of modifications to the text per session for Van Rijswijk.*
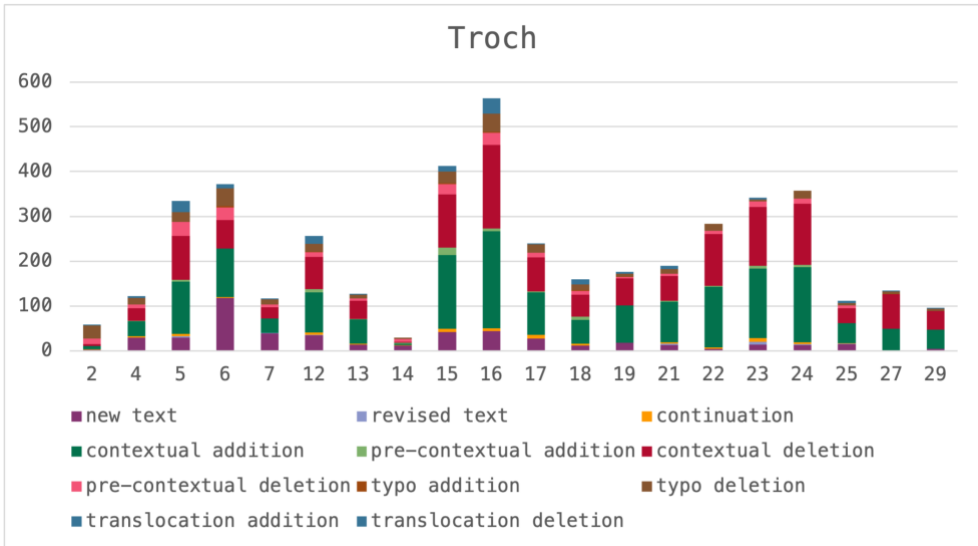
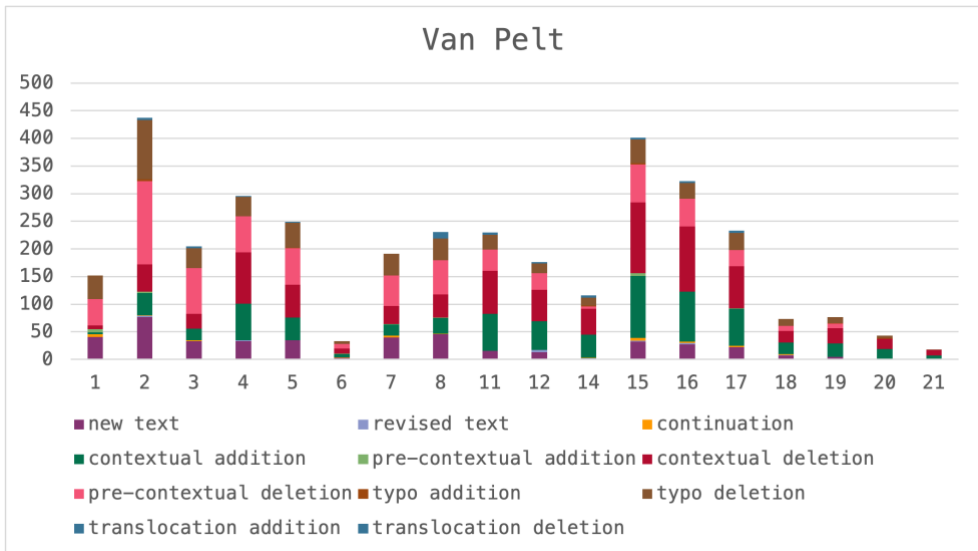*Figure 5.11: Number of the different types of modifications to the text per session for Troch.*

Figure 5.12: Number of the different types of modifications to the text per session for Posthuma.

Compared to the other authors, Van Rijswijk and Van Pelt made more revisions at the pre-contextual level. i.e. before the sentence was finished. These pre-contextual revisions were mostly pre-contextual deletions; as Van Rijswijk and Van Pelt wrote, they deleted the last typed parts of the sentence they were producing at that moment. This again contrasts with the revision habits of Posthuma and Troch. In the cases where Troch and Posthuma revised at the pre-contextual level, they also made pre-contextual additions, i.e. they left the leading edge of the sentence they were producing in order to make an addition. In general, Bogaert did not often make pre-contextual revisions while writing the sections narrated by Faraaz (except for Sessions 196 and 196), his revisions were almost always contextual. These differences were the result of different typing strategies (see Chapter 5.2).

In all the sessions – except the last session, which was mainly devoted to final revision (prepublication phase) – Van Rijswijk mostly revised the text she had just produced in that session, ranging between 60% and 100% of all the revisions. For the other authors, such numbers of revisions within the text produced in the same session tended to occur only in the first few writing sessions. This is logical, of course, since this is usually the only text that exists during these early sessions. Thus, at this level, revision habits are most often manifested in the varying use of pre-contextual and contextual revisions. Finally, at the level of the writing process itself, which is made visible in the visualisations of the reconstructions, it is worth discussing one main, commonly shared, revision habit that seems to be typical for digital writing. The authors began *most* of their writing sessions by revising their text produced so far, usually from the top of the document. At the beginning of the writing session, they probably (partly) re-read what they had written to 'get a feel' for the story again, and then came across formulations that they felt needed to be revised. This endless revision is made possible by the digital environment because previous revisions leave no trace; there is plenty of opportunity to revise again and again. With analogue writing, the habit of re-reading and revising also occurs, but the number of revisions is limited to the space available on the paper. Too many revisions in the same space would make the text illegible.

5.2 NANOGENESIS: A NEW ASPECT OF TEXTUAL GENESIS

With these types of revisions within the text produced so far as a starting point, the actual genesis of the texts can be analysed in relation to the movement through the text. Mahlow et al. (2022) point out that the production process of written text is highly non-linear: "during production, writers are free to modify the text at any place and at any point in time, without leaving any traces in the final product" (2022, 'Introduction'). When examining keystroke logging data, the geneticist can examine the writing process at an unprecedented level of granularity, including this non-linear behaviour and the sequentiality of text production and revision. The fine-grained data of keystroke logging

therefore enables a new type of what can be called "nanogenetic" research. I have coined this term to complement the already existing terms macrogenesis and microgenesis, because the keystroke logging data allows us to really focus on the sequentiality of the text production and revision. The prefix 'nano' comes from the Greek νάννος, meaning 'dwarf' and is currently used to denote something very small, usually quantities. It is also known through nanotechnology, which is defined as "the understanding and control of matter at the nanoscale, at dimensions between approximately 1 and 100 nanometers" (NNI n.d.). Of course, the prefix nano- is only used metaphorically in the term nanogenesis, especially if you imagine that a sheet of paper is about 100,000 nanometres thick (NNI n.d.). Nevertheless, it focuses on small – perhaps even the smallest – visible aspects of the writing process that we have not been able to perceive before. Central to the nanogenesis is the author's movement through the text, which can be deduced from the order in which the text was typed; the way in which letters and words were deleted; sentences that were abandoned mid-sentence but finished at a later stage; and the exact sequence of the production of new text and revisions. In addition, it is important to realise that a work's microgenesis and nanogenesis can be studied separately: one could, for example, study the variation of a particular paragraph without taking into account the fine-grained details of the exact sequence of the keystrokes (microgenesis), or, conversely, focus solely on the author's movement through the text (nanogenesis).

5.2.1 Creating coherence

The fine granularity of keystroke logging and how this can lead to an improved understanding of the writing process, can be illustrated through an example from the genesis of Troch's story 'Mondini'.[129] In the final version of the story, there is a clear connection between the second and the tenth paragraph. In the second paragraph, the narrator muses on how much the interaction between people has changed: people used to wink at each other, flirt, buy each other a drink, share a cigarette, and end up in bed with each other:

> Vroeger, veel, veel vroeger **knipoogde** je naar elkaar, flirtte je, trakteerde je op een glas, **deelde je een sigaret** en belandde je met elkaar in bed.[130]

---

[129] A previous version of this section has appeared in Bekius, L., and D. Van Hulle, "The literary draft in the 21st century: from paper notebooks to keystroke logging". *Toward a Comparative History of the Literary Draft in Europe. Comparative History of Literatures in European Languages Series.* (forthcoming). This chapter is co-authored, but the analysis and the writing for this section was done by Bekius, supported by comments from Van Hulle.

[130] Translation: "In the old days, much, much earlier, you **winked** at each other, flirted, had a drink, **shared a cigarette** and ended up in bed with each other".

In the tenth paragraph, the first-person narrator recalls a memory of his partner Evelina. One of the things he remembers is that she came home after buying cigarettes, and that, after a series of minor events, she lit a cigarette and offered one to him in exactly the same way as after the first wink she ever gave him:

> Evelina die opgelucht ademhaalt, toekijkt hoe ik de scherven bij elkaar veeg, een sigaret opsteekt, inhaleert, de rook uitblaast, **mij op identiek dezelfde wijze als na die eerste knipoog een sigaret aanbiedt**, mij vuur geeft, mij een tel afwezig aanstaart en zegt: 'Moet je nu eens wat weten?'[131]

When first reading the second paragraph, the reader could interpret it as a description of how people used to interact *in general*, like winking at each other and sharing a cigarette, before climate change affected the world and made people fear each other. Only when the reader gets to the tenth paragraph, does the second paragraph acquire additional meaning. After reading this *particular* memory of Evelina offering the first-person narrator a cigarette in exactly the same way as after that first wink, the reader understands that the sentence in the second paragraph was not just a description of old, long-gone – yet very recognisable – behaviour, but a description of how the first-person narrator met his partner. She was the one who winked at him, they flirted with each other and shared a cigarette and later the bed. This raises the question when the connection between these two paragraphs was made. Was this connection already present from the moment those paragraphs were both incorporated in the document or was it only established later in the writing process?

The reconstructions of Troch's writing process show that he only made this connection explicit in the 24th writing session. On 27 August 2020, Troch adds to the narrator's description of Evelina in the tenth paragraph that she offered him a cigarette in the same way as she did after that first wink.

> Evelina die opgelucht ademhaalt, toekijkt hoe ik de scherven bij elkaar veeg, een sigaret opsteekt, inhaleert, de rook uitblaast, mij ~~net~~[142]op identiek dezelfde wijze[143] als na [141]die eerste ~~sig~~[145][144]knipoog [146]een sigaret aa~~b~~[139][138]nbiedt, [147]~~die ik~~ [148][140]mij vuur geeft, mij [149]een tel afwezig aanstaart en zegt: 'Moet je nu eens wat weten?' (Session 24)

Only now does it become explicit that the cigarette mentioned in the second paragraph is not just any cigarette. The particular memory would be of little interest without the

---

[131] Translation: "Evelina, breathing a sigh of relief, watched as I wiped the shards together, lit a cigarette, inhaled, exhaled the smoke, **offered me a cigarette in exactly the same way as after that first wink**, gave me a fire, stared at me absently for a moment and said: 'Do you want to know something?'".

recognisability of its general character, while the general statement would have been rather bland without the *couleur locale* of the particular memory.

The nanogenetic analysis now makes it possible to see when the author made that connection, and it appears that this actually happened much earlier than Session 24, namely in Session 15. On 19 August 2020, Troch started this session by revising the second paragraph, making the description of the interaction between people more specific. Before the revisions, the paragraph stated in rather colourless, general terms that people used to greet each other, have a chat, and maybe even invite each other for a drink. After the revisions, the sentence specifies that people greeted each other, flirted with each other, offered each other a glass, and ended up in bed with each other. But there is no mention of the cigarette – yet.

> Vroeger, veel, veel vroeger groette je elkaar, flirte je, [4]maakte je een praatje, nodigde je elkaar misschien zelfs uit[6]trakteerde je de ander [11][12]op [7]voor [8]een glas, belandde je bij[9] met[10] elkaar in bed[5] (Session 15, after n4-12).

After these revisions in the second paragraph, Troch continues making revisions: he adds new sentences to the story, and makes a decision as to the name of the narrator's partner, she will be called Evelina. He adds her name to a sentence in the tenth paragraph about imagining that she has just gone out to buy some cigarettes.

> Evelina[359] Ze (naam?)[360] zou maar eens echt gewoon om sigaretten moeten zijn en straks uitbundig de deur openzwaaien. (Session 15, after n359-360)

Re-reading this sentence was therefore most likely the incentive for the subsequent revisions. 21 seconds after changing the name in the tenth paragraph, Troch moves his cursor back to the second paragraph to add: "shared a cigarette" and "winked at each other".

> Vroeger, veel, veel vroeger groette[366]knkip[368][367]ipoogde[369] je naar [370]elkaar, flirte je, trakteerde je op een glas, de[1][362][361]elde je een sigaret, [364][363]en [365]belandde je met elkaar in bed (Session 15, after n361-370)

By means of the nanogenesis, we can infer that the addition of sharing a cigarette was triggered by the rereading of the sentence about Evelina, which also mentions cigarettes. The moving of the cursor between these revisions allows us to see that Troch made a connection between the sentences in paragraph two and ten in Session 15, which means that he envisioned this relation way *before* this connection is made explicit in the text in Session 24. In this way, the nanogenetic analysis provides information about how this cohesion in the text is established during the writing process and, indirectly, it may also shed light on the cognitive processes involved in (creative) writing as studied in writing

OVERVIEW OF TROCH'S WRITING PROCESS

Monday, 10 August 2020	Start writing process; 3 sessions
Wednesday, 12 August 2020	3 sessions
Wednesday, 19 August 2020	3 sessions; **in the last session the connection is made implicitly**
Friday, 21 August 2020	4 sessions
Thursday, 27 August 2020	4 sessions; **in the last session the connection is made explicit**
Saturday, 29 August 2020	1 session
Wednesday, 16 September 2020	1 session
Thursday, 17 September 2020	End of writing process; 1 session

Table 5.3: Overview of Troch's writing process with a focus on the connection between the second and tenth paragraph.

studies according to models such as the ones proposed by Linda Flower and John R. Hayes.

The written traces produced by the author include, with digital writing, the text produced so far (Hayes 1996). In contrast to analogue versions of a text, the reconstructions of the writing process based on the keystrokes allows us to analyse the text at any given point in the writing process. As such, we may, for example, observe the state of the text when new ideas emerge, or when ideas about the text, or specific formulations, changes direction. This allows us to hypothesise about how certain formulations in the text produced so far prompted revisions or insertions of new aspects in the text.

In Hayes' 2012 model of the writing process, the various cognitive functions at work in a writing process are categorised as "proposer", "evaluator", "translator" and "transcriber" (Lindgren et al. 2019, 371). The "proposer" is the part of the mental writing process that suggests an idea; if the "evaluator" agrees to it, the "translator" translates it into a linguistic form, after which the "transcriber" is activated to create the actual written text (371). The (external) text produced so far is, in Hayes model (2012), therefore controlled by the internal transcriber. But as Lindgren et al. (2019) observe, the interaction between the text produced so far and the internal writing process remains ambiguous:

> According to Chenoweth and Hayes (2001) there are direct links (arrows) from the text produced so far to both the 'proposer' and the 'translator' and an arrow back from the transcriber to the text. In Hayes' (2012) and Leijten et al.'s (2014) models the directions between these processes and the text produced so far are less clear as there is only an arrow leading from the translator to the text produced so far with no direct indications of feedback. (Lindgren et al. 2019, 348)

Lindgren et al. (2019), then, state that there are dynamic ways in which the text produced so far may feed into the internal writing processes (348). After re-reading the text, the author might be dissatisfied with certain parts of it; in that case "the proposer would suggest a new idea that is evaluated and translated into linguistic form before being transcribed" (Lindgren et al. 2019, 348). It is this "feedback loop from the text produced so far into the internal writing process and back again" (348) that we see at work in the example above. Troch re-read the sentence about Evelina going for cigarettes in the tenth paragraph; this led to a suggestion of the 'proposer' that is then being evaluated, translated into linguistic form, and eventually transcribed into the second paragraph. In other words, without the possibility to 'enter the author's mind', the fine granularity of nanogenetic data does enable us to come closer to an understanding of the underlying cognitive processes.

Still, what the 'proposer' suggested remains internal and the approximation of these internal cognitive process may be asymptotic: We can come closer, but the curve may never meet the line at any finite distance. Numerous questions remain unanswered. Were the additions about the flirting meant to make this a personal memory, and did the 'proposer' suggest adding the cigarette to the second paragraph after Troch was reminded that Evelina smoked while re-reading the tenth paragraph? Or was the second paragraph still intended as a general description and was the addition of the cigarette primarily proposed to indicate that smoking was also something typical of the past? Or did the re-reading itself prompt the idea of colouring the general description with a personal memory? Or was it something completely different? This indicates that even when we have access to nanoscopic data of the writing process, internal parts of the writing process can still not be fully understood. However, this does not diminish the fact that we can get closer to the writer's decision-making process than ever before.

5.2.2 Textual awareness

The example of Troch's writing process shows how the nanogenesis can provide information about how an author creates coherence in their text during the writing process. Another aspect that the nanogenesis can shed light on is the 'textual awareness' that the author has while producing the text. In *Textual Awareness* (2004), Dirk Van Hulle uses the term 'textual awareness' for "a constant awareness of the textual aspects of the writing" in literary modernism (10). Here, I use a more specific implementation of the term, namely an awareness of the formulations used throughout the text produced so far as the writing process progresses.

This 'textual awareness' can be illustrated by Van Rijswijk's writing process. Halfway through Session 6, Van Rijswijk begins to write a sentence in which the I-narrator points out that (the ghost of) her grandmother is nowhere to be seen: "Mijn oma is in geen velden of" (n352). After producing this beginning of the sentence, she

pauses for 50 seconds and scrolls through the Word document without actually relocating the cursor. She then deletes "in geen velden of" (n353), and writes – four seconds later – "nog ste" (n354), but deletes this immediately. She scrolls through the document again, but this time she moves the cursor and substitutes "bekennen" with "zien" in the sentence "Mijn oma en haar aanhang zijn nergens meer te bekennen" (n356-357).[132] Van Rijswijk then returns to the unfinished sentence to complete it with "in geen velden of wegen te bekennen" (n358-360).[133] The two sentences – the one already produced and the one to-be-written – are clearly related, both in content and in time.

With this reconstruction of the production process of the sentence "Mijn oma is in geen velden of wegen te bekennen", we can hypothesise that Van Rijswijk – when she started writing the sentence – remembered that she had already written something similar. This indicates a high level of textual awareness; during the production of this sentence, she was aware of the gist of the text produced so far. To check how she had previously formulated that the ghost of the grandmother was nowhere to be seen, we can imagine that during the first pause Van Rijswijk scrolled to the sentence in question. Seeing that she had formulated it as "zijn nergens meer te bekennen" probably made her delete "in geen velden" in the sentence she had just started to write. She presumably did this to avoid the repetition of the word "bekennen", as her initial plan was probably to write the proverb: "in geen velden of wegen te bekennen". However, as she searched for another way of expressing the absence of the ghost of the grandmother, Van Rijswijk may have realised that she could also replace the word "bekennen" in the sentence earlier in the story (where she did not use the full proverb "in geen velden of wegen te bekennen") and use it instead in the sentence she was planning to write. As she did. The nanogenesis in this case shows the interaction with the text produced so far in order to arrive at the right formulations and to avoid repetitions.

5.2.3 Typing habits

Another aspect of digital writing that can be revealed by a nanogenetic analysis is the author's typing habits. As mentioned in Chapter 4.2, keystroke logging makes us aware of the peculiarities of digital writing. Just as each author has a unique handwriting, each author also seems to have a personal way of performing writing actions. In the same section, I already pointed out that Bogaert likes to recycle characters and words when he revises a sentence, as can be seen in the following two examples as well. In the first example, Bogaert made a contextual revision, replacing "zou je kunnen zeggen"[134] with

---

[132] Translation: "My grandmother and her entourage are nowhere to be seen".
[133] Translation: "nowhere to be seen".
[134] Translation: "you might say".

"zegt hij soms",[135] while keeping the letter "z". In the second example, Bogaert changed "ongemerkt" into "opvallend", by simply replacing "gemerkt" with "opvallend".

1) [49]Hij leeft van zijn woorden, z̲e̲g̲t̲ ̲h̲i̲j̲ ̲s̲o̲m̲s̲[102]~~ou je kunnen zeggen~~[103] […] (Opstap, Session 30)

2) Hij moet beslissingen nemen, on̲o̲p̲v̲a̲l̲l̲e̲n̲d̲[23]~~gemerkt~~[24] maar doordacht en kordaat. (Faraaz 1, Session 199)

What the examples also show is that Bogaert mainly uses the delete-key to perform the deletion; the added text is given before the deleted text, because the deletion was actually preceded by the insertion. His preference for the delete key distinguishes him from other writers, who use the delete key only occasionally. In addition to using the backspace key, Troch often deletes text fragments by holding down the shift-key and using the arrow keys to select the text, which he then deletes using the backspace key. Posthuma, Van Rijswijk and Van Pelt mostly use the backspace key, or they select the fragment to be deleted – usually with the cursor – and then delete it with the backspace key.

Bogaert's recycling of characters and words may be related to his typing skills; he may not be able to type very fast and therefore wants to minimise the number of keystrokes. This is also something observed by Mathias Fürer, who analysed the writing process of "an experienced writer whose typing speed is considerably slow because he mostly uses only his two index fingers to type", which led to this writer being "able to develop a very concise and detailed text in his mind while writing out a story" (2017, 40). This may also explain why Bogaert made few pre-contextual revisions. Seen in this light, it seems that Van Rijswijk – on the contrary – is not afraid of re-typing. In all of her writing sessions, it can be seen that she often deleted substantial parts of sentences in order to change only one or a few words. In the first example, Van Rijswijk deleted "in de waterkou", because she wanted to replace "in de" with "bij". She could have used the cursor to delete only "in de", but it might have been more convenient for her to delete "waterkou" as well. The same applies to example 2, where she only replaced "zegt" with "vindt", but deleted "zegt mijn moeder" completely to achieve this. In example 3, she only needed to reposition "je", for which she could have used the cursor to delete the word and re-type it in the right place. But again, her typing skills might have led her to delete "weet je nog w" (n511). Nevertheless, it is easy to imagine that Bogaert would have chosen the first option.

---

[135] Translation: "he sometimes says".

1) […] het jaar, ~~in de waterkou~~[245][246]bij waterkou ons vertier […]" (Session 0)

2) […] vooruit, ~~zegt mijn moeder,~~ [483][484]vindt mijn moeder, […]" (Session 0)

3) […] want ~~weet je nog w~~[511] [512]je weet nog […]" (Session 0)

4) Ik weet niet waarom ik ~~niet meer die rilling~~[25]nooit meer een rilling langs mijn rug ~~voel~~[27][26] voel op begraafplaatsen." (Session 1)

5) […] en ik heb ze nooit gekend~~en~~[32] [33]~~daar raken we allebei licht gefrustreerd van~~[34]we raken allebei een beetje gefrustreerd[35]. (Session 1)

Examples 4 and 5 concern contextual revisions. Here she could have recycled a lot of words that would also be in the revised part of the sentence, but she still chose to delete it completely, which would perhaps be just as easy for Van Rijswijk.

Van Pelt is also an avid user of the backspace key, as she seems to be thinking with her fingers, making many pre-contextual deletions. Compared to Van Rijswijk's pre-contextual deletions, Van Pelt's concern relatively short bursts of text. As the examples below show, the deletions concern only small parts of sentences, sometimes barely a whole word.

1) [295]~~de~~[296] [297]~~de~~[298] [299] ~~D~~[300] [301]~~De~~ ~~temperatuur is er aan~~[302][303]~~sh~~[304][305]~~eb~~[306] [307]~~beschutting~~[308][309]~~Een beeje~~ [310] [311]~~t~~[312] [313]~~De b~~[314][315]Het doet me goed om ~~uite~~[316][317]t de zon te zijn […] (Session 1)

2) [329]~~Het~~[330][331] ~~Z~~[332] [333]~~Ik~~ [334] [335]~~Vor~~[336][337]~~or ons staat d~~[338] [339]~~een hee~~[340] [341]~~dame~~[342] [343]~~vrouw op, ze zwaait~~ ~~wil~~[344] [345]~~wil~~[346] [347]~~Enke~~[348] [349]'Twee rijen voor on~~as st~~[350][351] ~~st~~[352] [353]s staat een vrouw bruusk op. (Session 15)

The way in which Troch gradually built up some of his sentences reveals a typing strategy that he used repeatedly in his writing process. Daniel Chandler distinguishes various writing strategies, including the "bricklaying strategy", in which each sentence or paragraph is polished before moving on to the next (1995, 86). Troch seems to use a more extreme form of this strategy, in which he builds a sentences word by word. First, he writes down his initial impulse of inspiration, as if he wants to put the loose idea into concrete before it slips away. He then builds the sentence word by word, brick by brick, leaving gaps for words he has yet to decide on. Once he has decided, he fills the gaps with the appropriate words, like the grout that is used to fill the joints. This strategy is already present in the writing actions Troch took to write the sentence "Mijn geschuifel langs de heiligenbeelden heeft veel weg van het getrippel van een muis"[136] in

---

[136] Translation: "My shuffling past the statues of saints is very much like the pitter-patter of a mouse".

the 18th writing session, discussed in Chapter 4.3, but it occurs more often. Take, for example, the writing process of the sentence "Ik die altijd zo'n fantastische appetijt had, heb nu de grootste moeite om een volledig speculaaskoekje door mijn keel te krijgen"[137] in Session 16:

> Ik,[193] die altijd zo'n [180] fantastische a[186][176]~~A~~[187]ppetijd[177][178]t [188]had, ~~niet nu deg~~[189][190] heb nu de groot~~ts~~[191][192]ste moeite om een volledig specu~~y~~[196][195]laaskoekje [197][194]door mijn keel te krijgen. (Session 16)

Troch first writes "Appetijd" (n176), but immediately spots the spelling error and corrects it to "Appetijt" (n177-178). Again, by typing the keyword 'appetijt', he concretised his first impulse of inspiration. He then briefly googles the word, perhaps to check that he has spelt it correctly. He returns to the Word document and writes "Ik die altijd zo'n" (n180) in front of "Appetijt". He probably already had an idea of how to continue this sentence, because he looks up synonyms of the words "onevenaarbaar" (n181), "evenaarbaar" (n182), "evenaren" (n183), "vergelijkbaar" (n184), and "onberispelijk" (n185). Although it was not in the list of synonyms for the word "onberispelijk", he adds "fantastische a" (n186) before "Appetijt" and deletes the capital "A" (n187). He then continues finishing the sentence with "had, niet nu deg" (n188), immediately deletes the part after the comma (n189), and continues with "heb nu de grootste moeite" (n190-192). Then he makes another pre-contextual addition – a comma between "Ik" and "die" (n193) – after which he finishes the sentence: "om door mijn keel te krijgen" (n194). Just as with the sentence that began with "Trippelen", Troch still has to add something to make the sentence syntactically correct, in this case: what is it that the narrator finds difficult to swallow? This is what Troch adds in the following writing action: one whole *speculaas* biscuit ("een volledig speculaaskoekje", n195-197). This is yet another example of how Troch tends to build some of his sentences like a bricklayer, working out his ideas brick by brick.

The same principle can be found in Troch's revision of sentences, during which he transforms sentences step by step. During Session 18 (21 August 2020), Troch wrote an already syntactically correct sentence: "Mijn voeten maken geluid"[138] (n44-46), but before he types the full stop, he begins to revise it:

> Als ik over de straatstenen schuifel, maken m[57][44] ~~M~~[58]ijn ~~voeten~~[47] voeten in de [54] slippers ~~die over de~~ [55][48] ~~straats~~~~chenen~~[51][50]~~tenen,~~ [53][52][56]~~maken~~[59]een soppend [49]gele[45][46]uid[60]. (Session 18)

---

[137] Translation: "I, who used to have such a great appetite, now have difficulty getting an entire speculaas biscuit down my throat".

[138] Translation: "My feet are making noise".

WRITING ACTION	TEXT AFTER THE WRITING ACTION
adds *Mijn voeten maken geluid* (n44-46)	**Mijn voeten maken geluid**
Deletes *voeten* (n47) and replaces it with *slippers die over de* (n48)	Mijn **slippers die over de** maken geluid
adds *een soppend* (n49)	Mijn slippers die over de maken **een soppend** geluid
adds *de straatstenen* (n50-53)	Mijn slippers die over **de straatstenen** maken een soppend geluid
adds *voeten in de* (n54)	Mijn **voeten in de** slippers die over de straatstenen maken een soppend geluid
deletes *die over de straatstenen* (n55-56)	Mijn voeten in de slippers maken een soppend geluid
adds *Als ik over de straatstenen schuifel, maken* m (n57) and deletes *M* (n58) and *maken* (n59)	**Als ik over de straatstenen schuifel, maken** m**ijn voeten in de slippers een soppend geluid
adds the full stop (n60)	Als ik over de straatstenen schuifel, maken mijn voeten in de slippers een soppend geluid**.**

Table 5.4: Troch's writing actions in the composition of the sentence "Als ik over de straatstenen schuifel, maken mijn voeten in de slippers een soppend geluid."

Table 5.4 gives an overview of all the writing actions involved in the revision of the sentence "Mijn voeten maken geluid" into "Als ik over de straatstenen schuifel, maken mijn voeten in de slippers een soppend geluid" (excluding the immediate correction of typing errors).[139] Here, again, Troch leaves room for the insertion of words for which he has yet to make a decision. On a small scale, the way Troch revised the sentence adheres to what Alamargot and Lebrave have termed "creativeness after writing" (2010, 17; see Chapter 5.4).

Posthuma is a very concise writer, as she carefully revises and rewrites certain passages that she has already written. Her story includes a description by the male friend of his childhood; he talks about his mother and especially about his stepfather, who was always watching porn. Posthuma wrote this passage in Session 12. In Session 18, she revised some words in this passage, but in Session 21, she returned to the fragment to rewrite it:

> Hij vertelt over zijn kettingrokende moeder en zijn stiefvader die voortdurend porno keek. Als hij thuiskwam uit school rinkelde hij hard met zijn sleutels en maakte hij veel lawaai in de gang om zijn stiefvader de kans te geven snel de computer uit te zetten en met zijn onderbroek om zijn enkels naar zijn

---

[139] Translation: "As I shuffle across the paving stones, my feet in the slippers make a sopping sound".

slaapkamer te rennen. O shit, hoorde hij dan, en dan gestommel. (Fragment at the start of Session 21)[140]

Hij vertelt over zijn kettingrokende moeder die drie slechtbetaalde baantjes had en daardoor nooit thuis was en hoe hij als hij uit school kwam extra hard met zijn sleutelbos rammelde zodat zijn stiefvader Larry, die altijd thuis was, gauw alle pornofilmpjes kon weg klikken en naar zijn slaapkamer rennen om daar wat kleren aan te trekken, want van die ene keer dat zijn broek nog om zijn enkels zat toen hij wegvluchtte, waardoor hij op de glazen salontafel smakte en een diepe snijwond opliep, had hij zijn lesje geleerd. O shit, hoorde ik altijd, en dan gestommel, zegt onze vriend. (Fragment at the end of Session 21)[141]

The revised paragraph includes some new details, such as the mother's three jobs and the cut the stepfather suffered when he fell on the glass coffee table. The revision of the paragraph, however, mainly concerned the formulation. Nevertheless, Posthuma spent almost half an hour – two times fifteen minutes, with a break in between – fiddling with and revising the paragraph, trying out different formulations. Her concise revision of the paragraph is revealed by the reconstruction of the writing process:

Hij vertelt over zijn kettingrokende moeder die drie slechtbetaalde[128] baantjes had [127]en daardoor nooit thuis was en [131] ~~over [129]~~zijn stiefvader ~~Larry [142]~~~~die [235] altijd thuis was[236]~~ ~~en hoe hij[206][210]~~ hoe hij [237]als hij ~~terug[133][132]~~uit school kwam[134] ~~en[226],[227]~~ ~~hoe hij dan[238][211]~~ extra hard met zijn sleutelbos rammelde ~~en [208][207]~~ ~~zodat[209].[212]om[213][214]~~ zodat[215] [130] ~~Hij [143]~~ zijn stiefvader [233] Larry[144], die altijd thuis was,[234] ~~'o shit' kon roepen en [242][277]~~ ~~wist dat hij er was en[239]~~ gauw [228] ~~op tijd [229] de computer uit kon zetten [217][216]~~ ~~de[279]~~ alle[280] porno~~webs[219][218]~~~~site[222]~~filmpjes[223] ~~die hij [221]~~ ~~zat te bekijken [224][230]~~ ~~kon[243]~~ kon[278] wegklikken en naar zijn slaapkamer ~~rennen[231]~~ ~~vluchten[232][248]~~ rennen[249], [240] om daar [244]wat kleren aan te trekken[241], want van die ene keer dat ~~hij [246][245]~~ zijn broek nog om zijn enkels zat [247] toen hij weg[251]vluchtte [250] ~~omdat[255][225]~~ ~~vluchten[256][220]~~ ~~zat altijd[136][137]~~ ~~de hele dag[138]~~~~in de woonkamer~~ ~~naar porno te kijken[145][146]~~~~, waar de enige computer stond ,e[147]~~en keek ~~daar [140]~~ ~~naar[141][139] [257][135]~~ ~~porno[148][149]~~ allerlei soorten porno, maar het liefste

---

[140] Translation: "He tells about his chain-smoking mother and his stepfather who constantly watched porn. When he came home from school, he would ring his keys loudly and make a lot of noise in the hallway to give his stepfather the chance to quickly turn off the computer and run to his bedroom with his pants around his ankles. Oh shit, he then heard, and then stumbling".

[141] Translation: "He tells about his chain-smoking mother who had three low-paying jobs and was therefore never at home and how when he came home from school he would rattle his keychain extra hard so that his stepfather Larry, who was always at home, could quickly click away all the porn movies and run to his bedroom to put on some clothes, because from the one time his trousers were still around his ankles when he fled, causing him to smash on the glass coffee table and suffer a deep cut, he had learned his lesson. Oh shit, I used to hear, and then stumbling, says our friend".

~~vrouw met vrouw.~~ [150] ~~Onze vriend rammelde altijd hard met zijn sleutelbos als hij thuiskwam, en~~ ~~hiujh~~ [152][195][151]~~j.~~[196]maakte expres [175]~~veel lawaai in de gang,~~ ~~zodat Larr~~[154][153] ~~zodat~~[159] ~~om~~[160][171] ~~en dan hoorde hij Larry 'o shit' zeggen en~~ ~~weg~~[173][172] ~~naar zo~~[179]ijn slaapkamer vluchten.[176][174] ~~Larry de kans~~ ~~kreeg~~ [161] ~~te geven~~ [162] de computer uit te zetten [156] ~~om~~[157] ~~en~~[158] ~~naar zijn slaapkamer~~ [177][155] [163] ~~te vluchten~~[178]~~, want meestal had hij geen kleren aan~~, dan kon hij beter rennen[188][197]. [258] ~~Van die ene keer dat~~ hij ~~een~~[189] zijn[190] onder[191]broek ~~droeg die~~[192] nog om zijn ~~enkels zat toen hij vluchtte~~[259], waardoor hij op de...[199] glazen [201]salontafel[200] ~~Viel~~[183] smakte[184] en zijn [203] ~~enkel~~[184][180] scheenbeen brak[182][202] een ~~grote~~[252] diepe[253] snijwond opliep[204][185]. [86] had hij zijn lesje [254]geleerd.[187] [164]~~Een keer droeg hij een onderbroek die nog om zijn e~~ ~~ngel=~~[165][166]nkels zat, waardoor hij ~~viel en~~ [167][168]~~over~~ [169][170]struikelde en voortdurend porno keek.[193] ~~Als hij thuiskwam uit school rinkelde hij hard met zijn sleutels en maakte hij veel lawaai in de gang om zijn stiefvader de kans te geven snel de computer uit te zetten en met zijn onderbroek om zijn enkels naar zijn slaapkamer te rennen. O shit, hoorde hij dan, en dan gestommel.~~[194] O shit, hoorde ik altijd,en dan gestommel,[261] zegt onze vriend[260][263]. (Session 21)

In Sessions 22 and 25, the passage is again heavily revised and in Sessions 26, 28, 29, 31, 34 and 35 it is again slightly reworded. The content remains more or less the same, but again Posthuma mostly fiddles with the wording. So Posthuma has a habit of rewriting paragraphs intensively, trying out different formulations, weighing up words carefully. Sometimes, after such a rewrite, she might decide on a formulation that is very similar to the one she started with. This incisive time spent with paragraphs is typical of Posthuma, but 'over-editing' is something that may also be typical of digital writing. Chandler notes that when using a word processor, some writers find it difficult to know when to stop revising, or that "they do too much 'tinkering' at a 'local' word or sentence level" (1995, 145). This is due to the fact that so much re-writing is possible because of the 'suspended inscription': "With a typewriter or a pen the amount of revision one does is limited at least by decreasing legibility or by one's willingness to re-write pages to keep them looking tidy" (Chandler 1995, 145).

In summary, what all of the above makes clear is that Word documents may not appear to have the same personality as handwritten manuscripts, but when we make the typing habits visible, the singularity of the word-processed document manifests itself. The way the author types during revision also reveals a writing personality.

5.3 DISCOVERY IN WRITING

The statement that we cannot know when an idea *came to mind* is related to the notion of discovery in writing. Discovery is often used as a metaphor to describe the writer's creative process (Flower and Hayes 1980; Galbraith 2009). And indeed, writers' personal accounts of their writing process are full of such descriptions. The Moroccan-

Dutch poet Mustafa Stitou, for example, states that he is often surprised while writing (of poetry in this case):

> Het schrijven van poëzie is een mengeling van wet en anarchie, van beheersing en dat wat je overkomt. Als je niet wordt verrast tijdens het schrijven, levert dat geen goede poëzie op. Soms word ik ook veel later nog door mijn eigen tekst verrast. Tijdens het schrijven voel ik dan dat er een lading in de tekst zit, ik begrijp dat een woord op een bepaalde plek moet staan, zonder dat ik precies weet waarom. Dat je verrast kan worden door wat je schrijft is in de eerste plaats de verdienste van de taal, van de woorden zelf, want woorden zijn ontzettend veelkoppig, ze kunnen vele kanten op kijken. (Stitou in Hasselt and Sistermans 2021, 26)[142]

Heeks (2012) distinguishes between three related, but slightly different notions of discovery writing. The first is 'pantsing', a term used by creative writers in the United States and to 'pants' means "to fly by the seat of your pants" (Heeks 2012, 28). It relates to free-writing: "pantsing tends to mean to plan ahead very little, and to make up a story while writing" (28). The second, 'writing as generating ideas', is closely related to pantsing, but the "focus is on the kinds of 'ideas' that emerge through writing" (28). The last is 'discovery as a post writing reflection', or 'ah' moment: "This is a process and moment of recognition, of writing something and then seeing that there is something about it which is a discovery, or seeing it in a new light" (29). The example in Troch's writing process above – in which he found a connection between the second and the tenth paragraph – can be interpreted as an example of the latter: the 'ah' moment. Daniel Ferrer describes such processes of discovery in his 'fourth model' for genetic criticism: the gradualist model, which focuses on the process that takes place in time (2022). Here, Ferrer reflects on the moment, after the words have been put to paper, when the author retrospectively discovers the potential of what has been written. The cognitive processes involved in such moments of discovery have been theorised further within the field of cognitive writing process research.

Chandler notes that writers "often say of an idea that 'it came to me', and may add, 'out of the blue' (presumably from the sky)" (1995, 61). This description of discovery, as Chandler points out, "seems to deny the writer's actively 'creative' role, suggesting that the writer has little or nothing to do with the words and ideas" (61). This is precisely

---

[142] Translation: "Writing poetry is a mixture of law and anarchy, of control and that which happens to you. If you are not surprised while writing, it doesn't result in good poetry. Sometimes I am surprised by my own text even much later. During the writing I feel that there is a charge in the text, I understand that a word has to be in a certain place, without knowing exactly why. That you can be surprised by what you write is first and foremost the merit of language, of the words themselves, because words are incredibly multifaceted, they can look in many directions".

what Flower and Hayes find problematic about the metaphor of discovery. They point out that the metaphor implicitly suggests that "somewhere in the mind's recesses or in data outside the mind, there is something waiting to be discovered, and that writing is a way to bring that something out" (Flower and Hayes 1980, 21). They argue that this emphasis on the 'Eureka, now I see it'-experience overshadows the fact that meaning is not something that writers *find*, but rather something that writers *make* (Flower and Hayes 1980, 21).

Flower and Hayes sought to "probe the cognition of discovery, the process itself, by studying the way writers initiate and guide themselves through the act of making meaning" (1980, 22). Their approach was to study writing as "a problem-solving, cognitive process" and they set out to "explore the problem-solving or discovery process that produces new insights and new ideas" (22). In their study, Flower and Hayes observed that good writers continued to develop "their image of the reader, the situation, and their own goals with increasing detail and specificity" as they write (30). Seen in this way, Troch's revisions in the second and tenth paragraph in the example given above, could possibly be linked to his aim of either attempting to build a coherent network of his ideas – to create *meaning* – or to create coherence as a formal feature of the short story. But again, this cannot be verified on the basis of the textual changes alone and we (in the 'Track Changes' project) do not have think aloud protocols to help us interpret this further, as we feel that this would interfere too much with the writing process. Nevertheless, I will argue that the keystroke logging data can give us insight into the process of discovery in writing.

As stated above, the composition process is often seen as a form of problem solving: "The problem of content – what to say – and the problem of rhetoric – how to say it – consumes the writer's attention and other resources of working memory" (Kellogg 2008, 2). However, Galbraith noticed accounts of expert writers – like that of Shirley Hazzard: "I think that one is constantly startled by things that appear before you on the page when you're writing," and that of E. M. Forster: "How do I know what I think until I see what I say?" – which makes no mention of any "thinking behind the text", but rather emphasise the role of writing in constituting thought: "For these writers, 'writing is discovery' means that what one thinks emerges in the text as it is produced, rather than being something which lies behind the text directing its production" (Galbraith 1999, 138). In these cases, new ideas are discovered through the initial spontaneous draft. This is also the point of the dual drafting strategies proposed by Elbow (1973) and Wason (1980), among others. These strategies imply that the writer does not plan in advance but articulates "their thought spontaneous in text without worrying how well expressed or well organised the text is" (Galbraith 1999, 138). In the second stage, the initial draft is transformed into well-formed text through explicit problem-solving operations, but it is the spontaneous production of text that leads to

the discovery of new ideas. Moreover, the problem-solving may not be an appropriate metaphor for modelling literary fiction writing (Heeks 2012). The problem space in literary writing is more open, and literary writers are less bound by pragmatic constraints (Alamargot and Lebrave 2010).

5.3.1 Evolving knowledge

In their model of the writing process, Bereiter and Scardamalia (1987) make a distinction between a knowledge-telling model of writing and a knowledge-transforming model of writing. They illustrate this difference with a description of the writing process by a 12-year-old student and by the author Aldous Huxley:

> I have a whole bunch of ideas and write down until my supply of ideas is exhausted. Then I might try to think of more ideas up to the point when you can't get any more ideas that are worth putting down on paper and then I would end it. (Bereiter and Scardamalia 1987, 9)

> Generally, I write everything many times over. All my thoughts are second thoughts. And I correct each page a great deal, or rewrite it several times as I go along…. Things come to me in driblets, and when the dribbles come I have to work hard to make them into something coherent. (Huxley quoted in Bereiter and Scardamalia 1987, 10)

The quote from the 12-year-old student illustrates the 'knowledge-telling' process, which is characterised by a relatively simple 'think-say' approach to writing, in which content is retrieved from long-term memory and translated directly into writing. The Huxley quote, on the other hand, "is one in which the thoughts come into existence through the composing process itself, beginning as inchoate entities ('driblets') and gradually, by dint of much rethinking and restating, taking the form of fully developed thoughts" (Bereiter and Scardamalia 1987, 10). This is what Bereiter and Scardamalia term the 'knowledge-transforming' process, in which thoughts are actively reworked. In this 'knowledge transforming' model, the knowledge-telling process is "embedded in a problem-solving process involving two different kinds of problem spaces" (11). These are the content space – where "problems of belief and knowledge are worked out" – and the rhetorical space – where "problems of achieving goals of the composition are dealt with" (11).

According to Ronald T. Kellogg (2008), knowledge-telling and knowledge-transforming are only the first two of the three stages involved in the development of written composition skills. In the most advanced stage, which is the knowledge-crafting stage, "the writer is able to hold in mind the author's ideas, the words of the text itself, and the imagined reader's interpretation of the text" (Kellogg 2008, 5). Because the writer's primary concern in the initial knowledge-telling stage is to generate text to say

what they want to say, the writer at this stage may not be able "to comprehend what the text actually says at a given point in the composition" (6). At this stage, the working memory is primarily concerned with the writer's representation of ideas, and not so much with the representation of the text itself. This is only acquired in the second stage of knowledge-transforming, which "involves changing what the author wants to say as a result of generating the text" (6). In the third stage – the knowledge crafting stage – the reader enters the stage: "In knowledge-crafting, the writer shapes what to say and how to say it with the potential reader fully in mind" (7). Throughout composition and revision, the writer tries to keep in mind the different ways in which the reader might interpret the text. This advanced stage is therefore characterised by a three-way interaction between the author's representation, text representation and reader's representation held in the working memory (9). The writer constantly tries to anticipate the response of the reader to "make certain that readers see matters the way the author does" (9).[143] In sum, "[a]udience anticipation, therefore, means the comparison of the mental model built by a potential reader with the mental model of the intended text and the mental model of the already written text" (Linnemann 2019, 329).

The focus on 'knowledge' may seem problematic in relation to fiction writing. Does writing fiction involve gaining a better understanding of a subject? In some cases, it does. Take, for example, the writer's understanding of the fictional world. An example from Troch's writing process illustrates how 'knowledge' can be involved in writing fiction, and how this knowledge also changes and evolves during the writing process. By reconstructing his writing process, we can see how his knowledge of the fictional world he is creating still needs to develop as he writes. As mentioned above, the story "Mondini" is set in a dystopian world that has been greatly altered by climate change.

---

[143] Kellogg's concept of knowledge-crafting is based on the work of Walter Ong (1978), who argued that "a skilled author creates a fictional audience for the text to understand its meaning from the prospective readers' point of view" (Kellogg 2008, 9). According to Ong, the writer anticipates the different interpretations the reader may have of a statement to "correspondingly clarify meaning and to cover it suitably" (Ong quoted in Kellogg 2008, 9). In doing so, the writer must imagine what the reader already knows and adjust their text to adhere to this knowledge, for instance by making something less or more explicit. Taking into account the reader, of course, also applies to writing fiction and it therefore relates to the concept of the *implied reader* within reception studies, as coined by Iser (1974). For Iser, the term 'implied reader' "incorporates both the prestructuring of the potential meaning by the text, and the reader's actualization of this potential through the reading process" (Iser 1974, xii). But even if the author of fiction has a reader in mind, they do not always have to facilitate the reader. As Hannah Sullivan points out in *The Work of Revision* (2013), revisions such as additions and deletions may even make the text more difficult. In the case of excisive revision in the writing of Modernist writers, the text becomes more complicated since the explanatory frames and syntactic connections are removed. And in case of revision by extension, the reader also has more work to do: the overdescription and parataxis tend to make the sentences less grammatical and comprehensible (Sullivan 2013, 149).

In the first writing sessions, however, Troch did not yet fully conceive how his dystopian world differs from the one we know. In Session 4, he had written that all appliances that run on electricity have lost their practical use, which means that there is no more electricity. In the following session, Troch writes that the first-person narrator tries to maintain a pattern in his days. Since many people start their day with a shower, this was probably also one of the first things Troch thought of when writing about how the narrator starts his day as he writes: "De dag begint met een douche" (n186).[144] Then he adds "Zolang er stromend water is, is er hoop" (n187-189).[145] He probably remembers that he had previously mentioned that there is no electricity, so he adds: "Hoe het komt dat er geen elektriciteit maar nog wel stromend water is" (n190-196).[146] This can be interpreted both as a reminder to address this in the story and as a question he is asking himself. A little later he does a Google search to find out if it is possible to have running water without electricity: "kan er stromend water zijn zonder elektriciteit" (n203).[147] This leads him to the webpage "Leven zonder stromend water en elektriciteit" (n204).[148] Eventually, Troch clarifies the situation by explaining that the narrator has to go outside to get water; it no longer comes out of the tap.

This seems to suggest that Troch was freely producing text in order to describe the daily whereabouts of the narrator when he made the initial statement that the narrator starts his day with a shower and that there is still hope when there is running water. As he was writing, he probably realised that it was very unlikely that there would be running water without electricity. His search on the internet shows that he was in fact not quite sure, so he updated his knowledge on this topic of electricity and running water with some information he found online. Through writing, therefore, he 'discovered' an aspect of his fictional dystopian world that he had not yet accounted for. By expanding his knowledge of the subject, he also gained a subject to write about. Having discovered that it is impossible to take a shower without electricity, he could elaborate on the actions the narrator needs to take to wash himself. Thus, by focusing on the aspect of discovery in writing, we can broaden our understanding of the genesis of the texts, and see how knowledge of a particular topic develops throughout a creative writing process.

5.3.2 Associations

To account for discovery through writing, Galbraith (1999; 2009) developed the knowledge-constituting model. A central role in this knowledge-constituting model is given to the writer's disposition, since the source of content involves "a dispositional

---

[144] Translation: "The day begins with a shower".
[145] Translation: "As long as there is running water, there is hope".
[146] Translation: "How come there is no electricity but there is still running water".
[147] Translation: "can there be running water without electricity".
[148] Translation: "Living without running water and electricity"

dialectic in which content is produced as a consequence of a dialectic between the writer's implicit disposition and the emerging text" (Galbraith 1999, 146). The writer's disposition is their implicit knowledge concerning a particular topic. This is represented as a semantic network of interconnected units, in which the units are linked by positive or negative connections of fixed weight – also known as the principles of neural network theories or parallel distributed processing (PDP). According to Galbraith, it is the 'dispositional dialectic' that is referred to when writers describe writing as a process of discovery (Galbraith 1999). Yet, he acknowledges that content can also be produced by retrieving it from episodic memory.[149] Galbraith's knowledge-constituting model therefore denotes a dual process model. There are two ways to produce content in the knowledge-constituting model: by retrieval from episodic memory, or by synthesis during translation (Galbraith 1999, 146). The second involves the aforementioned dispositional dialectic in which a dialectic between the writer's disposition and the emerging text leads to the creation of new content (Galbraith 1999, 146).

While Galbraith already referred to the knowledge-constituting model as a dual process model, ten years later (2009) he made this more explicit by distinguishing between a knowledge-retrieval system and a knowledge-constituting system within a dual-process model of writing. In the knowledge-retrieval system, knowledge is retrieved from long-term memory, which "may or may not be evaluated and manipulated in working memory in order to satisfy rhetorical goals" (Galbraith 2009, 21). Since this involves only existing stored knowledge it "cannot by itself lead to the discovery of new ideas" (21). The second component – the knowledge-constituting system – on the other hand, involves "the synthesis rather than the retrieval of content" and operates "by using an implicit representation of knowledge" (22). It is capable not only of "synthesizing individual utterances but also, through inhibitory feedback from working memory to the writer's disposition, of producing a sequence of content without requiring external goals to drive the process" (22). It is therefore a process of discovery: "the knowledge implicit in the connections between units is only realised as explicit content once it has been formulated as potential text" (22). Galbraith and Baaijen (2018) later referred to the same two processes – knowledge-retrieval and knowledge-constituting – as respectively the knowledge-transforming process and the knowledge constituting process. Like Galbraith (2009), they note that both these processes are necessary for effective writing and that they complement each other.

We can relate the knowledge-constituting process to the process of making associations. In the Dutch-Flemish literary television programme *Brommer op zee*, the Dutch author Niña Weijers reflected on her writing process, in which she emphasised the process of association: "Je gaat als je schrijft… Het is een soort natuurlijk proces.

---

[149] As Galbraith and Baaijen point out, the episodic memory "consists of a memory of ideas as individual, already created objects" (Galbraith and Baaijen 2015, 262).

Je associeert heel veel. Je denkt… En die associaties zijn niet altijd logisch of realistisch. Maar die zijn er wel" (Weijers in Wechgelaer 2021).[150] This associative process could be aligned with Galbraith's notion of disposition: the network of the writer's implicit knowledge that can be activated through the act of writing.

Two examples from Van Rijswijk's writing process illustrate how this can work in the case of fictional writing. In the first writing session (Session 0), Van Rijswijk writes a passage in which the I-narrator – who is walking through the cemetery Zorgvlied with her mother – describes some of the gravestones. The narrator and her mother pass stones with pictures and stones with bad poetry and stones without names and mossy patches with only a number. Then, after a short pause of just over one second, Van Rijswijk continues that the narrator thinks how you might save space by being buried upright. To this sentence, she adds, 18 seconds later, that the narrator thinks this because she suddenly 'becomes economical' (n361). She then goes on to write that burying upright is, of course, undesirable for all kinds of reasons (n362-366). The first reasons are of a practical nature: the graves would be too deep and cumbersome (n367), and a corpse would decompose upright, causing everything to sag (n368-372). It took Van Rijswijk between two and three seconds come up with these practical reasons. After another pause of 4 seconds, Van Rijswijk continues writing another reason why it is undesirable (n373-379). This reason is of a more associative nature: someone who dies in times of standing desks and stand-up brainstorms, deserves the time to lie down and never get up again.

Van Rijswijk thus ended up interpreting the 'upright position' or 'standing position' as a sign of our time in which we are motivated to 'stand upright' more – something the I-narrator seems to find ridiculous. This interpretation allowed her to formulate things that she associated with this 'standing culture', which for Van Rijswijk were standing desks and stand-up meetings (later revised into stand-up brainstorms). But to finally arrive at these associations, she first had to formulate more obvious reasons for the undesirability of upright burying. We might understand this as an illustration of Galbraith's model of the 'knowledge constituting process', which accounts for the aspect of discovery in writing. As described above, in the knowledge constituting process an utterance is produced by the activation of connections between units in the writer's disposition. An input to the network – here the question of why upright burying is undesirable – activate the units and this activation "is then passed round the network according to the strength of the connection between units" (Galbraith 2009, 20). The pattern of activation that the network eventually settles on "represents the writer's response to the external constraints and corresponds to the message they want to convey", and which is subsequently translated into language and written down as an

---

[150] Translation: "You go when you write… It's a kind of natural process. You associate a lot. You think… And those associations are not always logical or realistic. But they are there".

utterance (21). In this case, the first practical reason, that the graves would be too deep and impractical, was the connection that initially received the most strength (and is also the most obvious). It is important to note that this initial utterance can "pass inhibitory activation back to the writer's disposition" (21). As such, the activation of units corresponding to the initial message will be reduced, which will result in a different activation pattern for the same input to the network but will thus create a different output. For Van Rijswijk this was first the upright decomposition of the corpse, and finally, when these practical reasons no longer stood in the way of more creative expressions, she was able to 'discover' the association of standing desks.

The second example concerns the first-person narrator's ability to perceive the ghosts of her (great-great)-grandmother. In Session 4, Van Rijswijk writes a passage in which the narrator describes the moment when she saw her grandmother's ghost for the first time during a visit to the optician. After a pause of more than 14 seconds, Van Rijswijk begins this sentence by writing "Ze zat naast" (n49), but immediately deletes 'naast' (n50). She pauses for 4 seconds, and then writes "zomaar naast me" (n51) and pauses for another 18 seconds before deleting: "zat zomaar naast me" (n52). She then starts again and writes "stond zomaar naast me toen mijn ogen werden gecontroleerd door de opticiën" (n53).[151] Although this sentence appears to have been produced in one flow, Van Rijswijk actually paused for two seconds before writing "toen mijn ogen werden gecontroleerd door de opticien". Finally, after about 10 seconds, she deletes the full stop to produce a new end to the sentence: "die haar uiteraard niet waarnam".

> [49]Ze ~~zat~~ ~~naast~~[50] ~~zomaar naast me~~[51][52] [53]stond zomaar naast me toen mijn ogen 's ochtends vroeg[94] werden gecontroleerd door de opticiën.[54], die haar uiteraard niet waarnam.[55] (Session 4)

The association involved in producing this sentence is clear: where better to perceive a ghost for the first time than at the optician, who helps you to 'see' better. Although we do not have think aloud protocols to determine the initial moment when Van Rijswijk made the association, the pauses in the composition of the sentence indicate that she was thinking about the formulation while writing. We can therefore hypothesise that Van Rijswijk discovered that the narrator first saw her grandmother's ghost at the optician's because she was thinking about 'seeing', which activated the connection in her disposition with 'opticians'. This is not verifiable – this association could well be made long before the particular writing session – and thus remains a hypothesis. For genetic criticism, however, this has always been the primary goal, "to construct a series of hypotheses on the operation of writing" (Grésillon 1997, 106). We cannot see into the mind of the writer, and that is just as well. Nevertheless, these examples from the

---

[151] Translation: "suddenly stood next to me when my eyes were checked by the optician".

writing process of Van Rijswijk illustrate how the keystroke logging data, and the theories of the cognitive processes involved in 'discovery' in writing, can be relevant to genetic criticism to help us understand the creative process, or in this case, how associations are being made.

5.3.3 Simulations and revisions

Galbraith's knowledge constituting process coincides with the argument made by Patrick Colm Hogan in *How Authors' Minds Make Stories* (2013). Hogan argues that simulation is a pivotal means by which authors create stories, and he uses the framework of narratology to construct this argument. By simulation, Hogan refers to "our ordinary cognitive process of following out counterfactual or hypothetical trajectories of actions and events in imagination" (2013, xiii). Thus, for example, imagining what you would do if you had to survive alone in a dystopian world due to climate change – or what your character would do in such a situation – as Troch did for his story "Mondini". To explore how simulative processes operate, Hogan uses some ideas from neural network theories or PDP (parallel distributed processing), which are "simplified versions of brain operations" (2013, 15) and thus draws on the same principles as Galbraith's notion of the writer's disposition. As Hogan points out, these neural networks consist of circuits, which in turn consist of sequences of connections between neuron-like units (15). An input can then activate certain units within the network:

> Each unit has some level of activation. When that level of activation reaches a certain threshold, then the unit fires. The activation of the firing unit is then sent along the connection to the next unit in the circuit. Connections may be excitatory or inhibitory. If the connection is inhibitory, then the recipient unit loses activation. Any given unit will receive inputs from many afferent (or incoming) units and convey output to many efferent (or outgoing, thus recipient) units. Finally, connections between units may differ in strength. (Hogan 2013, 15)

As this can become very complex, rules can be devised that attempt to approximate the processes of neural-network activation (Hogan 2013, 15). Hogan discusses the final parts of William Faulkner's *The Sound and the Fury* as an example of how such simulations, in terms of neural networks and rules, work with regard to literature: he illustrates how a biblical account of Jesus' resurrection "has in part guided Faulkner's particularization of the somewhat schematic story about a young woman pregnant out of wedlock" (Hogan 2013, 19; see also Chapter 6).

An example of such a process of simulation can also be found in Van Pelt's writing process as she writes about a harmless incident during the lunch of the excursion to Flamingo Island. In this particular case, Van Pelt's simulation of this situation enables

new text production, but also initiates revision. At a certain point in the seventh writing session, Van Pelt begins to write new text at the end of the text produced so far to describe the moment when the first-person narrator and her son Finn sit down at a table to eat their lunch. The provided Tunisian lunch consists of fish, couscous and grilled vegetables, and Finn has already expressed that he does not like it and therefore has no intention of eating it. So the narrator is sitting with two plates in front of her, when Mustafa joins them. Suddenly, sand falls into their plates.

While writing this scene, Van Pelt has to simulate the first-person narrator's reaction: the narrator looks up with the intention of discovering the source of the sand. Then Van Pelt has to simulate Mustafa's reaction; she has to imagine how he will react now that his lunch is covered in sand. She does this by first writing that Mustafa looks up as well: "Ook Mustafa kijkt omhoog" (n107-111).[152] Van Pelt then changes this to Mustafa looking around, followed by a deep sigh: "Ook Mustafa kijkt rond en zucht dan diep" (n112-116).[153] This change already introduces Mustafa's slight annoyance; in her simulation of the different ways in which Mustafa might react, Van Pelt thus opts for an annoyed reaction. She then chooses to further emphasise this reaction by adding that Mustafa also looks at the narrator angrily (n120-123), which she then moderates to sullen (n124-125): "Ook Mustafa kijkt rond, vindt de oorzaak van de zandregen en kijkt me nors aan".[154] At an earlier stage in the writing process, Van Pelt had already described Mustafa as a man who 'seems to be looking sullenly ahead', but she had deleted this description in Session 5. In terms of the neural network, therefore, Mustafa still activates units that relate to a 'sullen-looking man'.

The description of Mustafa's facial expression then also immediately triggers the addition of a new sentence in the paragraph above. This new sentence also describes Mustafa's appearance, he has dark and watery eyes: "Zijn donkere ogen zijn koud en waterig" (n126).[155] This revision (in the form of a new sentence) thus clearly relates to the text that was produced at the leading edge, the place where Van Pelt wrote new content. The new sentence partly explains Mustafa's facial expression: the addition of his dark and watery eyes ensures that his gruff reaction does not seem to come out of the blue.

Van Pelt then continues to produce text at the leading edge: Finn is the instigator of the flying sand, as he is digging a tunnel in the ground. Like a little dog, he kicks up the sand, some of which ends up in the plates. The narrator reprimands her son, who begins to cry. Again, Van Pelt has to imagine Mustafa's possible reactions. She begins

---

[152] Translation: "Mustafa also looks up".
[153] Translation: "Mustafa looks around as well and then sighs deeply".
[154] Translation: "Mustafa looks around as well, finds the cause of the sand rain and looks at me sullenly".
[155] Translation: "His eyes are cold and watery".

with "Mustafa houdt zijn d" (n153).[156] She then deletes "houdt zijn d" (n154), continues with "schenkt ons geen blik" (n155-157),[157] which she also deletes (n158). Then "Mustafa heeft zijn bord opzij geschoven" (n159-161).[158] The first pre-contextual deletion is hard to interpret; it is difficult to ascertain what Van Pelt intended when she wrote "houdt zijn d". The second revision provides more clues: Mustafa does not intend to deal with the situation any further and therefore does not pay any more attention to it. With this reaction, however, Van Pelt also limits her possibilities to develop the scene further. This may have been the reason for her decision to delete it and focus on an action Mustafa takes: he pushes his plate aside.

In the simulation of the situation, this action seems to activate a unit related to the plates, which evokes a new idea in Van Pelt's mind, as she turns her attention to the text produced so far. She replaces "onze borden" with "zijn bord" (n162-163).[159] And Finn digs a tunnel near Mustafa, not near their table (n164-172). Van Pelt also deletes "onze borden in" (n173), the sand ending up in *their* plates.[160] This implies that it is only Mustafa's plate that gets 'sanded'. Now, the narrator can offer her second plate to Mustafa, since her plates are not covered with sand. In this way, Van Pelt has also found a 'purpose' for the two plates in front of the I-narrator. As his characterisation already suggests, he refuses the offer. The scene continues with Finn firmly telling Mustafa that his father is dead, looking at Mustafa warily. Mustafa looks at the narrator for a translation of what Finn has said, then stands up, puts his hand briefly on Finn's hair and disappears. The scene has come to its end.

Of course, mental processes may have unfolded differently, and the decisions I have described here may or may not have been made in reality. However, hope that this example, and the ones above, illustrate how the theories of discovery in writing in the geneticist's 'toolkit' (especially, but not limited to, those working with keystroke logging data) can highlight aspects of the writing process that offer a new perspective on how the text developed (or might have developed) during the writing process.

5.4 EXOGENESIS 2.0

The text produced so far and the author's movement through the text can enhance our understanding of the endogenesis because it allows for a nanogenetic perspective on the textual development, likewise the nanogenetic aspect applies to the exogenesis as well. It is stating the obvious, but in the age of the Internet, authors have unlimited access to all kinds of information, from synonyms to detailed information on any

---

[156] Translation: "Mustafa keeps his d".
[157] Translation: "does not give us a glance".
[158] Translation: "Mustafa has pushed his plate aside".
[159] Translation: "our plates" into "his plate".
[160] Translation: "into our plates".

subject, from Google Books to Google Maps – to name but a few. The author Tom McCarthy calls this "a blessing and a curse at once", since "you can find out instantly which year Egypt won independence or who Persephone's mother was, but that essential solitude you need to write gets more and more elusive …" (McCarthy quoted in Johncock 2011a). As Knopse et al. point out, "this [continuous access to sources] has implications for the conceptualisation of the writing process" (2019, 258).

Within the field of writing studies, the use of online sources has therefore already gained some attention, including its use in translation processes (e.g. Ehrensberger-Dow and Perrin 2009; Knopse et al. 2019; Schrijver and Van Waes 2022). Knopse et al., for example, observed the use of online resources among Swedish students when writing in German, and they distinguish between students *controlling the sources* and *controlled by the sources* (2019, 265). Schrijver and Van Waes investigated a translation process of a Japanese novel: *Sekai kara neko ga kieta nara* (1979) by Genki Kawamura, which was translated into Dutch by Luk Van Haute. The complete translation process was logged with Inputlog. During the examination of the data, they could distinguish three phases in the translation process, and in each of these phases Van Haute spent large parts of the time consulting sources. In the first two phases around 40% of the time, and in the third phase 30%. In total, Van Haute consulted over 2000 sources and Van Waes and Schrijver also noted the high interaction with the sources – in general one interaction per minute, both for task related and non-task related consultations. Most task related sources were related to realia and frequencies of formulations (Schrijver and Van Waes 2022).

Another way in which sources have been studied is in source-based writing (e.g. Perrin 2003; Leijten et al. 2014; 2019; Vandermeulen et al. 2020). These studies focus, among other things, on how sources are transformed by professional writers, such as journalists and communication designers. Journalists often get source texts from other news agencies, which they process and transform into new texts (Perrin 2003). In line with this, Leijten et al. observe that "professionals writing from sources tend to focus on analyzing what others have done textually and visually, distilling best practices for the genre and gleaning ideas for invention" (2014, 288). As such, source texts are paraphrased, and visual and verbal content of others is sampled for their own projects (288).

This touches on the way in which genetic criticism has examined sources. Within genetic criticism, the use and appropriation of source texts in literary writing processes has always been an important aspect of analysis, usually referred to as the exogenetic research. De Biasi, for example, noted that authors frequently borrow from other literary texts, but that these source texts can become invisible during the endogenesis:

> After undergoing its multiple transformations, the initial exogenetic element (for example, a topographical comment, a detail or a situation borrowed from

a literary work) may have become perfectly untraceable: it has metamorphosed into an organic part of the text, which, for the reader, points only to the writer's imaginary and to the internal logic of the fiction, just like any other element of the work. (De Biasi and Wassenaar 1996, 46)

Dirk Van Hulle distinguishes between two phases in the study of the exogenesis: an 'upstream' and a 'downstream' phase. This first phase is about uncovering what caught the eye of the author, what interested them and thus "to uncover the text's possible exogenetic dimensions" (Van Hulle 2022, 134). It is therefore primarily concerned with discovering and identifying the source texts. With the second phase, the 'downstream' research, genetic criticism distinguishes itself from *critiques des sources* because it involves the investigation of how the author processes this exogenetic material (137). This is where for the author the act between revealing and hiding comes into play, and it is up to genetic criticism "to map the gradual appropriation and integration of material in the text, which involves a process of mixing and merging materials from a variety of sources" (137).

While there are many studies investigating the exogenetic dimensions in the work of modernist writers, for example James Joyce (e.g. Crowley and Van Hulle 2016), no research has been carried out into how present-day literary authors make use of sources, including those on the Internet. Since Inputlog also logs all the windows opened and Internet pages visited during the writing session, this offers the possibility for a detailed study in the usage of sources in digital literary writing processes.

I distinguish four categories in which the keystroke logging data can contribute to the study of exogenesis, taking into account both 'upstream' and 'downstream' exogenetic research. To clarify these categories, I propose to align them with parts of an abstract representation of a tree. Imagine the final text as being the leaves of the tree, the trunk and the branches of the tree refer to the textual witnesses of the writing process (the versions and the keystroke data), and the roots are the actual sources that were used to write the text.

Botanical metaphors have been used more often to describe the act of writing (Van Hulle 2006; 2017), for example with reference to the 'rhizome': "[t]he usually horizontal underground plant stem that produce shoots above and roots below" (Van Hulle 2006, 217). The focus, when making use of a botanical metaphor, lies most often "on the underground aspects of the writing process" (217). Yet, as Van Hulle points out, the writing process should not be reduced to 'underground' activities', since literary composition involves both the downward and the upward movement. He states that the search for, and reading of source material is part of the 'underground' aspect of the text, but when this material is incorporated and processed in the drafts, it becomes a matter of inflorescence:

> The 'root' metaphor may be applicable to the search for sources of information exterior to the writing. But it is perhaps not quite adequate to visualize the processing of pre-textual material in the drafts. Once the author has decided that he may use a specific source text, the subsequent process is the most complex aspect: the incorporation of this foreign material in the drafts. This process should be visualized from the other direction – not top down, but bottom up – to show how a foreign element is integrated and where it ended up in the published text; or how it eventually did *not* make it into the final version by ending up in a textual cul-de-sac or aborted section that fell out of the direct line of textual descent. (Van Hulle 2006, 220)

Van Hulle (2017), has therefore used a plant metaphor to describe the genesis of a text, including its epigenesis. Here, he relates the endogenesis to the endodermis of a plant: the endodermis regulates the passage of substances drawn from the soil, just like the endogenesis involves the selection, incorporation, and transformation of exogenetic material (Van Hulle 2017, 417). In the same way, the inner workings of a tree are more complex than the eye can observe: the functions of the trunk of the tree therefore perfectly lends itself to the processing of the source material as present in the versions and the keystroke logging data.

The first category – the roots of the tree – focuses primarily on the sources and looks quantitatively at how many and which sources were used in the writing process. Like the roots of a tree, these sources may not be directly visible, but lie beneath the surface. At times, however, the roots come to the surface of the ground; this is when the intertextuality may get noticed by the reader. In this category, all the roots are uncovered. The keystroke logging data can also be instrumental in determining how a source has been used, thus shifting the focus to the trunk and the branches of the tree. The trunk of a tree has several functions, including two layers responsible for transporting nutrients and water. Just below the outer bark is the inner bark, called the phloem. This tissue moves sap up and down the tree, and nutrients produced in the leaves are transported through the tree to the roots. The sapwood, or living xylem cells, carry water and minerals from the roots throughout the tree.

The second category – the sapwood of the tree – takes the roots (the sources) as its starting point, but then examines how the sapwood transports water from the roots (the 'inspiration') through the tree (the textual traces of the writing process) up to the leaves (the final text). This category thus examines how the sources were transformed and appropriated during the writing process, and what the moment at which the sources were consulted, and how the sources were finally incorporated into the text.

The third category – the inner bark of the tree – is more related to 'downstream' exogenetic research. Sometimes there is a source and a final text, but it is uncertain whether the source was actually used. By starting with the final text and the sources,

and reconstructing the route between them, insights can be gained into whether the author indeed incorporated the source text into their own text.

The fourth category – the branches and twigs of the tree – starts with specific, detailed parts of the writing process (the branches) and examines whether the original source (the roots) can be traced and how it is ultimately used in the final text (the leaves). Authors may not always have referenced the actual original source, or may not have visited the website during the logged writing process. The keystroke logging data as well as the text produced so far, however, can provide indications of the actual source, even after the fragment referring to the source has finally been removed from the text – when the leaf has fallen from the tree.
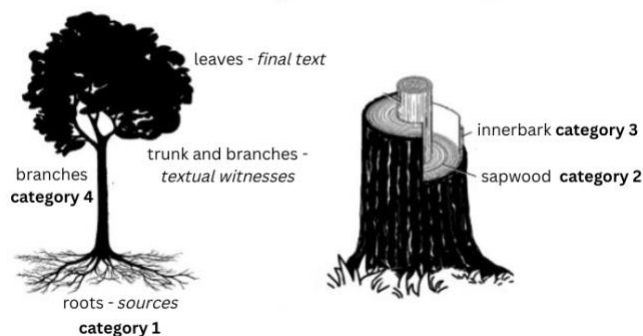

Figure 5.13: The tree as a metaphor to understand how keystroke logging enhances our knowledge of the exogenesis, image of tree taken from Baughman (2019, 33) and modified to include the different exogenetic categories.

5.4.1 Category one: the roots of the tree

The first category serves to obtain insight into the actual use of the Internet during writing: the roots of the tree. Since Inputlog logged all the websites visited by the authors during the writing process, we have a wealth of data to study the exogenesis. What served as inspiration and what sources did the authors consult before and during writing? What remains of the exogenesis of modernist writers is often just the tip of the iceberg, but the exogenetic keystroke logging data seems to provide almost the entire 'iceberg' – *almost*, that is, since there could have been many influences that have not been documented. Or as Van Hulle puts it: "What they [the scholarly editors] have uncovered is not necessarily all there is" (2022, 141). This remains the same with keystroke logged digital writing processes.

A separate category in the actual use of the Internet during writing are the websites that the author visits as part of the downtime. These websites cannot be related to the writing of the text, but as mentioned above, Leijten et al. (2014) still consider it to be an important part of the writing process, as it could increase productivity when finally returning to the text and could even stimulate new ideas. This also comes to the fore in writers' personal accounts. For example, Joanne Harris notes that briefly scrolling

through Twitter can actually increase her concentration: "I use Twitter as a kind of mini-break for the mind – a minute or two away from the writing often helps me concentrate better" (Harris in Johncock 2012b). Louise Doughty, then, referred to what Leijten et al. term 'involuntary downtime': "What really distracts me when I'm writing are the concrete realities of life; doing my tax return, or that urgent email to my agent, or the school run. What's completely disastrous is if I open my email" (Doughty in Johncock 2012a). According to Leijten et al., this involuntary downtime could also have a positive impact on the writing, but for Doughty specifically, it does not.

The first way to understand the use of resources is to take a quantitative approach: to get an idea of how many sources the authors used. For Bogaert, there are two sources for his use of sources: the Atoma notebook and the keystroke logging data. The Atoma notebook contains 63 sources and references to sources, which can be divided into the following categories: 26 literary sources, 24 sources from newspapers, 6 printed sources originating from the Internet, 3 quotations, an information brochure about the planned construction work at the actual Franklin Roosevelt square in Antwerp, a photograph, a source from a magazine and a map taken from Google Maps. Nine characters are also 'given a face' by Bogaert by including portraits of people he found in newspapers. In addition to external sources, 'reality' was an important source for Bogaert. He went to Franklin Roosevelt Square several times to sit and watch the action and the people for a couple of hours. He noted down everything that caught his eye, which allowed him to write realistic descriptions of the square. However, as these are his own impressions, they are not included in the classification of external sources.
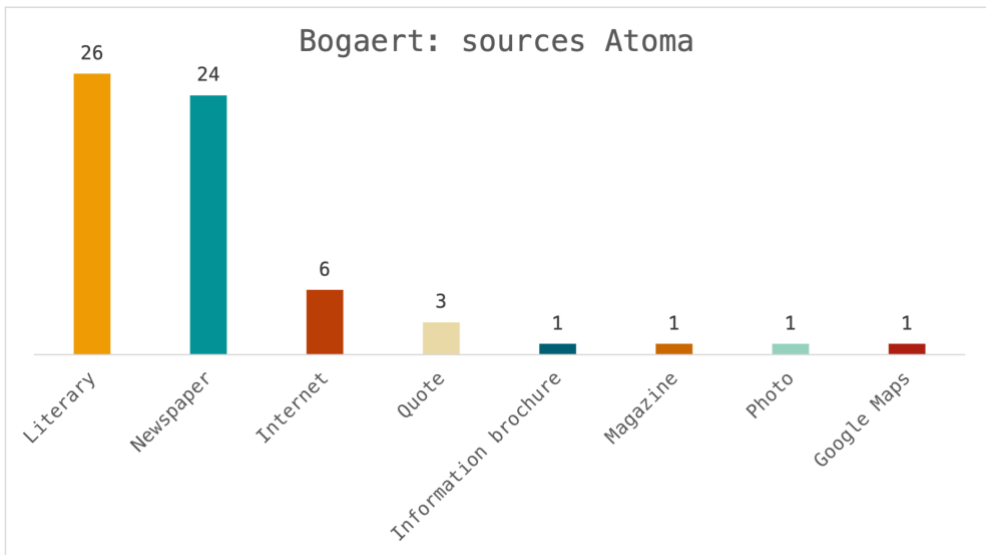

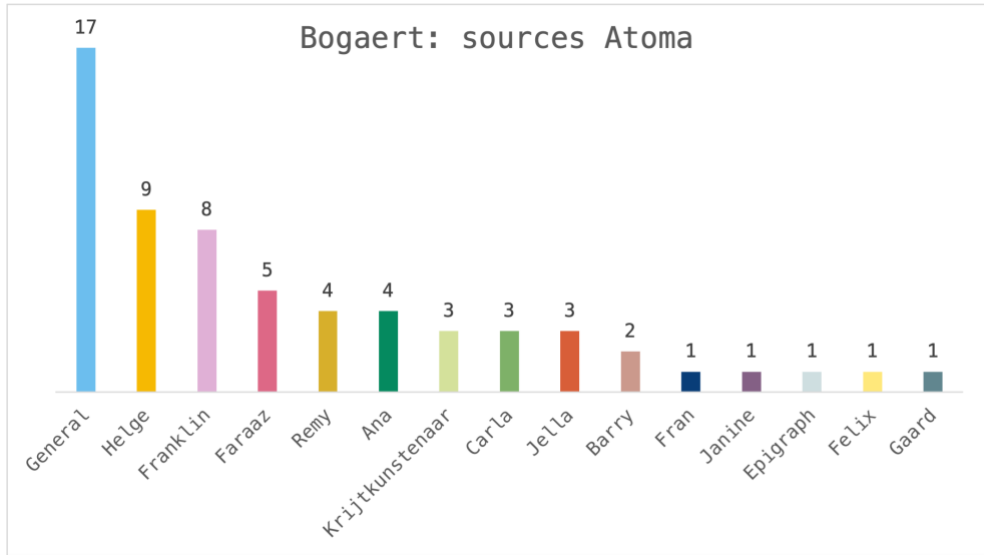Figure 5.14: Categories of sources in the Atoma notebook.

Figure 5.15: Sources in the Atoma notebook per character.

But what were Bogaert's intentions with these sources? Figure 5.15 shows that most of the sources were related to the novel in general. This is mainly due to a reading list with several novels and references to novels that related to ideas Bogaert had in mind for *Roosevelt*. In terms of the characters, Bogaert included most sources for the character Helge, followed by the character Franklin. For these characters Bogaert had already gathered a lot of information in advance that could support him during the writing process.

Regarding Bogaert's online behaviour, Inputlog logged a total of 2183 relevant 'focus events' during the entire writing process, i.e. the unique websites and other windows that Bogaert opened per writing session. This means that if Bogaert googled 'Murakami' in one session, and googled it again in the same session, this would only count as one focus event; if Bogaert googled it again in the following session, then this specific Google search would appear twice in the list of focus events.

Bogaert however, did not always visit websites during writing. He only did so in about 55% of the sessions, which means that during the remaining 45% of the sessions he was only writing in Microsoft Word. In addition, 76% of all windows were related to subjects that had nothing to do with writing the novel, compared to 24% that did.
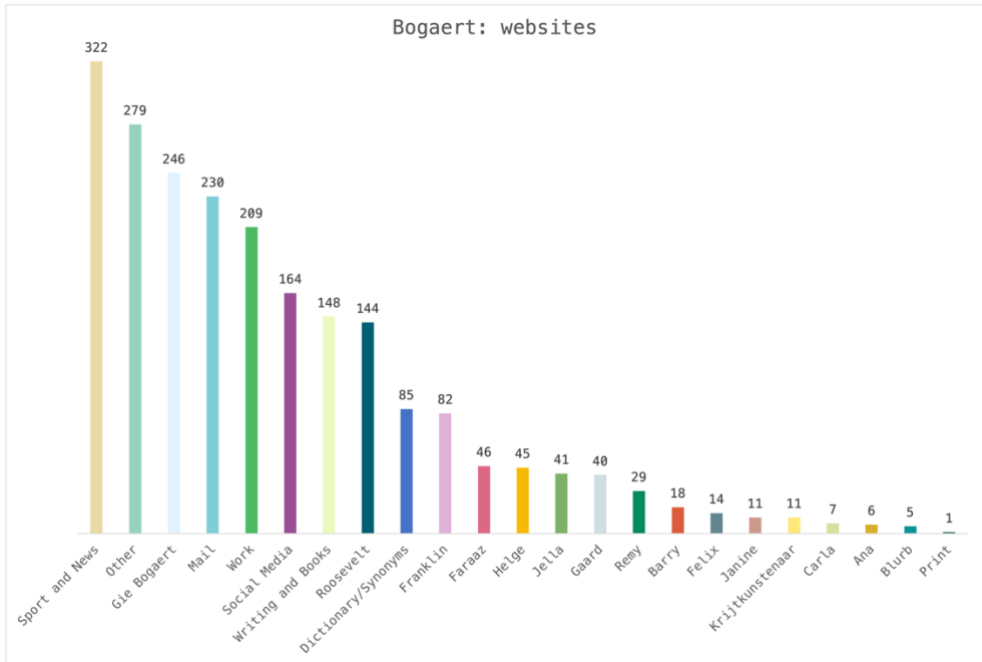
Figure 5.16: Online exogenesis per category.

Figure 5.16 shows that Bogaert visited the most websites for topics unrelated to the writing of the novel. He visited mainly sports and news websites, and there is also a large category of 'other', with all sorts of things that could not be further categorised, such as internet banking and looking for a new house. He visited a lot of websites related to himself, such as the backend of his website and reviews of his previous novels, which can be seen as a form of self-regulatory behaviour (cf. Chapter 5.1), and often checked his mail. Other distractions from writing were work-related, including his work as a teacher, but also making arrangements to speak at literary events. Social media is also a fairly large category, and he visited websites related to writing and books. Only then do we come to the websites that Bogaert visited that were related to writing. The category 'Roosevelt' consists of everything that has something to do with writing the novel, but not so much with finding information about the content. For example, he visited the website of his publisher De Bezige Bij, note apps and other files related to the novel. Bogaert also visited 85 websites that would lead him to the 'right word'. These are all the home pages of websites with dictionaries, spelling tips and synonyms. The actual searches for specific words are categorised according to the section of the novel – the characters – for which Bogaert was searching. In total, he visited 185 websites in 62 sessions related to spelling, words and synonyms. A closer look at the writing actions that followed these searches showed that Bogaert did not blindly choose the first best synonym. Often he would look at the synonyms, choose one to try, and then choose another.

The other categories are all characters in *Roosevelt*. Bogaert searched most often for information for the character Franklin, for which he visited 82 websites. This is not surprising, given that Franklin's sections take up most of the space in the novel and Bogaert therefore also spent most of the sessions writing these chapters. For Franklin, the search terms are the most eclectic. He did a lot of searching on Google, or directly on websites for synonyms. Among other things, he looked for the correct spelling of iPhone, the names of fast-food chains, and the tram line in Antwerp. Almost all of the terms are related to the descriptions of the square at each of the times described, which Franklin mainly does.

Bogaert consulted 46 websites to write Faraaz's sections. Most of the search terms for Faraaz are related to Islam and Iranian culture; and to describe Faraaz's nocturnal walks through Antwerp, Bogaert had to look up some street names. In one of the sections of Faraaz, Bogaert included a reference to the dog in Japan that, after the death of its owner, continued to go to the train station to wait for him, as it had always done. Bogaert had read about this dog in an article on animal behaviour in *Time* magazine, "The Mystery of Animal Grief" (April 15, 2013), which he had included in the Atoma notebook. One of the sentences in the article describing this dog is: "One Akita in Japan went to a train station looking for its deceased owner every day for a decade" (36). To include this in the novel – as something that Faraaz remembers – Bogaert first wrote: "Er is het verhaal van een hond" (Session 203).[161] Then he searched on Google for "akita", where he would read that the FCI (Fédération cynologique internationale/International Canine Federation) classifies the Akita as a Keeshond. With this information in mind, he changed "hond" into the two options "akitahond/keeshond". He then opted for "akitahond", but in Session 442 he generalised it back to "hond".

The 45 websites that Bogaert visited to write the sections narrated by Helge are more language-oriented; Bogaert often looked for synonyms. He also searched for words for birthmark and some medical terms, as Helge is a general practitioner. For the sections on Jella, he visited 41 websites, including searches for street theatre in Avignon, Greenpeace, and a lot of information about abortion. For the literature professor Gaard, Bogaert visited 40 websites. He searched, for example, for restaurants in Antwerp with a Michelin star; for the line "April is the cruellest month" by T.S. Eliot; for the authors Murakami and Salinger; and for synonyms, the meaning, and the correct use of words. The 29 websites that Bogaert consulted for Remy's sections (the tram driver) were related to food, as Remy is rather fond of food. Among other things, Bogaert searched for words describing food, and especially sweets: "chipolata", "snoepgoed", "snoepgigant", "snoepwinkel", "geldhof", and "marshmallows". Bogaert visited 18 websites for Barry's sections. Barry's heart is broken and he takes his revenge by

---

[161] Translation: "There is the story of a dog".

shooting people in the square from his hotel room. To be able to write this, Bogaert looked for the right weapon and whether it could be available in Antwerp. The 14 websites he used to write the sections of Felix had to do with pickpocketing tricks and mobile phone brands. Janine is the toilet attendant at Hotel Terminus, which is also reflected in the 11 search terms and names of the websites Bogaert visited: the words urinal, Dettol, and bleach, among others. For the sentences "Ik moet morgen nog bleekwater kopen. Vijfennegentig cent voor een vijfliterbus bij Aldi",[162] Bogaert first had to look up the price of bleach. To be able to write the sections narrated by Carla, Bogaert visited 7 websites; on Google Maps he searched for the district of Hoboken in Antwerp, and he consulted the correct spelling of some words. For Ana, he consulted 6 websites: a list of Moroccan boys' names and ways to hurt oneself without leaving marks.

Overall, there is a significant difference between the online consultations and the sources in the Atoma notebook. In general, the way Bogaert uses the sources corresponds to his general working method. The content-related sources are mostly found in the Atoma notebook, while the sources for details, fact-checking, language and spelling are consulted online. The former therefore fits perfectly into the 'creative' process and the latter into the 'linguistic-creative' process. The usage of the internet sources that were included in the Atoma notebook and the internet sources that he consulted while writing also differs: the digital, but hard-copy sources are used more extensively and in more detail than the fact-checking information that he searched for later.

For Posthuma, Van Rijswijk, Troch and Van Pelt, there is only the keystroke logging data – and the eventual story – to work with. Van Rijswijk was the only author who did not consult any sources while she was writing her story in Microsoft Word. To categorise the sources used for the other three stories, the websites were grouped into the following categories: 'Setting', for sources that can be related to the description of the setting of the story; 'Character', for websites that were used to describe a character, or to obtain knowledge that the character is supposed to have; 'Event', for sources that were used to describe a specific situation or event; 'Object', for sources that were used to describe a specific object; 'Style', for websites that were used for the style of the story; and 'Dictionary/Synonyms', for the meaning, spelling, or synonyms of words. In addition to these content related categories, the authors also visited websites as part of their downtime, for example to check their mail, listen to music, visit social media, read the news, or organise something for their work.

---

[162] Translation: "I have to buy bleach tomorrow. Ninety-five cents for a five-litre can at Aldi"
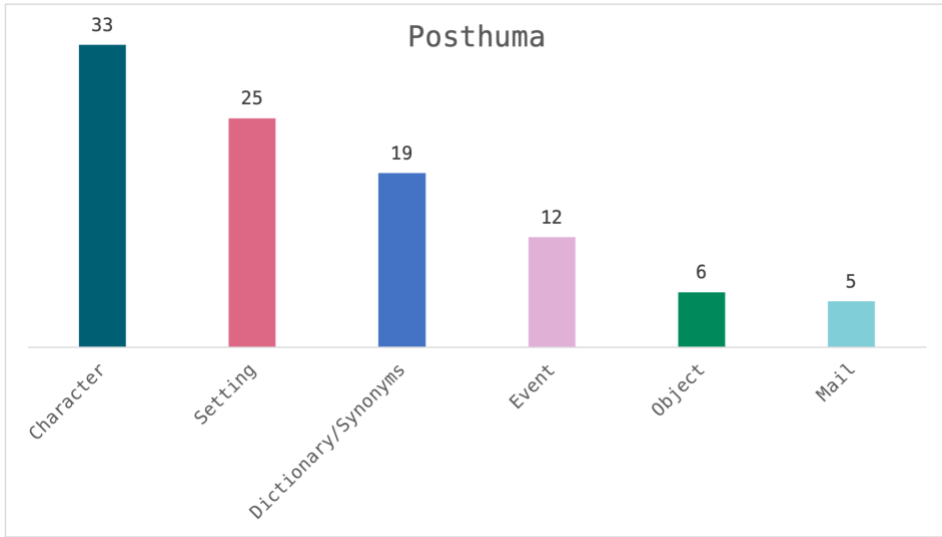
Figure 5.17: Online exogenesis per category for Posthuma.

Posthuma went online during nine out of 38 sessions and visited 102 relevant websites (unique websites or Google searches per session). Most of the websites consulted, 33 in total, are related to the writing of the characters in the story, especially to the description of the I-narrator's friends. The male friend is a researcher specialised in the ecological humanities, and for writing the descriptions of his job, Posthuma googled, among other things, "ecologie", "ecologie geesteswetenschappen", "ecologische geesteswetenschappen", "sun energy frederick douglass", "frederick douglass solar energy", and "the end of boring weather". For the female friend she googled, among other things, Andrea Long Chu, Valerie Solanas, and the theory of desire. Posthuma consulted 24 websites for the setting of her story, she googled for family summer house and visited websites of hotels in Croatia. She visited a website related to spelling and synonyms 19 times, among other things to look for synonyms for "spijt" and "zelfspot". For the description of certain events, Posthuma consulted 12 websites. After the 'stuck-finger incident', in which her son's finger got stuck in a pedalo, the I-narrator has to think about a brochure in the hospital about the treatment of children in the emergency room and the way nurses recognise child abuse, for which Posthuma googled "eerste hulp onderzoek kind mishandeling", "spoedeisende hulp informatie folder onderzoek kind mishandeling", "spoedeisende hulp kind folder", and "spoedeisende hulp kind folder olvg", which led her to the pages "Behandeling op de SEH" on the website of the hospital OLVG in Amsterdam, and "Behandeling van kinderen op de SEH" on the website of ZorgSaam. For the 'near-suffocation incident', Posthuma googled "verslikken" and 'Heimlich manoeuvre'. Finally, for the description of specific objects, Posthuma consulted six websites by googling the words "waterfiets glijbaan", "leunstoel massagestoel", 'massagestoel', and the ice lolly "calippo".
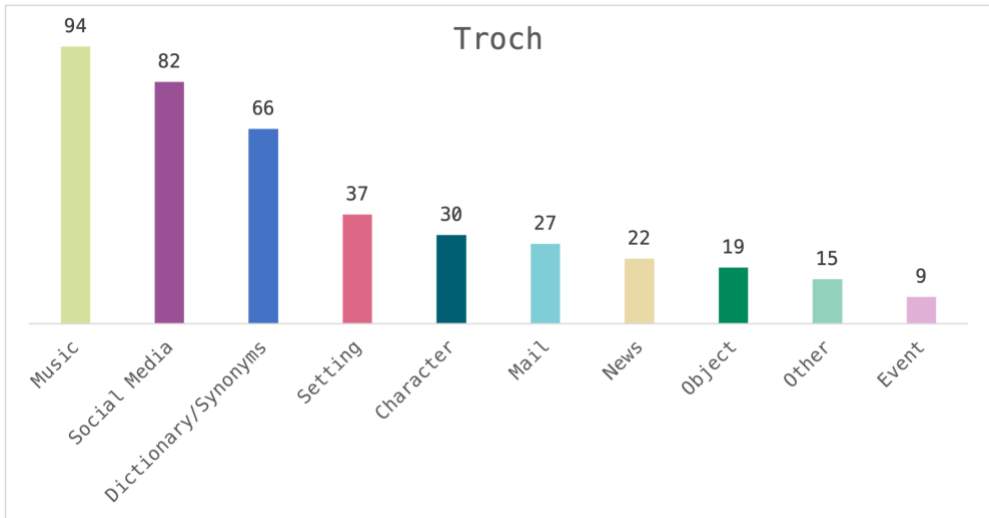
Figure 5.18: Online exogenesis per category for Troch.

Apart from logging into to her email account during one writing session, the keystroke logging data shows very little downtime. However, Posthuma admitted in an interview during the Track Changes symposium that she deliberately did not log websites that were not related to the writing of the story. Her writing process also includes some long pauses – during which she may have had some downtime – and she used "Schermvergrendeling" (the lock-screen) for between fifteen minutes and one hour and 45 minutes.

Troch visited websites in 18 out of 20 sessions and consulted 401 relevant websites (unique websites or Google searches per session). The two largest categories are related to downtime. 94 websites involve searching for, and listening to, music, mainly on YouTube. The music Troch listened to included songs and tracks by Beirut, OP8, Morphine, Angèle, CocoRosie, Beck, The Jon Spencer Blues Explosion, and L7.[163] Troch was also very active on social media while writing the story, but this did not seem to interfere with the writing process; he could write, interrupt it by answering a chat message, and then resume writing. He also checked his email quite often, read some news and visited some websites that could not be categorised further.

---

[163] Troch explained in an interview at the working conference of the Flemish Authors Association (Vlaamse Auteurs Vereniging; VAV) on November 26, 2022, that each story requires its own playlist.
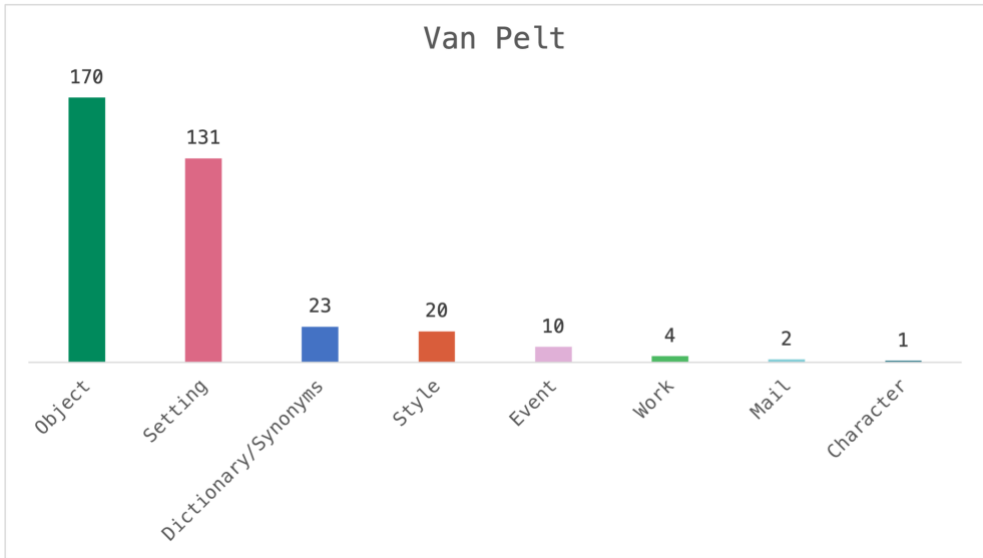
Figure 5.19: Online exogenesis per category for Van Pelt.

The category regarding synonyms is also fairly large, as he searched for synonyms for words such as "terechtkomen", "onevenaarbaar", "likken", "pietluttig", and "oogappel". For the setting, Troch consulted 37 websites. He searched for the highest mountain and ski areas in Italy; he used GoogleMaps to visit Mondini Cuneo Piëmont, which is only a small hamlet, and he gathered information about life without electricity, heat, the Stone Age and how much salt sweat contains. He used 30 sources for the description of the characters, most of which concerned the search for the names of the partner and daughter. For the I-narrator, Troch looked for information about Michel Lotito, the man who ate an aeroplane. For the description of objects, Troch used 19 websites and searches on Google. He looked for hieroglyphs, cave drawings, cuneiform script, calf's tongue, ski equipment, shop bell, and water bottle, among other things. Writing descriptions of some events required 9 searches and websites: what to do if you have a broken collarbone, what makes good snow for skiing, and how to ring the church bells. The information Troch sought online was therefore very practical.

For Van Pelt, 362 relevant websites (unique websites or Google searches per session) were logged in 14 of the 18 sessions. Van Pelt was only briefly distracted from writing the story to make arrangements for an interview and to check her mail. The largest category is finding information to describe objects. Most of it concerns the names of seashells. Van Pelt also consulted many sources (131 websites and Google searches) to describe the setting, as she searched for pirate ships, specific parts of ships, hotels in Djerba, Flamingo Island, and the music from *The Pirates of the Caribbean*. Most of these sources were essential for Van Pelt to be able to write the story; without them, she could not describe the setting in as much detail and the story could not progress.

Van Pelt also looked for synonyms, for example for the words "touwwerk", "krakkemikkig", "piraat", "steiger", and "danspas". One event described in the story required more attention from Van Pelt. On their way back from the excursion, the tourists on the pirate ship see some dolphins. The narrator then reflects on the fact that dolphins are the only animals that can commit suicide. So Van Pelt consulted some sources on dolphin suicides. The one search that Van Pelt made in the category 'Character' was a search for first names in Tunisia. A specific category in the sources used by Van Pelt is 'style'. During the writing process, Van Pelt looked for phrases and expressions from the maritime sector that she could weave into the story. In Session 8, for example, Van Pelt googles the expressions from the maritime sector, during this search she includes a list of maritime expressions and words in the Word document:

> ruimschoots
> Aanklampen
> Afschepen
> "Beter een anker kwijt dan het gehele schip". "Overstag gaan". "Driemaal is scheepsrecht".
> "Geen land mee te bezeilen". Schipperen "Hij is schipper te voet". "Olie op de golven". "Tussen kaai en schip gevallen". (Session 8)

Van Pelt used various expressions from the maritime sector throughout the story, including some that are not listed in this note, but she did not use the longer expressions between quotation marks. However, the words "ruimschoots", "aanklampen" and "afschepen" were included in Session 15. Van Pelt had prepared this when she was revising the text produced so far with pen and paper. To include the word "afschepen", Van Pelt revised an already existing sentence and added a new clause to it:

> Ik had <u>niet eens veel moeite gedaan om hem af te schepen</u> ~~had[33][32]~~ en had[34] hem het geld cash gegeven. (Session 15)

The word "aanklampen" was included by changing the sentence "Hij probeert het wat verderop" into "Hij klampt een jong koppeltje naast ons aan" (n69-76).[164] And Van Pelt wrote a new sentence to include "ruimschoots" as well: "We hebben ruimschoots de tijd om van dat paradijs te genieten" (n98-102).[165] The keystroke logging data thus provides a wealth of data for studying the (online) exogenesis, and with the example of how Van Pelt included some of the expressions from the maritime sector, we are already touching on the second category: the trunk of the tree, or how and why the sources are used during the writing process.

---

[164] Translation: "He tries it a bit further on" into "He accosts a young couple next to us".
[165] Translation: "We have plenty of time to enjoy that paradise".

5.4.2 Category two: the sapwood of the tree

Even more important than knowing which sources were used is the question of how these sources were transformed and appropriated. This is the focus of the second category: the sapwood of the tree. Van Pelt, for example, searched online for accurate descriptions of the setting of her story. In the reconstruction of the writing process captured in the first session, we can see that Van Pelt starts by jotting down some ideas for the story: a dolphinarium, the main character is (maybe) divorced, and (maybe) travelling with her parents. She mentions a hotel lobby, waiting for a bus that stops at various hotels, and a boat on the sea. So the idea of the boat was there from the start. Then she starts to draft a sentence: "We zaten op de zijbeuken van de boot" (n23).[166] She is probably unsure about the terminology of boats because this sentence prompts the first search on the Internet for "onderdelen piratenboot". This search then leads to a number of searches through which she visits a couple of websites:

> **Google**: onderdelen piratenboot
> **Google**: piratenboot
>   ➡ **Website**: Piratenboot | Wikipedia
> **Google**: geschiedenis piratenboot
>   ➡ **Website**: Piraat | Wikikids Hoe zijn piraten ontstaan? | willemwever
> **Google**: piratenschip delen van het schip
>   ➡ **Website**: Fragmenten en delen van een piratenschip, geïsoleerde beeldelementen | Piratenschip, Zeilen, Afbeeldingen
> **Google**: terminologie piratenschip
>   ➡ **Website**: Deze termen moet je kennen, om te begrijpen wat er in de wereld gaande is – complete lijst | Ellaster
>   ➡ **PDF**: Piraterij: Het Jack Sparrow complex - Laurent Samyn - Thesis Master in de Maritieme Wetenschappen
> **Google**: piratenschip tunesie - Google search
> **Google**: piratenschip djerba - Google search

The list shows that the successive search terms are initially quite specific, with a search for *parts* of a pirate ship; then they become more general with 'pirate ship', which in turn leads to a search for the 'history' of pirate ships. After this detour from the initial search for information – information about the terminology used to describe pirate ships – she searches for 'parts of the ship' and 'terminology pirate ship'. Then the search terms suddenly become specific to a particular location: first Tunisia, then Djerba. The question arises as to whether she already had this country – or even this island in Tunisia – in mind as the setting of the story, without having externalised it yet, or whether this idea arose during the online searches.

---

[166] Translation: "We sat on the side aisles of the boat".

Even with access to the keystroke logging data we cannot know what Van Pelt had in mind when she started the first writing session, but we can still look for clues that might have prompted this direction of locating the story in Tunisia. The page "Piraterij: Het Jack Sparrow complex", which Van Pelt visited before searching for 'pirate ship tunisia', is a pdf-file of a master thesis on how the international community deals with piracy today. She scrolls through the document for about 30 seconds (13:47:48-13:48:18). Interestingly, this thesis mentions Tunisia on pages 7 and 31 (Samyn 2011, 7, 31). Although it is uncertain whether this caught her eye within thirty seconds, it is a possible link we can make between this document and the search for pirate ships in Tunisia.

In any case, Van Pelt specified the location of her story during this search, and subsequently narrowed it down to Djerba. A search for 'piratenschip tunesie' returns many results that mention Djerba, which could explain the specification in 'piratenschip djerba'. For one of the results, Google previews the content of a page describing a trip to 'Flamingo Island', which Van Pelt subsequently notes in the Word document: "het eiland van de flamingo's" (n37-41).

After adding this note about Flamingo Island, Van Pelt returns to the sentence that prompted the search for parts of a pirate ship. She deletes it and starts over a few times. Then she writes "Het binnendek", searches for 'binnendek schip' and visits the Wikipedia page about 'Dek'. This makes her change 'binnendek' to 'bovendek'. After another visit to a page about decks, she continues her research on pirate ships in Djerba. This leads her to TripAdvisor pages that describe a tourist activity of going with a pirate ship to Flamingo Island. One of these pages shows some pictures of the trip. The clicks we can see in the keystroke logging data might indicate that she clicks through at least a couple of them. Also, some of the pictures show 'pirates' climbing into the ship's ropes. Seeing these pictures probably led to the insertion of the note: "piratenmuziek, kunstjes in de touwen". After inserting this in the Word document, Van Pelt visits the page "Djerba (Tunesië): informatie & reizen" from the online travel agency Reisgraag.nl. There, she finds another description of a boat trip to Flamingo Island. She copies it and pastes it in the Word document. This trip to Flamingo Island, which is also described on the TripAdvisor pages, will serve as the outline for the story.

Now that Van Pelt has set the story in Djerba, she starts searching for more information about the island. First, she visits Google Maps to locate the island and then she searches for hotels. This search leads to the hotel Fiesta Beach, which Van Pelt notes in the Word document as the name of the hotel where the narrator and her son are staying. She continues her search on Google Maps and visits "L'île aux Flamants Roses". By locating the Flamingo Island on the map, she could see that it is a peninsula, which she then notes in the Word document: "het is zelfs geen eiland, maar een

schiereiland" (n80).[167] She continues her search for information about Fiesta Beach, Djerba and the tourist trip to Flamingo Island. She visits TripAdvisor and the Facebook pages of the pirate trips, where she sees some pictures again. She interrupts her search to make some notes: "Kolonne van piratenschepen, volgestouwd met toeristen" (n94-102).[168]

During the remainder of the first session and some subsequent sessions, Van Pelt occasionally returns to the TripAdvisor pages to get more details about the excursion. General descriptions and pictures of the "tourist circus" – as the narrator calls it – sparked Van Pelt's imagination and helped her to develop her ideas for the story. The information Van Pelt gathered on the internet about the excursion to Flamingo Island served as a frame for the story, which she could fill with the narrator's perspective to tell a story about going on with life after the loss of a partner. The fact that the structure of the story is based on the excursion becomes particularly clear when we consider the content of the TripAdvisor page "Pirate Ship in Djerba by Vacances Promo Tunisie" into account. This page describes the programme of the excursion:

> Programma: Om 9.00 uur verlaat de boot de jachthaven en vaart naar zee met een aangename muzikale sfeer aan boord. Tijdens de reis zullen verschillende animaties worden voorgesteld. Als je geluk hebt, kun je wat dolfijnen spotten en je cameralenzen zien! Je zult ook een traditionele vissessie bijwonen. Op het eiland Flamingo's heb je voldoende tijd om te zwemmen en te zonnebaden. 'S Middags wordt een heerlijke lunch aangeboden in een prachtige omgeving. Het menu bestaat uit: Tunesische salade, baksteen Tunesische couscous, gegrilde vis, cola, muntthee, fruit van het seizoen. Na de lunch kun je een sessie-animatie bijwonen of terugkeren naar de kust. Om 14.00 uur voel je je weer opgetogen en klinkt het geluid van de lucht als je aankomt in de toeristenhaven van Houmet Souk. (TripAdvisor Ship in Djerba by Vacances Promo Tunesie)[169]

With this timetable with activities Van Pelt did not have to pay much attention to the aspects of events and setting – these were predefined by the 'real' excursion – but could instead focus on the narrator's experience. The writing process of Van Pelt is therefore

---

[167] Translation: "it is not even an island, but a peninsula".

[168] Translation: "Procession of pirate ships, packed with tourists".

[169] Translation on TripAdvisor: "Program: At 9h, the boat leaves the marina and heads to sea with a pleasant musical atmosphere on board. During the trip several animations will be proposed. If you're lucky, you can spot some dolphins and feats your camera lenses! You will also attend a traditional fishing session. On the Island of Flamingos, you will have enough time to swim and tan. At noon, a delightful lunch will be offered to you in a lovely setting. The menu includes: Tunisian salad, brick Tunisian couscous, grilled fish, cola, mint tea, fruits of the season. After lunch you can either attend a session animation or return to the shore. At 14h you return feeling elated and the sounds of cheer fill the air as you arrive at the tourist port of Houmet Souk." See: https://www.tripadvisor.com/AttractionProductReview-g2629138-d16653961-Pirate_ship_in_Djerba-Gabes_Governorate.html

characterised by an interaction with the pre-defined structure of the excursion, and by analysing the genetic development of the story we can get a sense of the creative development.

5.4.3 Category three: the inner bark of the tree

The third category – the inner bark of the tree – is more related to the 'downstream' exogenetic research; it gives us insight into whether and how the author incorporated source texts into their own text, and it also provides knowledge about the intertextuality of the text. According to Van Hulle, "a text's intertextual condition denotes the way in which it shows or hides its connections to other texts" (2022, 133). He terms the "balancing act between revealing and hiding, telling and not telling" the text's "genetic code" (133). In cases of invisible intertextuality – when there are no direct traces of the external source text in the published version – the intertext may only be perceived in the drafts and manuscripts (133). This also applies to the keystroke logged digital writing process, where the keystroke logging data itself may even indicate whether a source was actually used or not.

This is the case with the photocopy of a couple of pages from *Grote Europese roman* (2007) by the Flemish author Koen Peeters, which can be found in the notebook that Bogaert prepared before 'putting the text into words'. Bogaert placed it in the 'Decor' section of the notebook, indicating that he wanted to use it to describe the setting, but he did not specify which part of the novel it was for. In Session 179, however, Bogaert wrote a note to himself referring to Peeters' novel in the section narrated by Franklin, the square, in the overarching chapter 10 o'clock. In Session 174, Bogaert had already written another note to himself within this specific section: "(Hierin wat minder exterior. (Want minder observaties in K 73-83 over dit uur, evt. elementen uit andere 'uren' gebruiken. Toon: oude man!)".[170] This means that he only had a few observations of his own from the square made at ten o'clock in the morning, and as a solution he suggests getting observations from other hours. In Session 176, he adds another solution, which is to look at the fragment of *Grote Europese roman*, as he noted: "K.Peeters K73-4 /82-83".

In earlier references to this novel, Bogaert already expressed his interest in the 'impressionistic city descriptions' in this novel, but are there any similarities between the two texts that might indicate whether Bogaert actually used them? There are three details that are mentioned in both texts. Koen Peeters, for example, wrote a paragraph about a beggar who shouts "Jew" at the narrator, which makes the narrator linger about his Jewish origins: "'Wandelende Jood!' riep de tweede bedelaar. Hoe wist hij dat? Het

---

[170] Translation: "(Herein somewhat less exterior. (Because less observations in K 73-83 about this hour, possibly use elements from other 'hours'. Tone: old man!)". K. 73-83 refers to pages in the Atoma notebook.

was de eerste keer dat iemand hem zo noemde" (Peeters 2007, 248).[171] Bogaert also included a Jewish man in Franklin's section in the chapter 'Ten o'clock': "De jood met zijn hoge hoed en in zijn lange zwarte jas die ginder voorbij de Russische delicatessenzaak wandelt, lijkt wel de enige die geen last heeft van de warmte" (Bogaert 2016, 51).[172] Furthermore, Peeters mentions a dog – or perhaps two – that whines: "Ergens jankte een hond ingehouden. Of waren het twee honden?" (248).[173] In the section of Franklin, Bogaert also mentions a dog barking at a pigeon – or perhaps two: "Een hond blaft bezeten naar een duif. Of zijn het twee duiven?" (52).[174] And finally, *Grote Europese roman* mentions swallows flying around: "Er vlogen zwaluwen naar binnen en naar buiten" (249).[175] This detail is also present in *Roosevelt*: "Vlakbij duikt een vroege zwaluw naar een vlieg" (52).[176] Both thus speak of a dog, a Jewish man and swallows. But besides the striking similarity between "Of waren het twee honden?" (Peeters) and "Of zijn het twee duiven?" (Bogaert), these are such generalities that it hardly seems evident that Bogaert found these ideas in *Grote Europese roman*. It seems much more likely that Bogaert saw a Jewish man walking on Franklin Roosevelt Square, since Antwerp has a large Orthodox Jewish community, but none of these observations were written down in the notebook. Did the idea for these elements arise while writing, or was it the fragment from Peeters' novel that inspired him?

The keystroke material seems to point to the latter. In Session 178, Bogaert writes, consecutively but at different places in the text: "Een hond blaft bezeten naar een duif", then "Een verdwaalde jood in een … wandelt langs de/voorbij de" and lastly, "een vroege zwaluw duikt naar een insect/mug".[177] Immediately afterwards he removes the note "K. Peeters K73-4". Because this deletion takes place immediately after the insertion of *all* the text elements that also actually appear in the fragment of *Grote Europese roman*, I want to hypothesise that these elements from *Roosevelt* have an origin in the novel by Peeters. However, the elements are so general and small that this would not be possible to extract from the text alone, but by including the keystroke material in the analysis, we can in fact make a stronger connection between the two fragments. In this example, we knew beforehand that Bogaert intended to use *Grote Europese roman* while writing, and the keystroke logging data in turn assisted us in deciding whether,

---

[171] Translation: "'Walking Jew!' shouted the second beggar. How did he know? It was the first time than anyone called him that". This could, in turn, be an intertextual reference to *De wandelende Jood* (1906) by August Vermeylen or to the Christian legend of the wandering Jew.

[172] Translation: "The Jew with his top hat and in his long black coat walking past the Russian deli over there seems to be the only one not bothered by the heat".

[173] Translation: "Somewhere, a dog whined restrainedly. Or was it two dogs?"

[174] Translation: "A dog barks possessively at a pigeon. Or is it two pigeons?"

[175] Translation: "Swallows were flying in and out".

[176] Translation: "Nearby, an early swallow dives for a fly".

[177] Translations: "A dog barks obsessively at a pigeon", "A lost Jew in a ... walks past the" and "A early swallow dives at an insect/mosquito".

and in which part of the novel he used this source. This brings us to the third way in which the keystroke logging data provides information about the exogenesis, since the author may also have used sources without explicitly referring to these sources in the drafts or visiting the specific website during the writing session.

5.4.4 Category four: the branches of the tree

The fourth category – the branches of the tree – attempts to use the keystroke logging data and the session versions to determine the exact source used during the writing process, even if it is not explicitly listed in this material. An example can be found in the keystroke logging material from Posthuma. In the published version of "En daarom haten ze zichzelf", the narrator's friend briefly explained some 'theories of desire'. However, during the genesis of the story, this part about desire was much more elaborate. In earlier session versions of the text, the friend refers to the work of Andrea Long Chu, and Posthuma incorporates some of her statements. During the writing process, Posthuma alternates between translating and directly quoting Chu's ideas. Posthuma began incorporating Chu's ideas as early as Session 14 (17 September 2020). In this session, she first wrote "Je hebt …, zei ze, die zegt dat…" (n91), indicating that she did not have the name at hand.[178] After a long pause, during which she locked the screen, Posthuma added the name: Andrea Long Chu (n132). Then, in the next session, on the same day, she added the first quotation:

> [6]~~Z~~[7] [9]Chu zegt ook: I dont think we believe th~~oin~~[10][11]ings because ~~we think~~ [13]they are true.[12] ~~W~~[14] <u>I think w</u>[15]e believe th~~ings~~[16]<u>em</u>[17] because they make us feel a particular way. (Session 15)

Although it appears to be a direct quote, the exact source is not visible in the keystroke logging data. She either has it written down on paper or on some other device. In Session 19 (24 September 2020), Posthuma expands on this passage about Chu's ideas by adding, among other things, "You could sooner give a cat a bath" (n43). Again, no direct source was logged, but since this is a very specific sentence, a search on Google on 'Andrea Long Chu You could sooner give a cat a bath' quickly gives us the relevant article, which was published in the journal *n+1*. In this case, it is the text that leads us to the source. Then, in Session 21 (28 September 2020), Posthuma again adds quotations from Chu, the complete reconstruction of the writing process of this paragraph is given in Table 5.5.

During this session, Posthuma actually turns to the Internet to make use of Google and Google Translate. All the pages that Posthuma visited are listed below.

---

[178] Translation: "You have …, she said, who says that …"

> **Google**: valerie solanas
> **Google**: google translate
> ➡ **Google Translate**: outlandisch claim
> **Google**: sceptical about the existence of men
> **Google**: sceptical about the existence of men andrea long chu
> **Google**: valerie solanas
> **Google**: has doubts about the existence of men andrea long chu
> **Google**: valerie solanas
> **Google**: andy warhol
> **Google Translate**: tendrils

It is noteworthy that she only searches on Google – or that only these searches are actually logged. Therefore, the 'focus events' in keystroke logging data does not provide any information about which source text she used. Nevertheless, the text provided by keystroke logging data *does* lead to its source. Quite early in writing this paragraph, Posthuma actually writes "In een filmpje vertelde ze […]" (n10-22).[179] This is a clue that, together with the Google search "sceptical about the existence of men andrea long chu", leads to a video on YouTube. In this video, Chu explains her ideas and mentions Valerie Solanas, who shot Andy Warhol. So this video seems to be the source that Posthuma found before this session, because the quote "I dont think we believe things because they are true. I think we believe them because they make us feel a particular way" is also from this video. The clause that pointed us to the direct source – the reference to the video – was deleted from the sentence during this writing session; the only way we had access to this information was through the keystroke logging data. Without it, we were still left in the dark as to the exact source, as googling the sentences does not yield any results as they were not textualised – the words were only uttered in the video.

As can be seen in the visualisation of the writing process of this paragraph, Posthuma directly quotes or directly translates Chu's statements – and for the translation of some words she used Google Translate. The result at the end of this session is a mixture of the two, in fact a mixture of the source text and its appropriation. While the full passage is later moved to the end of the document, becoming part of the notes, and is eventually deleted from the story altogether, the genetic material still helped us "to figure out […] what the author was interested in" (Van Hulle 2022, 135). Having uncovered this, the passage still resonates when reading the published story as the friend mentioned the 'theories of desire'. Using the keystroke logging data, we can thus analyse the exogenetic dimensions in great detail, even if the source is eventually not even used in the published text.

---

[179] Translation: "In a video she told […]"

JENTE POSTHUMA	ANDREA LONG CHU
[7]Toen ik weer alleen was googlede ik Andrea Long Chu en kwam ik erachter dat ze een witte trans-vrouw was ~~die~~ [89]en nog slimmer dan mijn vrienden. In ~~een filmpje [23] vertel~~[22]~~t~~de[1615]~~de ze over haar boek~~ [21] haar boek [24]Femai[1817]le stelt ze dat iedereen vrouwelijk is en dat iedereen dat haat en [66]dat ze skeptisch is op[26]~~ver~~ het bestaan van mannen [25]en over die bizarre clam[3635]~~ims~~, zoals ze die zelf noemt, die broad, indefensible, sweeping generalisaties over iedereen in de geschiedenis, zegt ze: het gaat mij erom een react[67]ie op tre[3837]e roepen bij de lezer.[39] ~~haar stelling dat iedereen vrouwelijk is~~[4019] ~~over ... zei ze~~[4110] [12]~~Chu zegt ook:~~[11] I dont think we believe things because they are true. I think we believe them because they make us feel a particular way. [20]En extreme provocatieve [42]uitspraken ~~maken~~[43] kunnen[44] ons bewust maken [46]van de emoties ~~die we~~ [47] [45]in argument or the production of truth[48]. [49]Je gh[68]eeft de lezer een kans om zich te realiseren: dit zit me dwars, niet omdat het verkeerd/fout[50] is[51], ~~maar~~ [52] [53]op inte~~klk~~[54] [55]llectueel niveau, maar omdat ~~ik~~ [56] [57]I have investmen~~s~~[69]ts[70] in some things being true and other things not being true, and those investments, that is the tendrils of desire. [59]P~~o~~[6061]rovication can be an education of the senses if you let it. [63]The trick of desire is not to ~~h~~[6465]get crushed by it. (Session 21)	I make lots of outlandish claims in the book. I say everyone is female, I say everyone hates it, I say I'm very skeptical about the existence of men, I make these broad, sort of indefensible, sweeping generalizations, about, like everyone in history, these crazy, crazy claims. And I think it's right to say that the point of that is to provoke a response in the reader. I don't think we believe things because they are true, I think we believe them because they make us feel a particular way. And what a shocking, or provocative statement can do, if done correctly, is bring to consciousness the fact of emotion in argument or in the production of truth. You give the reader an opportunity to realize this bothers me, not because I think it's wrong on an intellectual level, but because I have investments in some things being true and other things not being true, and those investments that is the tendrils of desire. […] It is, I think provocation can be an education of the senses, if you let it. The trick of desire is not to get crushed by it. (Chu in Verso Books 2019, 4:43-6:58)

Table 5.5: Comparison between the text written by Jente Posthuma, and the transcription of the video of Andrea Long Chu, which was used as source.

5.4.5 Some notes on the leaves

Of course, it cannot be ruled out that there are more sources that have informed the writing, but which have not left a trace in the genetic material. In *Roosevelt*, for example, the character Gaard – who is a literature professor at the University of Antwerp – quotes various authors. Most of them are not mentioned in the Atoma notebook. For example, a quotation from a novella by Saki, which Gaard knows by heart:

> Ze herinneren me aan een passage uit een novelle van Saki. Ik ken die uit mijn hoofd: 'Clyde deed zijn best, en heel erg goed zijn best, iets van een feestmaal in hun voortdurende woestijnpicknicks te leggen, maar zelfs sneeuwgekoelde Heidsieck verliest zijn bouquet wanneer je de overtuiging bezit dat de donkere wijnkelner die hem met zoveel neerbuigende charme serveert, slechts wacht op de goede gelegenheid om je hals af te snijden.' Het was een voorbeeldzin in mijn vroegere cursus narratologie. (Bogaert 2016, 99)[180]

The reference to this citation occurs for the first time in the work document in Session 256. During this session, Bogaert pasted a draft into the section of Gaard. This draft already contains the citation from the work of Saki:

> Verwijzen naar de donkere kelners bij Gustav. Faraaz. Citaat van Saki, pseudoniem van Hector Hugh Munro over donkere kelner! Ik ken het, kan citeren uit mijn hoofd: 'Clyde deed zijn best, en heel erg goed zijn best, iets van een feestmaal in hun voortdurende woestijnpicknicks te leggen, maar zelfs sneeuwgekoelde Heidsieck verliest zijn bouquet wanneer je de overtuiging bezit dat de donkere wijnkelner die hem met zoveel neerbuigende charme serveert slechts wacht op de goede gelegenheid om je hals af te snijden.' Het is een van mijn favoriete zinnen. (Session 256, 45)[181]

The citation comes from the story 'Cross Currents' from *Reginald in Russia* (1910). We are not able to reconstruct the exact writing process of this passage, as it was not logged. Yet, we can see a glimpse of the moment the idea for the citation came to mind: "Citaat van Saki, pseudoniem van Hector Hugo Munro over donkere kelner!" This is clearly a

---

[180] Translation: "They remind me of a passage from a novella by Saki. I know it by heart: 'Clyde did his best, and very, very well his best, to put something of a feast into their prolonged desert picnics, but even snow-cooled Heidsieck loses its bouquet when you are convinced that the dark wine waiter serving it with such condescending charm is only waiting for the right opportunity to cut your neck.' It was a sample sentence in my earlier narratology course."

[181] Translation: "Referring to the dark waiters at Gustav's. Faraaz. Quote from Saki, pseudonym of Hector Hugh Munro about dark waiter! I know it, can quote from memory: 'Clyde did his best, and very, very well his best, to put something of a feast into their prolonged desert picnics, but even snow-cooled Heidsieck loses its bouquet when you are convinced that the dark wine waiter serving it with such condescending charm is only waiting for the right opportunity to cut your neck.' It is one of my favourite phrases."

note Bogaert wrote to himself. The citation that was eventually included also raises some questions. Did Bogaert – like his character Gaard – know it by heart and retrieve it from his long-term memory? Did he have a copy of the book? Or did he search for it online? To arrive at a possible answer, we thus need to consult the original passage in English and the Dutch translation of the story by Stella Bromet.

> Clyde deed zijn best, **en zijn best was niet gering**, om hun langdurige woestijn-picknicks enigszins op echte diners te laten lijken, maar zelfs sneeuw-gekoelde Heidsieck verliest zijn bouquet wanneer je **ervan overtuigd bent** dat de donkere wijn**schenker** die hem met zoveel eerbiedige stijl serveert, slechts wacht op **een gunstige** gelegenheid om je **de strot** af te snijden. (Saki 1983, 105)

> Clyde did his best, and a very good best it was, to infuse something of the banquet into their prolonged desert picnics, but even snow-cooled Heidsieck lost its flavour when you were convinced that the dusky cupbearer who served it with such reverent elegance was only waiting a convenient opportunity to cut your throat. (Saki 1910, 99)

It appears that the quotation given by Gaard does not quite correspond to the Dutch translation – the differences are indicated in bold. The differences are subtle but may indicate that Bogaert did not copy the quotation directly from the Dutch translation. Two hypotheses remain: either Bogaert translated the quotation himself, based on the English original, or he deliberately changed some of the words to show that recalling a quotation from long-term memory is very error-prone; even a professor of literature makes mistakes in his quotations. This indicates that not all questions can be answered with the keystroke logging data.

Another example of sources that were not present in the keystroke logging data relates to the story "Zorgvlied". As mentioned above, Van Rijswijk's keystroke logging data does not contain any references to possible sources. However, this does not mean that there has been no 'seed incident'. Charlotte Doyle uses the term 'seed incident' to the events "which writers identified as the events that had begun their stories" (Doyle 1998, 30). Like Bogaert, she actually went to the setting of her story. In an interview with the magazine *De Groene Amsterdammer*, Van Rijswijk clarified that she had indeed taken the described walk through the cemetery with her mother: "Ik heb die wandeling over de begraafplaats daadwerkelijk gemaakt met mijn moeder, en ik heb als een vanzelfsprekendheid die geesten om ons heen geschreven" (Van Rijswijk in Meines 2022).[182] The real cemetery and her impressions of it thus functioned as a source – as

---

[182] Translation: "I actually took that walk through the cemetery with my mother, and I wrote about the ghosts around us as if it were self-evident".

inspiration – during the writing process. Another, more conceptual source is mentioned in this interview as well.

> In *De terugkeer,* van Esther Gerritsen, komt een man voor die vanuit het hiernamaals naar zijn kinderen kijkt. Hij kan ze echter alleen maar zien als zij aan hem denken. Ik ben niet religieus maar als het hiernamaals zou bestaan, zou het misschien wel zo werken. Ik heb van *De terugkeer* geleerd dat je zo'n idee ook in je literatuur kan verwerken. Daardoor kon ik mijn voorouders achter mij aan laten lopen in 'Zorgvlied'. (Van Rijswijk in Meines 2022)[183]

These examples from the sources used by Van Rijswijk thus show that even when we have all the keystroke logging data for an exogenetic analysis, some sources will remain unnoticed. This also applies to the writing process of "Dauphin" by Van Pelt. In a personal conversation, she clarified that she had actually taken the holiday to Djerba and had also taken a pirate ship to Flamingo Island. The story was thus based on her personal experiences, and she searched for information on TripAdvisor to recall the memories of the trip. It is clear that we have more than just the tip of the iceberg, but there is still ice below the surface that goes unnoticed. Anything that was not consulted online during the logged writing sessions, or otherwise documented and shared, cannot be observed by the geneticists. This can only be uncovered by consulting the author.

5.5 CONCLUDING REMARKS

The detailed temporal dimension of writing, which can be studied through keystroke logging, was the focus of this chapter. It is through this temporal dimension that we can make claims about authors' writing habits. While authors' personal accounts of their writing habits have always been an important source for studying working methods (e.g. Zimmerman and Risemberg 1997), keystroke logging now allows us to study them without having to resort to the authors' own narratives about how they approach their writing tasks. Using keystroke logging data, we can, for example, observe when and for how long the author worked on the literary text. The types of revision used to reconstruct the writing processes provide additional information about how each of the authors wrote and revised their text. The clearest distinction can be made between those writers who revise a lot during text production (pre-contextual revisions) and those who revise mostly after the sentence has already been completed (contextual revisions). The keystroke logging data can also shed light on more unconscious habits. Ellen Van Pelt,

---

[183] Translation: "*De terugkeer*, by Esther Gerritsen, features a man who watches his children from the afterlife. However, he can only see them when they are thinking of him. I'm not religious but if the afterlife existed, maybe that's how it would work. I learnt from *De terugkeer* that you can also incorporate such an idea into your literature. This allowed me to have my ancestors walk behind me in 'Zorgvlied'".

for example, said in a personal conversation that she was surprised of how many revisions she made during the production of a sentence; she herself had the impression that she had a sentence well in mind before she started typing it.

The keystroke logging data makes it possible to study the (so far) smallest aspect of the writing process, thus enabling a new type of text genetic research: the study of the nanogenesis. Nanogenetic research is mainly concerned with the sequence of the writing actions, and to establish – were possible – connections between them. In addition, it allows us to distinguish typing styles, which reintroduces the singularity of the 'word processed' document. As I have illustrated with examples from the writing process of David Troch and Roos van Rijswijk, the nanogenesis makes it possible to examine how the literary text is made coherent and how the text produced so far guides new text production. Moreover, nanogenesis also allows the study of typing habits, which highlights the uniqueness of each word-processed document.

The question now is how useful this nanogenesis is for genetic criticism. Is it relevant to know that Troch typed the word 'knipoogde' at 4:07 p.m. – or even at 4:07 and 18 seconds to be precise – but made a typo (he wrote 'knkip), which he immediately corrected? Probably not, but that alone is not the reason why a nanogenetic analysis might be interesting for genetic criticism. It *would* be interesting, for example, if a nanogenetic analysis could show that this typo actually triggered the insertion of a new word. However, I would like to emphasise that the nanogenesis should not be reduced to the time at which typos are made, or other individual writing actions. The main possibility that I want to address with nanogenesis is precisely the *possibility* to look beyond individual writing actions and to start looking for connections between them; it is precisely the sequentiality – which is given in the keystroke logging data – that allows us to trace the development of the (literary) text in more detail. Thus, we are able to make hypotheses about the choices made during the writing process. In fact, it is precisely nanogenetic analysis that increases the probability of finding the needle, such as a meaningful revision campaign, in the haystack of keystroke logging data. The author's movement through the text makes us aware of the interconnectedness of textual elements that would otherwise be easily overlooked. Not only can the nanogenesis put forward the author's textual awareness, it can also create textual awareness on the part of the reader/geneticist.

Nanogenesis also contributes to our understanding of discovery in writing. Galbraith (2022) noted that it may be time to consider the actual text production in order to improve our knowledge of discovery, and the reconstructions of the writing process allow us to do just that. The reconstructions of the writing processes enable us to make hypotheses about how knowledge expands during writing, how associations underlie the text production, and how simulation is part of the writing process. An example from Troch's writing process illustrates how knowledge about a story develops

during the writing process. Two examples from Van Rijswijk's writing process show how associations can drive text production. And an example from Van Pelt's writing process demonstrates how we may reveal the processes of simulation – in which the author imagines a particular situation in the story – and how the characters could react to it.

Let us return to my tree metaphor one last time. A nanogenetic analysis can also be applied to the use of sources; the keystroke logging data allows us to see how many sources the author used during the logged writing process and, more importantly, *when*. The metaphor of the tree highlights how keystroke logging data has significant implications for our understanding of exogenesis. Not only do we have access to most of the tree's roots (the sources), but we can also examine the sapwood: just as water is transported from the roots throughout the tree, the sources are transformed to become part of the text. Starting with the leaves – parts of the text – and the roots, we can also move back and forth across the inner bark of the tree to see if these parts can be connected. And from certain parts of the writing process – the branches of the tree – we can find clues to the real roots. But sometimes we see a leaf but are unable to see the branches that connect it to the tree and down to the roots. The same applies to keystroke logging; we cannot know all the sources and inspirations used by the authors.

CHAPTER 6. NANOGENETIC NARRATOLOGY

WHERE NARRATOLOGY MEETS KEYSTROKE LOGGING DATA

When you write, you lay out a line of words. The line of words is a miner's pick, a woodcarver's gouge, a surgeon's probe. You wield it, and it digs a path you follow. Soon you find yourself deep in new territory. Is it a dead end, or have you located the real subject? You will know tomorrow, or this time next year. (Dillard 1990, 3)

How are stories and novels written? And what takes place in the mind of the author in the moment of writing? In *How Authors' Minds Make Stories*, one of the examples Patrick Colm Hogan discusses regarding simulation in terms of neural networks concerns William Faulkner's *The Sound and the Fury* (Hogan 2013, 17). As Hogan points out, the final chapters of *The Sound and the Fury* deal with, among other things, the characters Quentin Compson and her uncle, Jason. As the illegitimate daughter of Caddy Compson, Quentin lives with her uncle who keeps her away from her mother. Jason treats Quentin cruelly, stealing the money her mother sends her, for instance. Quentin, in turn, is characterised as "rebellious and somewhat wild", and seems to be pregnant, signalled "by her great need for money and her unwillingness to explain that need" (Hogan 2013, 17). After summarising the final chapters as such, Hogan argues that seeing the pregnancy within a neural network account of Faulkner's simulation makes it unsurprising: "There is presumably a strong associative connection between Quentin and Caddy, a strong connection between Caddy and premarital pregnancy, and a strong connection between wild behavior and pregnancy, making pregnancy an obvious narrative possibility for Quentin" (17). Hogan also suggests that the pregnancy offers two "simulative possibilities", which are based on episodic memories of (past) cases: "having the child or having an abortion" (17). According to Hogan, both will have strong connections within a PDP account (Parallel Distributed Processing account; see Chapter 5.3), and that there may also be "connections between having the child and suffering familial cruelty" (18). This leads him to the following analysis:

> Because there is already cruelty from Jason toward Quentin, this idea (or unit) of 'suffering familial cruelty' should receive particular activation. This should occur at two levels: in the author's simulation of future possibilities for Quentin and in the author's simulation of Quentin's simulation of future possibilities. […] In more rule-like terms, then, we could say that the author would expect Jason to be cruel if he discovers the pregnancy, and the author would envision that Quentin has the same expectation. The latter simulation would be strongly aversive for Quentin. Thus one would expect her emotional response to incline toward an abortion. The abortion would, of course, be connected with the acquisition of money as well as some sort of escape from Jason's supervision, perhaps along with travel to another city where she would not be recognized. The difficulty – indeed, near impossibility – of an abortion in these circumstances is likely to activate a further possibility at each level (i.e., for Quentin, for the author, and for the reader). That further possibility is suicide. (Hogan 2013, 18)

According to Hogan, Faulkner needs to choose which option he will follow and decide how to develop it. Some circumstances, such as the time in which the story takes place, "have direct, causal impact" on the specification, while others do not have such a direct, causal impact but still "guide the particularization of the narrative" (18). An example of the latter is the story taking place on Easter Sunday, having "the effect of partially activating or priming a number of religious stories relating to Jesus' resurrection" (18). After analysing the set of activations caused by the biblical model, Hogan concludes that this seems "to have served as model in Faulkner's development of Quentin's escape from her family" (26). While Hogan's analysis of *The Sound and the Fury* in terms of neural networks leads to an apt interpretation of the cognitive processes involved in creating a story, the linearity of his interpretation does in fact merely make it an interpretation of how Faulkner eventually created the final version (worked out the storyline) – and not of how the story was originally created in the course of its genesis. This is probably because the analysis is based on the published version, and not on the manuscripts and drafts (if they exist).

It is plausible to imagine that Faulkner started with the biblical model of Jesus' resurrection which in turn activated associations that he could use to refine the story, but the first part of Hogan's analysis (of Quentin's pregnancy and subsequent decision to run away) raises a couple of questions, since this would entail a perfect alignment between the genesis and the course of the events in the story. Could it not be possible that he started associating 'backwards' from the biblical model, leading him eventually to introduce the idea of a pregnancy to give a possible reason for her disappearance? Or, that Faulkner's initial idea was to let Quentin disappear *because* of her (apparent) pregnancy, without needing to simulate these other possibilities in the first place? In other words, the simulation process can be deduced from the narrative, but did the genesis also develop along these lines? Hogan's interpretation belongs to one of the

possibilities of how the story was created, but the drafts could possibly help to substantiate this hypothesis – or reject it. And this is where narratology and genetic criticism, as well as the insights from writing studies, especially those from cognitive writing process research (such as the theories regarding discovery), may be combined to interpret the cognitive processes involved in the writing process of literary works.

The critical dimension of genetic criticism entails the reconstruction of the genesis of a text from a chosen point of view, like, for instance, a narratological perspective. In "Narrative across Versions: Narratology Meets Genetic Criticism" (2013), Lars Bernaerts and Dirk Van Hulle already advocate combining narratological analysis and genetic criticism. According to them, this can be productive in at least two ways:

> On the one hand, if one's object of study is a text's discourse, genetic criticism's reconstruction of the writing process can be an aid to the narratological analysis of texts: and not just of narrative text in their finished form but also in any of their preceding versions, since any textual version can be the object of narratological analysis. On the other hand, if one's aim is to study the development of a work's writing process, then narratological tools and insights can contribute to an understanding of how narratives evolve through notes, sketches, drafts, and other textual versions (such as fair copies or corrected proofs). (Bernaerts and Van Hulle 2013, 285)

More recently, Van Hulle (2022) proposed to add 'genetic narratology' to the list of postclassical narratology, which already includes intermedial, rhetorical, cognitive, feminist, queer, postcolonial, cultural, natural and unnatural narratology (Van Hulle 2022; Herman and Vervaeck 2019). He subsequently shows how genetic criticism can contribute to the aspects of classical, structuralist narrative analysis: story, narrative and narration. The story is what Russian formalists called the *fabula*, and refers to "the chronological sequence of events" (Herman and Vervaeck 2019, 48). The 'narrative' and 'narration' are both covered by what the Russian formalists called the *syuzet*, which is the specific way in which the story is presented in the text (48). According to Van Hulle, genetic narratology can be carried out by deploying a (partial) reconstruction and re-enactment of cognitive processes underlying the textual genesis. He states that, assuming that

> one accepts E.B. Holt's hypothesis that the mind is not a substance but a relation between the organism and its environment – reiterated in Richard Menary's 'Writing as Thinking' (2007) – and if manuscripts can be regarded as traces of a cognitive process, it may be possible for genetic critics to partially enact the consciousness of the writer at work [...]. (Van Hulle 2022, 148)

Of course, as Van Hulle acknowledges, "[it] may be impossible to relive the writer's 'interior' experience", but the writing process is not solely an interior affair: "It is an

interaction between the writer and their environment, including their written traces" (Van Hulle 2022, 148). The author "is aware of and responsive to his or her surroundings and acquires understanding through thought, experience, and interaction with the environment" (149). A genetic analysis, then, needs to reconstruct the writer's cognitive processes during the writing. Here, a link can be established with cognitive writing process research, which often deploys keystroke logging to investigate the cognitive processes during writing.

The written traces produced by the author are, in the digital genetic dossiers studied here, the session versions with the text produced so far and the keystroke logging data. In contrast to analogue versions of a text, the transcriptions of the keystrokes in the text allows us to exactly reconstruct the text at any given point in the writing process. As such, we may, for example, observe the state of the text when new ideas emerge, or when ideas about the text as a whole, or specific formulations, changes direction. This allows us to hypothesise about how certain formulations in the text produced so far prompted revisions or insertions of new aspects into the text, and to reflect on the "feedback loop from the text produced so far into the internal writing process and back again" (Lindgren et al. 2019, 348) with regard to the structuralist aspects of narratology. In this chapter, I examine what we can learn when narratology meets keystroke logging data. To what extent is it possible to use keystroke logging data to re-enact the author's cognition, and to trace how ideas about, and the fleshing out of, the focal points in narratological analysis arise and change during the writing process?

6.1 STORY

The first level of structuralist narratology is story, which is an abstract level and refers to the chronological sequence of the described events. Herman and Vervaeck (2019) subdivide the story into events, actants and setting. For each of these aspects, I address how keystroke logging provides insights into the integration of these aspects into the text and their development during writing.

6.1.1 Events

Bogaert: Defining the cause of the effect

Herman and Vervaeck (2019) outline several proposals for ordering events on the abstract level of the story. Roland Barthes, for example, makes the distinction between functions and indexes: "Functions are elements whose interrelatedness is responsible for the horizontal progress of events, that is, their linear development", while indexes "do not bring about the horizontal progress of events" (Herman and Vervaeck 2019, 49). On a very small scale, we find an example of a function in Bogaert's novel *Roosevelt*,

specifically in the narrative of Faraaz in the reason for Faraaz's family to leave Iran. This is the start of the fabula and the backstory in Faraaz's sections. Barthes makes a distinction between cardinal function and catalyzer functions: "A cardinal function implies a risk, which means it harbors a choice or a possibility" and the function of a catalyzer "does not involve a risk but instead merely assures the continuation of what the cardinal function has started" (Herman and Vervaeck 2019, 49–50). In general, most of the crucial events of the story are cardinal functions, as is the decision to leave Iran.

Functions are part of what Roman Jakobson terms a syntagram: "a horizontal sequence of contiguous elements" (Herman and Vervaeck 2019, 49). When there is a direct connection between elements, such as "part" and "whole", "cause" and "effect," "producer" and "product," "pole" and "opposite", then the elements are contiguous (49). The genetic material allows us to reconstruct Bogaert's decisions in terms of a small-scale syntagram: he searched for the 'cause' that produced the 'effect' of Faraaz's family fleeing from Iran.

The cause, in fact, changed over time. In the Atoma notebook, we can see that Bogaert originally planned the father as an engineer who ran into trouble with the authorities: "Met ouders uit Iran gevlucht in 1980 (zijn vader ingenieur, moeilijkheden met de autoriteiten)" (33).[184] In Session 307 (19 May 2015), Bogaert pastes 35 sentences into the Word document in the third section of Faraaz, including a paragraph about the reason for leaving Iran. In this first version of the paragraph, Faraaz's father came into trouble with the authorities while working on a road project in Tehran, causing him to lose his job. Also, according to Faraaz's mother, his father never talked about politics, though he came from a family of opponents of the regime. He was threatened and their house was burned down which was the decisive moment for fleeing the country.

This content was changed in Session 310 (21 May 2015). During this session, Bogaert altered the cause of the family leaving Iran. After this session, it is not the father whose family were opponents of the regime, but the mother. Therefore, the reason for fleeing shifted from the father to the mother.

Both a nanogenetic and a microgenetic analysis can serve to interpret the decision for this shift in the cause. Consider, firstly, the following nanogenetic narrative with a focus on the writing actions related to this change. Bogaert starts Session 310 by deleting the first sentence of the paragraph about the reason for leaving Iran: "Ik weet wat het betekent te worden uitgesloten/weggezonden, verbannen" (n1).[185] Then he deletes 'at a road project' – the specification of where the father worked as an engineer. The seventh writing operation is the deletion of the sentence: "Volgens mijn moeder sprak hij nooit over politiek, maar hij kwam uit een familie van tegenstanders van het regime"

---

[184] Translation: "Fled Iran with parents in 1980 (his father engineer, difficulties with authorities)"
[185] Translation: "I know what it means to be excluded/rejected, banished"

(n7).[186] Then Bogaert adjusts the sentence about their house being burnt down, and moves the year in which it took place, 1980, into a separate sentence (n8-19). He also makes some small modifications such as inserting a comma.

Bogaert then adjusts the next paragraph which narrates how Faraaz and his sister grew up with stories told by their mother. But Bogaert changes it to stories they told *each other*, and the mother is deleted from the paragraph. Immediately after this deletion, however, he changes his mind and undoes the revisions by clicking the undo button (n32).

Then Bogaert makes a small modification in the paragraph about leaving Iran (n33-34), after which he also makes small adjustments to the paragraph about the mother telling stories (n35-38). The next paragraph is meant to tell the reader that Faraaz's sister modified the stories. Five sentences were already written, as well as an unfinished "Ze…", which Bogaert begins to elaborate on (n39-44). This paragraph is not yet finished, but Bogaert moves back to the paragraph about the fleeing to make a couple of adjustments (n45-48). After this, he continues writing the paragraph about the sister changing her mother's stories (n49-58). And again, he returns to the paragraph on fleeing. There he re-inserts the detail that Faraaz's father came from a family of opponents of the regime, which was why he lost his job (n59-71).

After a pause of more than twenty seconds, Bogaert modifies this sentence again: now it is the mother that came from a family of opponents of the regime (n72-75). Then, he starts fiddling with a sentence, changing whether it was 'he' (the father), 'we' (the family) or 'they' (the father and the mother) who were being threatened:

Hij[87]W[81]Z[7882]W[79]e[7688]Hij[77] werden[8089]ook [48]een paar keer [83]bedreigd. (n76-83)

After adding a comma and correcting some typos in the word authorities (autoritei e[85]te[86]n, n85-86) he decides to change the subject from 'we' back to 'he': it was the father (he) who was threatened because the mother came from a family of opponents of the regime. At the end of the paragraph, Bogaert adds the sentence: "Mijn vader heeft het mijn moeder nooit vergeven" (n90).[187] Apparently, the father has never forgiven the mother that he was threatened, that their house burnt down, and that they eventually had to flee; all because of *her* family. During this session, Bogaert also added a paragraph that describes how the father did not do much after they arrived in Belgium; he sat on the couch, reading or watching the mother's movements as she cleaned or cooked.

---

[186] Translation: "According to my mother, he never talked about politics, but he came from a family of opponents of the regime."

[187] Translation: "My father has never forgiven my mother".

How can this nanogenetic narrative help us interpret the revision that transformed the father's family as opponents of the regime into that of the mother? The microgenesis shows that, during this session, Bogaert added text and made adjustments that created an opposition between the father and the mother. The mother is very active, and has a strong bond with her children, while the father is passive and blames the mother for their current situation. In this way, the father's passivity is explained by the fact that he blames the mother for having to leave Iran because of her family.

When we take the nanogenesis into account and focus on the sequence of the revisions we may hypothesise even further. First of all, we could see that Bogaert repeatedly returned to the paragraph, indicating that he was not yet satisfied with the paragraph in that moment. Even while working on other paragraphs, he was constantly in the process of revising the paragraph about leaving Iran.

Secondly, the order of the revisions further shows that Bogaert initially decided to delete most of the material about the mother from the chapter, deleting the sentence starting with 'according to my mother' and removing her from the paragraph about the stories. But just after this deletion of the mother from the paragraph about the stories, he changed his mind and reversed the revision. Not long after that, Bogaert decided to give her a more prominent place in the paragraph about the reason for fleeing Iran. These revisions may therefore be related to each other. The keystroke logging data show that Bogaert focussed on the role of the mother in the chapter, and specifically in the paragraph about the stories. These decisions may have led him to address the mother's presence in the paragraph that precedes it. In other words, working on the role of the mother may have activated other units related to the mother. So, by focussing on the nanogenesis, we can hypothesise about the decision-making process that led to this revision and see how other changes in the text produced so far prompted an emphasis on the mother in the text.

Van Pelt: The genesis of a pure index

Barthes also distinguishes between two types of indexes. The first are pure indexes that require interpretation by the reader. This includes, for example, information about clothing and other preferences that are to be interpreted by the reader as symbols for the character (Herman and Vervaeck 2019, 50). The second type are informative indexes, which are "mainly important for spatiotemporal description and which [do] not require symbolic interpretation or the solution of a mystery" (Herman and Vervaeck 2019, 50). However, an informative index may turn out to be a pure index after all.

An example of a pure index that acquires more importance with a genetic narratological analysis – and during the genesis itself – can be found in the genetic material of Van Pelt's story "Dauphin". Leisure time, during which tourists can relax in the sun and swim, is part of the Flamingo Island excursion. The I-narrator spends part

of this time trying to read the book *Heilbot op de maan* (*Halibut on the Moon*). She could have been reading any book, but the genetic material reveals that this book was in fact chosen for a specific reason. It could be interpreted by the reader as telling them something about the death of the narrator's partner, as the genesis will indicate. The genetic material also shows that Van Pelt revised this passage to emphasise the significance of this book.

Van Pelt starts writing this passage in Session 8 (9 December 2020) with the following sentence: "Terwijl Finn keer op keer van de blauwe glijbaan naar beneden in het water plenste, probeerde ik te lezen" (after n95).[188] Then she conducts a Google search: "roman zoon over zelfmoord vader vissen" (n96).[189] The results given by Google lead Van Pelt to a review of a book by David Vann in the Dutch newspaper *Trouw*: "Met zijn nieuwe boek rekent David Vann af met het schuldgevoel over de zelfmoord van zijn vader".[190] The name of the book appears to be *Halibut on the Moon*. She googles the author David Vann and visits his Wikipedia page (n98-99). Then she makes a new Google search for 'novels about suicide father': "roman over zelfmoord vader" (n101). This search is less specific – it does not mention fishing or fishes – and therefore gives more general results, such as a list of nine books about suicide ("9 boeken over zelfdoding") on the website Boekvinder.be (n102). Van Pelt visits this website, but when she returns to the Word document, she writes the title of the book by David Vann: "Heilbot op de maan" (n103).

The exogenesis thus shows that Van Pelt was intentionally looking for a book about suicidal fathers, and the specific terms of the initial search indicate that she had already considered this book but may have forgotten its name. The search terms also emphasise that the reader could interpret the inclusion of this book as providing information about the backstory: did her partner also commit suicide? Or does it refer to her partner's death in another way? The book thus establishes a link between an untold event from the backstory (the death of her partner) and the fictive present (reading *Halibut on the Moon*).

At the end of this session, a passage describes the narrator trying to read the book but having difficulty concentrating:

> ~~Ik~~ [225224][227226]Op mijn bui~~tk~~[229228]k in het zand probeer ik [230]~~TeFinn keer op keer van de blauwe glijbaan naar beneden in het water plenste, probeerde ik~~ [231]te lezen. *Heilbot op de maan*. Telkens ik een bladzijde omsl~~oeg, wist~~[232]a, weet[233] ik dat de woorden ~~die ik had gelezen,~~[234] alweer vervlogen ~~ware~~[235]

---

[188] Translation: "While Finn splashed down the blue slide into the water again and again, I tried to read."

[189] Translation: "novel son on suicide father fishing"

[190] Translation: "With his new book, David Vann settles the guilt over his father's suicide"

zij[236]n. Mijn ogen registree~~rde~~[237]re[238]n, maar ik n~~a~~[239]ee[240]m niets in me op[184]
(Session 8)

In this version, the book appears at first glance to be nothing more than an informative index, a detail describing the setting. In the course of the writing process, however, Van Pelt interacts with the text produced so far and makes revisions to indicate the importance of this book to the narrator. Between Sessions 8 and 17, the passage was slightly altered at the word level and Van Pelt also added a new sentence: "De hitte dreunt in mijn hoofd" (Session 16).[191] In Session 17 (31 December 2020), she reads the passage again, and this sentence about the heat rumbling in the narrator's head seems to act as the impetus for a new sentence, one that also seems to refer to the book's effect on the narrator: "Het verdriet dreunt in" (n133-135).[192] This sentence suggests that the book instils sadness in her, but Van Pelt deletes the sentence before it is finished (n136), perhaps because it is too explicit a reference to grief – a decision Van Pelt often makes during the writing process.

Three sessions later (Session 20; 10 January 2021), the passage is modified again. Van Pelt changes the start of the sentence about the words fading away (n37-40) by adding a clause:

Ik dacht da~~t het b~~[38][37]t het boek me midscheeps zou raken maar t[39]~~T~~[40]elkens ik een bladzijde omsla, weet ik dat de woorden alweer vervlogen zijn. (Session 20)

The insertion of the clause has two functions. Firstly, it can be classed among the revisions that emphasise the nautical theme. Throughout the writing process, she incorporates maritime vocabulary and idioms into the text. Although "midscheeps" ["midship"] was not in the list of expressions she had inserted in Session 8 (see Chapter 5.2), it fits with the theme. Secondly, it is also clear in this clause that the narrator thought the book would touch her deeply, which might indicate some correspondence with her own personal experience.

This revision thus makes the book, as a pure index, more prominent in the text. While a narratological analysis in itself can already help us understand many of the details in the text (Herman and Vervaeck 2019, 52), a (nano)genetic narratological analysis makes such an analysis even more profound, since it addresses the steps that were taken to implement or disregard such details.

---

[191] Translation: "The heat rumbles in my head"
[192] Translation: "The grief rumbles in".

6.1.2 Actants

Van Pelt: The origin of the helper

Events cannot be considered separately from the agents who are involved in them (Herman and Vervaeck 2019, 54). Following Greimas, these agents can be termed *actants*, which "does not refer to the actual manifestation of a character in the text but rather to the specific role a character plays as an abstract agent in a network of roles on the level of the story" (Herman and Vervaeck 2019, 54). Greimas' actantial model consists of six roles or actants. The subject carries out the action and strives for a specific object. This quest is inspired and triggered by the *destinateur*, or following Cor van der Voort, the sender. The *destinataire*, translated by Van der Voort as receiver, is the agent who benefits from the quest. During the quest, the subject can be assisted by the helper, or can be hindered by the opponent (Herman and Vervaeck 2019, 55). The roles can be played by real characters and by abstract actants (e.g. fate and death), one role can be played by many characters, and one character can play several or all the roles (56). The genetic material of "Dauphin" shows how such roles gradually develop during the writing process.

In "Dauphin", the I-narrator is the subject who, in this case, strives to survive her first summer holiday alone with her son. They travel to Djerba in Tunisia and take a day trip on a pirate ship to Flamingo Island. The narrator's father can be seen as the sender: he gifted her the vacation to Djerba. The role of the opponent may be the death of her partner, which forces her to take on all the 'parental' tasks and makes her the only person her son can lean on; she must simultaneously embody a 'father' and 'mother'. The I-narrator can also be seen as an opponent of herself, since she feels inadequate to fulfil both roles. Her son Finn is the agent who benefits from the quest. The role of the helper is played by the character Mustafa, who rather unexpectedly takes care of Finn during the excursion.

In the first writing session (23 November 2020), Van Pelt already makes a reference to the man that will eventually be Mustafa. The I-narrator sees a man sitting in the corner of a bench, in the sparse shade of the ship. He is wearing dark clothes and sunglasses, and seems to be staring sullenly ahead:

> [153]~~De~~ [154][155]~~Recht tegenover me zit een man in e~~[156] [157]~~er me zit een man.~~[158]
> [159]~~Op de ban~~[160] [161]~~R~~[162] [163]~~Se~~[164] [165]In de hoek van de zitband~~,~~[166] [167]k, in
> ee[228][227]de weinie~~[230][229]ge schaduw die ~~op~~[232][231]er op ~~td~~[234][233] dit schip is, [235] zit
> een man. [168]~~Zijn lange trainin~~[169] [170]~~trainingsborek~~[171] [172]~~roek en swetar~~[173] [174]~~ater
> met lange~~[175] [176]Met zijn ~~a~~[180][179] ~~zwarte~~[181]~~trainings~~[182] Ad~~disa
> s~~[184]~~is~~[185][186][183]~~sa~~[188][187]das-trainings[189]broek en ~~zw~~[177]~~gitzwar~~[190][191] zwarte [192]
> [178]sweater[193]~~,~~[194][195] v~~lat h~~[196][197]alt hi~~op t~~[198][199]j tussen al het blote bla~~k~~[200][201]nke
> ~~ble~~[202] [203]~~vee~~[204][205]lees. [206]~~Hij draagt een zonnebril,~~[207][208]j draagt een zonnebril
> en lijkt ~~dars~~[209] [210]nors voor zich uit te staren. [236]~~In~~[237][238]~~Hi~~[239] [240]Oortjes in.

²¹¹H̶i̶j̶²¹² ²¹³Is hij e̶²¹⁴ ²¹⁵iemand van d̶ ̶e̶ ²¹⁶ ²¹⁷ de crew? ²¹⁸e̶e̶ ̶p̶²¹⁹ ²²⁰e̶e̶²²¹ ²²²Een piraat m̶e̶t̶ ̶e̶e̶n̶ ̶d̶a̶g̶j̶e̶ ̶v̶e̶r̶l̶o̶f̶²²³ ²²⁴ die̶ ̶z̶i̶j̶²²⁵ ²²⁶zijn kostuum vergat? (Session 1)

Before Van Pelt wrote this passage describing the man with the dark clothing, she visited a couple of TripAdvisor pages and travel agency websites that describe similar excursions to Flamingo Island. The TripAdvisor pages that she visits also recommend another TripAdvisor page for an excursion to Flamingo Island called 'Bateau Pirate Elyssa Djerba'.[193] This specific TripAdvisor page contains various photographs of the excursion. One of these photographs depicts a man who could well have served as the inspiration for the character Mustafa. The photograph gives an overview shot of the deck and in the only shade on the ship – in the shadow of the mast – a man is sitting on a bench, wearing dark clothing and sunglasses (See Figure 6.1). Seeing this photograph may well have activated Van Pelt's imagination about what this man is doing on the pirate ship.

The question is, however, if Van Pelt actually visited this TripAdvisor page and saw this photograph during the writing session. The keystroke logging data does not provide direct evidence; it does not list this exact website. It is noteworthy that she includes the name of the ship – Elyssa – in the draft towards the end of the session: "De onze heet Elysa" (n242-244).[194] Yet, this does not mean that she has visited the specific TripAdvisor page of 'Bateau Pirate Elyssa Djerba', since the TripAdvisor page she *did* visit also contains a picture of a pirate ship called Elyssa. This may have inspired her to call the ship in the story 'Elyssa'. We can therefore conclude that she has probably not seen the TripAdvisor page 'Bateau Pirate Elyssa Djerba' and certainly not the specific photograph. The description of the man sitting in the shade is rather a notable coincidence. Indeed, Van Pelt later clarified that she had undertaken this excursion herself, and that she saw a man who fascinated her.

This illustrates one precaution one must take when analysing keystroke logging data. The level of detail makes it seem as if we know exactly how the writing process proceeded, but we can only infer the development based on what has actually been logged. With keystroke logging, we cannot trace the websites the author visited before or after the writing sessions, which also implies that we cannot determine exactly what was or was not seen. In addition, searching for websites visited by the author using the page titles given in the keystroke logging data leads to a plethora of websites where the boundary between what was visited and what was not is easily crossed. And since there

---

[193] See: https://www.tripadvisor.be/Attraction_Review-g1187636-d16852897-Reviews-Bateau_Pirate_Elyssa_Djerba-Houmt_Souk_Djerba_Island_Medenine_Governorate.html#photos;aggregationId=101&albumid=101&filter=7&ff=428467644

[194] Translation: "Ours is called Elys". The spelling of the name of the ship varied during the writing process. At first, Van Pelt wrote it with one 's', later she changed it to the way the ship is called on the websites, Elyssa, with double 's'.

is some time between the writer's search and the reconstructed search, pages can be removed or updated. The page given in the keystroke logging data as "Tripadvisor | Piratenschip excursie naar Flamands Island aangeboden door Depart Travel Services | Djerba, Gouvernement Médenine", for example, no longer exists online today.

Based on her own experience or not, Van Pelt incorporates a man that – through his clothing – stands out among the other tourists, as she writes from the narrator's perspective: "Is hij iemand van de crew?"[195] and "Een piraat met een dagje verlof",[196] which she immediately changes in "Een piraat die zijn kostuum vergat?".[197] Thus, Mustafa is included in the story, but his role has yet to be defined, after this first writing session. He is (just) a grumpy-looking man who raises questions in the narrator's mind. The question here is: when did Mustafa's role become more concrete?

Figure 6.1: The photograph on TripAdvisor that Van Pelt may have used as inspiration for the character Mustafa.

---

[195] Translation: "Is he someone from the crew?"
[196] Translation: "A pirate on a day's leave"
[197] Translation: "A pirate who forgot his costume?"

Until the fourth writing session, Mustafa's role is not elaborated. Van Pelt only makes small revisions in the description of his appearance – should he wear a hat or not? Then, in the fourth session (30 November 2020), Van Pelt starts working on a passage describing that the narrator and her son have to walk to a wooden hut on the beach. Not long after embarking their walk, Finn refuses to go any further. This is the moment Van Pelt lets Mustafa return to the stage and the reconstruction of the keystroke logging data offers us a glimpse of the role Van Pelt has in mind for this character. When Van Pelt introduces Mustafa back on the scene, she makes a couple of immediate (pre-contextual) deletions, one of which seems to reveal her thoughts on the role Mustafa should fulfil:

> [280]~~'C~~[281][282]~~Als een prins op het~~ [283] [284]~~Deman met de~~ [285][286]~~norse~~[287] [288]~~m~~[289] [290]an staat de man met de trainingsbroek naast ons. [291]Zijn ~~e~~[292] [293]~~zw~~[294] [295]swet~~er~~[296] [297]ater heeft hi~~om~~[298] [299]~~jm zij~~[300][301]om zijn middel~~t~~[302] [303]gebonden~~,~~[304] [305].
> (Session 4)

She starts writing a sentence with "Als een prins op het",[198] but before she finishes this sentence, she deletes it. Yet, it is clear what she meant to write: like a prince on a white horse. He is the hero who will save her from distress. In terms of the model of the writing process (Hayes 2012; Leijten et al. 2014), we might hypothesise that the proposer came up with the idea of the saviour, which activated the connection of the fairy-tale, and led the translator to transform this into language by addressing the prince on the white horse. While the transcriber was working on creating the written output, the evaluator quickly judged the formulation as too explicit and perhaps too clichéd, after which the sentence in production was deleted. The evaluator did in fact agree with the general idea of the 'saviour', and the translator translates it into more implicit language.

Van Pelt elaborates the scene as follows. The I-narrator tells Mustafa that her son is tired – realising too late how desperate it sounds. Mustafa crouches down in front of Finn, who immediately understands that he is allowed sit on his back. The I-narrator concisely describes the situation as follows: "Hij kent hem niet, ik ken deze man niet en mijn zoontje hangt nu op zijn rug" (n338).[199] Although Mustafa is described as an enigmatic figure, he takes care of Finn and, as such, helps the I-narrator.

The writing of the sentence in which Mustafa appears on the beach, also shows that Van Pelt considered calling Mustafa surly again. Van Pelt repeatedly refers to the alleged sullenness of Mustafa, which we could interpret as part of the contrast that Van Pelt implements in the text between how the I-narrator perceives Mustafa and his actions. As Herman and Vervaeck point out, "[t]he reader will expect certain actions

---

[198] Translation: "Like a prince on the"

[199] Translation: "He does not know him, I do not know this man and my son is now hanging on his back"

from a specific actant'', which often rely on the stereotypes that circulate in the reader's social and cultural context (2019, 48). The writer can then play with the reader's anticipations to create suspense or surprising turns (58). Van Pelt plays with the reader's anticipations through the I-narrator's descriptions of Mustafa as sullen and a bit antisocial.

Nonetheless, the role that Mustafa eventually fulfils is rather surprising, and his role as helper gradually becomes more elaborate during the writing process. Van Pelt works in Session 8, inter alia, on a passage about the act performed by the pirates. Just before that, she had included notes about how the story should proceed and end:

> Mustafa weer?
>
> Terugtocht - dolfijnen
>
> Einde?? -Ander begin? (Session 8, n253-256)

In this note, Van Pelt also wonders whether Mustafa should reappear in the story. She had thus already given Mustafa the role of the helper, but the note indicates that she had not yet decided if and how this role should be further developed in the story. In the next session she works on this paragraph about the pirates' performance (Session 12; 11 December 2020), she introduces Mustafa back on the scene. The I-narrator describes Finn resting his head against Mustafa:

> [164]H[165][166]Finns hoofdje maakt zich los van mij~n~ [167] [168].[172]n arm~,~ [174][173].[175][176]~To mij~[177] [178]~ot mijn verbazing~[179] [180]Mustafa zit nu naast hem en Finn~s h~[181] [182] leunt t~gen~[183] [184]egen hem aan. (Session 12).

These sentences were not written in one go. In fact, we can see a pre-contextual deletion (n179) which perhaps addresses the effect that the sentence she is writing can have on the reader: surprise. Seeing her son resting against Mustafa, the I-narrator contemplates whether she can take Mustafa home with her so that Finn does not have to lean on her alone: "Zou ik Mustafa mee naar huis kunnen nemen? Zodat Finn ook thuis niet enkel tegen mij aanleunt?" (Session 12, n186-195).[200] After writing this scene, Van Pelt removes the note "Mustafa weer?" from the document (n196).

The story describes more than once how Finn hides in his mother's armpits. Van Pelt revised one of these descriptions before writing the new paragraph where Finn leans against Mustafa. Could the attention she paid to hiding under an armpit during revision have led to the idea of Finn leaning against Mustafa? The text produced so far, can, after all, also "prompt content generation" (Wengelin, Leijten, and Van Waes 2010,

---

[200] Translation: "Could I take Mustafa home with me? So that Finn is not solely leaning on me at home too?"

737). About this, too, we can merely formulate hypotheses. Even though keystroke logging allows us to reconstruct the writing process on an unprecedented level of granularity, we are, of course, only able to analyse what has been externalised; we cannot know – only infer – what the author thought.

The role of Mustafa was already implemented early in the writing process – in Session 4 – but his role became more elaborate as the writing process continued. The idea for Mustafa may have originated from seeing a photograph on TripAdvisor, and while this is *not* the case, it shows that this fictive situation is not that far from the reality. Van Pelt actually based her story on her own experiences during a holiday in Djerba. The immediate revisions Van Pelt made during writing about Mustafa reveal her thinking process; she is more explicit about her ideas – or rather, the I-narrator is more direct during this process. In the first example, the I-narrator explicitly declares the role of Mustafa; the prince on a white horse that will save her, or more actually, Finn, from this situation on the beach. In the second example, the I-narrator at first cannot hide her surprise to see her son resting his head against Mustafa. The nanogenesis thus provides information about how the roles of actants gradually crystallise during the writing process.

Posthuma: Self-acceptance through others

In Posthuma's "Daarom haten ze zichzelf" one of the themes is 'self-hatred and self-acceptance'. Regarding this theme, several actants can be distinguished, for example in the narrator's quest for body positivity. In the roles distinguished by Greimas, the first-person narrator, the subject, strives for self-acceptance in terms of her appearance, although she finds it difficult to achieve. The acceptance of her body as it is, is therefore the object. The sender is also the I-narrator, as she – theoretically – wants to accept her 'pear-shaped' body. The receiver is the I-narrator as well because she will benefit from accepting herself, including her body. In her quest for self-acceptance, her husband helps her. He often tells her that he thinks she is beautiful and that she over-interprets other people's opinions. She is also, to a lesser extent, her own helper. For example, she teaches her son not to judge women by their size, which in the long run can help her and others to accept themselves. In addition, she reprimands her friend for self-deprecating remarks. At the same time, she is her own worst opponent as she finds it difficult to take her own advice. However, the first-person narrator originally had a much stronger role as a helper in achieving the goal of self-acceptance during the development of the story.

In the published text, there are two paragraphs in which body positivity and self-acceptance are addressed. Until Session 37 the content of these two paragraphs was still contained in a single paragraph, written in Session 14 (17 September 2020). In the first version of this paragraph, the I-narrator tells herself in the imperative mood that it is

sad to go along with the beauty ideal, that she has to love herself (which Posthuma later changed to that she *must be* herself), that she is a woman, that she must not be like her friend, and that she should not be guided by the thoughts of others, who may well think something different. This is also what her husband also tells her.

> Mijn vriendin is iets slanker dan ik en heeft niet, zoals ik, de vorm van een peer. Misschien had ik geen derde glas wijn moeten nemen. Dikke billen zijn mooi, houd ik mezelf voor. [67] [73]~~Dat wil ik vinden~~ [74] [75]Het is treurig om ~~je~~[76] [77]mee te gaan in het ... schoonheidsideaal.[78] ~~Hou van~~[114] W~~wes~~[116][115]ees[117] jezelf. [79]~~Probeer niet~~ [80] [81]Streef niet het lichaam van een tienermeisj~~e na, of nog e~~[82][83]e, of tienerjongen, na. [84]Je bent een vrouw. [100]Wees niet zoals je vriendin, die er stiekem van geniet dat ~~jij dikker~~[101] [102]zij slanker is dan jij, die heeft besloten dat zij de mooiste is van jullie twee, die jou vaak [123]stijlvol no~~i~~[120]em~~de~~[121]t[122], maar nooit mooi, [119]ga n[118]iet mee in de gedachten van anderen, bovendien, misschien denken ze ~~dit wel niet~~[104][103] iets heel anders, ~~zegt je man niet~~[106][105]dat zegt je man ook, die jou mooi vindt, [125]hij zegt [128]dat het je eigen gedachten zijn, dat [108]je mensen te veel invult[107], ~~vul~~[110][109] ~~jij~~[129]vul~~t~~[130] me niet [131]in, ~~zegt hij als ze~~[112][111] roept hij als ~~ze~~[126] jullie[127] ruzie hebben[113]. (Session 14)

In Session 19 (24 September 2020), Posthuma adds one rather important sentence to this paragraph: "Voed jezelf" (n52).[201] So when this section first appears in the story, the narrator has an important role to play in helping herself achieve self-acceptance, as she tells herself to 'be herself' and that it is important to eat – a thought that is echoed by her husband. It is also noteworthy that the beauty ideal is explicitly mentioned, which in this version also functions as an opponent.

In Session 24 (1 October 2020), Posthuma reduces the narrator's role as her own helper in achieving self-acceptance. Posthuma deletes all the imperatives that positively affirm the narrator's appearance. She also adds new sentences about teaching the son that heavy-weighted people can be beautiful. First, Posthuma writes that the I-narrator does not want him to hate himself (n341), which implies that she is teaching him this for his own benefit. But Posthuma almost immediately deletes this (n348). Now this sentence can only suggest that the I-narrator is teaching her son this lesson for the sake of women, and therefore for her own sake as well. Yet, Posthuma also adds that the son sees his mother inspecting her bottom in the mirror, which, in a way, nullifies what she teaches him (n342-344). The I-narrator tells people she has a healthy weight, but the context in which this sentence appears casts doubt on whether she really believes this. Moreover, whereas initially it was the *narrator* who said that she should not conform to other people's opinions – a belief confirmed by her husband – now it is the *husband* who makes this statement; the first-person narrator is merely recalls his opinion.

---

[201] Translation: "Nourish yourself"

Mijn vriendin heeft niet, zoals ik, de vorm van een peer. Misschien had ik geen derde glas wijn moeten nemen. Dik is ~~ook[335]~~ ~~net zo mooi of lelijk als[336][337]~~ kan mooi en lelijk zijn, net als dun[338] ~~mooi, houd ik mezelf voor[339]. Dat[346],[347]~~ leer ik ~~ook[340]~~mijn zoon. Hij mag geen man worden die dikke vrouwen afwijst. [342]Maar hij zi~~tbo[343][344]~~et ook hoe ik mijn achterkant in de spiegel bekijk~~, hoe ik frons en dan[345]~~de frons in mijn voorhoofd[349]. [341]~~Hij mag zichzelf niet gaan haten[348]~~ [350]Ik heb een gezond gewicht, hoor ik mezelf vaak tegen mensen[353] zeggen. [351]Daar gaat het om. [352]En[354] ~~ik[372]~~ ~~vind het[355]~~ ~~word[356]~~~~treurig[371]~~als ik[357]onze vriendin [358]hoor praten over de periode dat ze ~~depressief[359]~~ van de stress zo weinig at, hoe mager ze was en hoe goed dat voelde[360]~~was, hoe mager ze was en hoe goed [361]over.[362]HOE IK TREURIG WORD OM BIJ VROUWEN TE ZIEN HOE ZE EEN ONBEREIKBAAR SCHOONHEIDSIDEAAL NAJAGEN MAAR ZELF OOK JALOERS BEN OF ALS IK ZIEK BEN DENK: DAAR VAL IK VAN AF[363]~~, word ik [373] treurig ~~omdat[376]~~ [370]~~het niet gezond is om jezelf niet te voeden en omdat ik jaloers[374]~~~~maar dat komt ook[375]~~, om~~dat het[381][378]~~dat verwoestende en [389]on...[385]uitroeibare[386] schoonheids[383]ideaal [382]~~maar ook[379][377]~~doordat[380]en omdat[384] ik jaloers~~op[365][364]~~ben[366]. [367]~~Ik ga van[368][369]~~Als ik gestresst ben eet ik graag patat..~~Het is treurig om te lijden onder een onbereikbaar schoonheidsideaal. Wees jezelf. Voed jezelf. Je bent een vrouw. Streef niet het lichaam van een tienerjongen na.[387] Wees niet zoals je vriendin, die er[388]~~Ik denk dat~~mi[391][390]~~onze vriendin er[392] stiekem van geniet dat zi~~j[393]~~e[394] slanker is dan ~~jij[395]~~ik. [396]Ze noemt me vaak stijlvol,[397]~~, die jou vaak stijlvol noemt,[398]~~ maar nooit mooi. ~~Ga[399]~~Verlies jezelf niet zo in de[400] ~~niet mee in de[401]~~ gedachtengang van anderen, zegt mijn man altijd. [402]~~bovendien, want m[403]~~M[404]isschien denken ze iets heel anders dan je vermoedt~~, dat zegt ook je man, die jou trouwens wel mooi vindt[405]~~. ~~Hij zegt dat het venijn niet in het hoofd van anderen zit, maar in jezelf, dat je mensen te veel invult.[406]~~ Het venijn zit in je ei~~tgen[408][407]~~gen hoofd.[409] Vul me niet in, roept hij als ~~jullie[410]~~ we[411] ruzie hebben. (Session 24)

Posthuma revises the paragraph multiple times in eleven sessions. Sometimes she only changes a word and other times the revisions have more effect. For example, in Session 28 (6 October 2020), she changes the narrator saying she has a healthy weight to saying she is just healthy: "Ik ~~heb een[5]~~ben [5]gezond ~~gewicht[6]~~, hoor ik mezelf vaak tegen mensen zeggen" (Session 28). Then, this sentence is deleted in Session 34. These revisions further reduce the self-accepting statements of the narrator.

This example illustrates how the roles of the actants evolve and change during the writing process. A genetic reading of this paragraph instantiates a more cynical reading of the story, which seems to imply that no matter how often you positively acknowledge yourself, self-acceptance is always conditioned by others. This is exactly what becomes more explicit after the deletion of the positive, self-accepting statements the narrator tells herself in Session 24. In Session 25 (2 October 2020), Posthuma added a paragraph in which the narrator admits that lately she finds older lesbians attractive. She adds that

the narrator tells her friend that she has always experienced herself through the eyes of a man, and now that she no longer conforms to this norm, lesbians have suddenly become attractive to her.

> [174]Dat ik mezelf altijd [175] [176]ijn seksualiteit altijd indirect heb beleegd, [177][178]fd, via de ogen van[179] [180]door de ogen van een man naar mezelf heb gekeken, en nu ik ouder ben en volgens de norm niet meer aantrekkelijk ben ine[181][182]plotseling mijn eigen seksualiteit ontdek. (Session 25)

Posthuma deletes this in the same session. The idea of 'looking at yourself through the eyes of others' returns in a different form in the following session: "Want er moet altijd iemand zijn door wiens ogen ze naar zichzelf kunnen kijken" (Session 26; n240).[202] This sentence is then deleted yet again in Session 27 (6 October 2020), but the idea reappears in a sub-clause – "die alleen begeerte kan voelen als een ander haar begeerlijk vindt" – within the passage about the attractiveness of lesbian women that Posthuma also re-inserted:

> Een heteroseksuele vrouw van boven de veertig die zichzelf niet meer aantrekkelijk vindt omdat ze buiten [81][80]mannen van [83][82]liever naar jongere vrouwen kijken, die alleen begeerte kan voelen als een ander haar begeerlijk vindt[104]en [105][106], [107]die daarom haar aandacht verlegt naar vrouwen die eruit zien als een man.[89][84].[91],[92] die [103]pas [95] begeerte [88] een libido heeft [94] alleen begeerte kan voelen als een ander haar begeerlijk vindt[102]. [96](Session 27)

The shift in the role of the helpers to achieve self-acceptance, from primarily the self to others, gave Posthuma the opportunity to explicitly address the issue of 'judging yourself through the eyes of others' in the story. Revisions in the text produced so far enable exploration of new territory and can therefore steer the text in a new direction. Thus, this shows that a shift in the role an actant has in achieving a specific goal can substantially influence the further development of the text.

6.1.3 Setting

Bogaert: Narrating setting

The specific time and place also belong to the level of the story, since events take place in a setting.[203] Mikhail Bakhtin calls this spatiotemporal setting the 'chronotope', which

---

[202] Translation: "For there must always be someone through whose eyes they can look at themselves"

[203] This section is based on a section published in: L. Bekius, "De geboorte van een plein. De analoge en digitale tekstgenese van het *incipit* in Gie Bogaerts roman Roosevelt", *Tijdschrift voor Nederlandse taal- en letterkunde* 137.2 (2021).

– according to him – "constitutes the narrative and ideological center of the text because it shapes figures and actions" (Herman and Vervaeck 2019, 60). The setting can both function as an index for the action and for the actants. In *Roosevelt*, the setting functions, literally, as the 'common ground' for all the characters in the novel; no matter how different their lives may be, they share the experience of being present on the Franklin Roosevelt square in Antwerp on that warm day in June. They act within the same chronotope. Although they notice each other *en passant*, they are brought together by the setting in yet another way: through the narrator. Franklin, the square, is both the setting (in the fictive present) and the main extradiegetic narrator in *Roosevelt*.

How do you convey, in a natural way, that a *square* is the narrator? This question is especially relevant for the first chapter, which can be seen as the *incipit* of the novel (See, for a more elaborate discussion of the genesis of the incipit Bekius 2021b). There are several functions an incipit can have. According to Niels Buch Leander, the incipit on the one hand prepares the reader for what is to follow and provides the information for the unfolding of the story, and on the other hand seduces the reader to enter the story world and to keep on reading (2018, 7). These two functions are also noted by Andrea Del Lungo, who distinguishes a seducing function of the incipit among four other functions:

1. The initiation of the text (codification function/*fonction codifiante*)
2. Presenting the subject of the text (thematic function/*fonction thématique*)
3. Staging the story (informative function/*fonction informative*)
4. Initiating the action (dramatic function/*fonction dramatique*) (Del Lungo 2003, 154–55)

The parts of the incipit that explicitly address the chronotope of the novel have an informative function, which is part of the exposition: "exposition concerns setting (time, place) and characterization relevant to the fictive present [...] [and] can be roughly likened to the frame in which the narrative unfolds" (Hentea 2010, 348).

The incipit contains constitutive information – information about the fictional world, about time, space, and characters (Del Lungo 2003, 167). This is also the case in *Roosevelt*, in which the square, as narrator, explicitly introduces himself to the reader as Roosevelt, but 'you' may also address him as Franklin. This adds to the performative aspect of the beginning: it asks the reader to accept the narrator's name.

> Wat aarzel je? Schrikt mijn oudemannenhuid je af? De schrammen en die paar putten in mijn vel? Kijk niet te nauw. Ik heb een gouden hart.
> Mijn naam is Roosevelt. Maar dat weet je natuurlijk, anders was je hier niet. Je mag me ook Franklin noemen als dat gemakkelijker is.
> Honderdvijftig jaar maken er iemand niet beter op zoals je ziet. Ik kan nauwelijks nog bewegen. En ik heb last van jicht. Zo nu en dan spelen ook mijn ingewanden op. Soms is het alsof er een tram doorheen rijdt.
> Ik zou best een nieuw pak kunnen gebruiken. Moderner. Modieuzer. Meer

> gestroomlijnd. Ik had het al jaren geleden moeten krijgen. Maar beloftes aan een oude man worden zelden ingelost. (Bogaert 2016, 9)[204]

Through the context of the sentence in which Franklin introduces himself, the referential information and the proper names, the reader discovers that the narrator coincides with the location. Bogaert chose to make the reader aware of Franklin's 'square-ness' through a number of descriptions of Franklin, such as the scratches and pits in the narrator's skin and the feeling that a tram is passing through his stomach. These are already present in the first draft that Bogaert pasted in the Word document in Session 12 (6 July 2013). In terms of their content, these sentences are not changed throughout the writing process; Bogaert only performs several revisions that focus on the formulations.

In the incipit, an appeal is made to the reader's spatial, or even sensory, imagination. The reader is asked to imagine the textual world; attention is drawn, among other things, to the smell on the square and the sun that shines on the buildings:

> Ruik je de ochtend? Proef je de dag? Heb je gezien dat achter de kantoren de zon allang is opgestaan? Een rechthoek van haar gloed ligt tegen de zijgevel van de Opera. Merk je hoe het roze licht daar over het terras van Brasserie Gustav valt? (Bogaert 2016, 9)[205]

As such, the incipit also has the function of providing referential information – information that refers to the real world and in which the staging takes place – by referring to the extralinguistic, to general or specific knowledge passed on to the reader (Del Lungo 2003, 167). By means of descriptions of recognisable elements of Roosevelt square (the Opera and Café Gustav), Bogaert refers to the real world and thus situates the story on the square.

To account for such realistic descriptions of the square, Bogaert went to the actual Franklin Roosevelt square in Antwerp to observe and make notes of everything that

---

[204] Translation: "What are you hesitating about? Does my old man skin put you off? The scratches and those few pits in my skin? Don't look too closely. I have a golden heart.
My name is Roosevelt. But of course you know that or else you wouldn't be here. You can also call me Franklin if that's easier.
One hundred and fifty years don't make anyone any better as you can see. I can barely move. And I suffer from gout. Every now and then my intestines also play up. Sometimes it's like a tram is passing through them.
I could do with a new suit. More modern. More fashionable. More streamlined. I should have got it years ago. But promises to an old man are rarely kept."

[205] Translation: "Do you smell the morning? Taste the day? Have you noticed that behind the offices, the sun has long since risen? A rectangle of its glow lies against the Opera House's side wall. Notice how the pink light falls over the terrace of Brasserie Gustav there?"

caught his eye. For the timeframe of the first chapter (6:30am) Bogaert observed the square during the morning of 7 June 2013. He noted, among other things:

> 6.30 Het is al helemaal xxxlicht – > Is ook om 5.30 al licht! Ook de zon is er al. Nog niet zichtbaar door de hoogbouw, maar je ziet de gloed al <tegen de zijgevel van de Opera en het terras van de Gustav, maar boven de bomen ook vol tegen de gevel v.d. Theaterbuilding>; merkt zelfs al hoe warm de dag zal worden.
>
> Er is al wat verkeer. De eerste trams passeren. Er staan al wat bussen te draaien; nog weinig auto's; zeven taxi's staan op rij te wachten. Wel al wat voetgangers en enkele wachtende mensen op perrons. Ook al behoorlijk wat volk op de trams. Een bus arriveert uit Turnhout, bijna vol (?) De Oranje poetswagen v.d. stad komt de goten schoonmaken. Een dikke, slonzige vrouw loopt met een borstel vooraan. De zon komt <6.50 u> boven het dak van 'Geuzenhof' café tussen twee hogere gevels kijken en giet geel licht over het plein.[206]

In the seventeenth session (23 August 2013) Bogaert adds these impressions to the Word document. In processing the impressions, he transforms them into Franklin's observations in which he addresses Carla and the reader: "Heb je gemerkt dat […]"[207] or "Kijk hoe die […]".[208] This inserts the referential information. Because the descriptions of the square are based on observations, the material shows that Bogaert wanted to sketch a realistic image of the square at half past six on a warm summer's day.

> [270]Ruik je de ochtend?+++[271] [272]Hoor je de dag beginnen [331][396]aanbre[273]ken[332]? [210]Heb je gemerkt dat de zon er [211][226] al [251]al [256]achter de hoogbouw moet zijn [231]is [234]geklommen? [227]is?[228] [225]Je kan haar [229] Kijk hoe haar [230]gloed al[257]al[274][333][232] zien tegen de zij[258]gevel van de Opera. [235] [233]En [369] K[370]k[371][252] K[253]ijk[237] hoe[238] die[276] hij[277] al [275][249] die [250][236] [239]over hett[240][241]t[242] g[243]erras van de[244] Gustav valt en [245]boven de bomen ook tegen de hoogste kantoren van de 'Theaterbuild[246]ing[254][255]. er valt[247] 5 *6DECORbeschrijving. [212]

---

[206] Translation: "6.30 It's already xxx light -> Is also light at 5.30! Even the sun is already there. Not yet visible through the high-rise buildings, but you can already see the glow <against the side façade of the Opera House and the terrace of the Gustav, but above the trees also full against the façade of the Theatre Building>; even notice how warm the day will be.
There is already some traffic. The first trams pass. Some buses are already turning; few cars yet; seven taxis are waiting in line. Some pedestrians and a few people waiting on platforms already, though. Also already quite a few people on the trams. A bus arrives from Turnhout, almost full (?)
The city's Orange cleaning van comes to clean the gutters. A fat, slovenly woman walks in front with a brush. The sun rises <6.50am> above the roof of 'Geuzenhof' café looking between two taller facades, casting yellow light over the square."

[207] Translation: "Have you noticed that […]"

[208] Translation: "Look how that […]"

> ++++ [213]~~Het plein om half zeven.~~[248] [259]~~Je m~~[260] [261]~~Het wordt~~ [334]~~Je boelt~~[279][278] ~~voelt~~ ~~aan al~~ ~~a~~[286]~~les~~ [285][317] ~~dat het~~ [280][335] ~~kan al~~ [281]~~we~~[264]~~d~~[262][265]~~q~~[263]~~er~~ ~~bakke~~ [374][372]~~en~~ ~~stra~~[375][373]~~aks~~[376] ~~zo'n~~ ~~warme~~[315]~~bloedhete~~[313][316]~~warme~~[314] ~~d~~[283]~~ag~~[377][282]~~, let maar op~~[336] ~~ten~~[266]~~heo warm de~~[267] ~~dag~~ [284][268]~~zal~~ [310]~~wordt~~[311][337] ~~en~~[312].[393] [381]<u>Heb je ook</u> <u>g</u>[397]~~G~~[398]<u>ezien</u> <u>hoe</u>[399]~~dat~~[400] het digitale bord boe[382]<u>ve</u>[383]n de [384] ~~k~~[386][385]~~hoek~~[395][387]apotheek[388]al tween[389][390]ëntwintig[391]graden[392] aangeeft? <u>Het</u> <u>wordt weer bakken straks, let maar op.</u>[394] (Session 17)

The genetic material also shows Bogaert's reflections on whether or not to explicitly mention Franklin's square-ness. In the fourteenth session (2 August 2013), Bogaert writes a note to himself in the Word document: "Hier al iets over plein(en) - voor de duidelijkheid?"[209] Then, in Session 17 (23 August 2013), he starts writing 'Squares are', with a few plus signs behind it.

> [308]Plei~~e~~[309]nen <u>zijn</u> [403]+++ [348]Vuilnis. [294]Over pleinen. [326]IK (Session 17)

He deletes the note itself in this session. This fragment is finished in the next session with an entry about Franklin's siblings. At the end of the session this fragment consists of:

> -- Elders? Pleinen hebben de gave van het overzicht. Wij maken geschiedenis. Vraag het aan mijn broers en zussen hier in de stad of elders in het land, of beter nog: aan mijn beroemde en beruchte neven en nichten in Moskou, Peking, Rome of Caïro.++ (Session 18)[210]

This excerpt is based on a reflection by Hilde Van den Eynde in *De Standaard* about the role of squares in revolutions, which is included in the notebook. She states: "De revolutie vraagt om een plein".[211] Bogaert takes this statement one step further and has Franklin say: "Wij maken geschiedenis".[212] Squares play an important role in revolutions and therefore make history. The addition of "-- Elders?"[213] shows that Bogaert is considering moving the fragment.

Another reference to being a square is the addition of Franklin's complaint that people rarely have an eye for what a square – and therefore he – has to offer:

---

[209] Translation: "Something about square(s) here - just to be clear?"

[210] Translation: "-- Elsewhere? Squares have the gift of overview. We make history. Ask my brothers and sisters here in the city or elsewhere in the country, or better still, my famous and infamous cousins in Moscow, Beijing, Rome or Cairo.++"

[211] Translation: "The revolution demands a square"

[212] Translation: "We make history"

[213] Translation: "-- Elsewhere?"

[54]Voor wat [166]een [195]plein~~en~~[196]kan[198]~~heeft te~~[197][199]bieden[167] hebben ze[168]nauwelijks~~/geen~~[200] oog...[169]. (Session 21)

In Session 23 (23 August 2013), Bogaert moves the fragment about squares and the making of history to Franklin's second chapter and deletes the sentence "Voor wat een plein kan bieden hebben ze nauwelijks oog".[214] Even without these explicit references to squares, it is already evident that Franklin is a square.

The text-genetic material thus shows the dynamic between making Franklin's 'square-ness' explicit and implicit in the first chapter. Instead of explicitly referring to squares, Bogaert chooses to include more subtle, implicit references to the Franklin Roosevelt square. This indicates a form of interaction with the text produced so far. Bogaert first made his ideas and formulations very explicit, to help him convey the purport of the novel. Through a process of re-reading, the information density could gradually be reduced or concealed.

Troch: Killing setting?

Herman and Vervaeck point out that the spatiotemporal background against which the story develops is generally seen as relatively fixed – as a static motif (Russian formalists) or pure index (Barthes). They state that "[b]oth terms are appropriate since the fictional universe does not cause the story to develop", but they do emphasise that "story development is inconceivable without the setting, which makes it possible for actions to take place and actants to become involved in them" (Herman and Vervaeck 2019, 61). Actants and events are impossible to imagine without their embedding in time and space (61). There will always be a fundamental connection between actions, actants and setting (61). In Troch's "Mondini", however, the setting does seem to cause the story to develop. The short story belongs to the genre of climate fiction: it describes the daily activities of the I-narrator, who shelters from the heat and away from other humans in a mountain village called Mondini, while the world has altered considerably due to climate change. Parts of the *syuzet* give insight to how the ever-increasing heat forced him to attempt to survive, alone, in the mountain village. In "Mondini" the setting has changed so considerably due to climate change that it has also changed people's worldview and the way they interact with each other in the fictive present. The reconstruction of the writing process allows us to follow how Toch implemented the change in worldview into the short story and eventually used it to describe what happened to the partner and daughter of the I-narrator.

In the first session in which Troch starts working on the story (Session 2, 10 August 2020) – he types the approximately 700-word long opening sentence. In this version, the I-narrator is already hiding out in a desolate mountain village, but Troch

---

[214] Translation: "For what a square can offer, they hardly have an eye"

does not yet provide any information about the exact reason *why* he is hiding. The world the I-narrator inhabits still has running water and electricity, and although he tries to avoid people, when he does meet others, they are rather friendly – too friendly for his liking:

> maar als je dan toch iemand tegenkomt, heeft die zo goed als altijd zin om je staande te houden voor een praatje en de praatjes vrolijken me niet bepaald op, want veel diepgang kennen ze niet, er worden geen vragen gesteld, men durft de huidige toestand niet aan de kaak te stellen en dat zou men nochtans beter soms wel eens doen, (Session 2)[215]

In this world, people still want to talk to each other, preferably about light-hearted topics. The reference to "the current situation"[216], however, seems to imply that there is something going on in the world, but this has not really affected people's behaviour and worldview – apart from, perhaps, that of the I-narrator, who acts as if he is "the last human being on this godforsaken planet".[217]

The world already changes a bit in the fourth session, which took place only a couple of hours later. Whereas in the previous version it was only very unlikely that it would snow again, skiing is now something from the 'distant past'. Also, all appliances that run on electricity have lost their practical use. As the I-narrator puts it: "We zijn terug naar het stenen tijdperk geworpen. Wat er van we nog overblijft".[218] People no longer form a community; they are all on their own. People no longer talk to each other; they just look at each other cautiously from a distance before finally running away. Troch is thus already describing a tiny shift in worldview, one that affects human interaction.

At the end of the text, Troch writes a conceptual note suggesting that the story is set in a world in which there seems to be no life. It also raises the question of how the narrator got there and what he is saying about the past. Firstly, this can be interpreted as two questions that Troch wants to explore in the short story. However, they can also be interpreted as questions that Troch is asking himself – he needs to think about these questions before he can elaborate on the story at all:

> En man alleen in een appartementje boven een leegstaand winkelpand in een uitgestorven bergdorpje - met de naam Mondini, wat Italiaans zou kunnen

---

[215] Translation: "but when you do meet someone, they almost always want to stop you for a chat and the chats don't exactly cheer me up, because they don't go in depth, no questions are asked, people don't dare to denounce the current situation and they should do that sometimes,"

[216] Translation: "de huidige toestand"

[217] Translation: "de laatste mens op deze godvergeten planeet"

[218] Translation: "We have been thrown back to the Stone Age. Whatever is left of we"

zijn. Hij tuurt angstig naar buiten. Er is geen leven meer, er lijkt geen leven meer te zijn. Hoe is hij daar terechtgekomen, wat vertelt hij over het verleden?[219]

As already discussed in Chapter 5.3, Troch has not yet fully conceived his dystopian world. In Session 5, however, it becomes increasingly clear that the narrator lives in a world that is different from the one we know: there is no electricity nor running water and it is stiflingly hot even in January.

In Session 6 (12 August 2020), Troch writes a part of the backstory that briefly summarises how the world changed due to climate change, in particular the consequences of the ever-increasing heat. Troch mainly focuses on the effects of the heat on the world, resources and the well-being of the people. A couple of hours later, in Session 7, he elaborates this passage mainly by describing how people's behaviour changed. With these revisions by extension (see Chapter 4), he includes a whole new level of consequences:

> Eerst stegen de temperatuurgemiddelden gestaag, toen exponentieel. Groen gras werd zeldzaam, niet al te voedzaam voor vee. [68]Koeien kalveren niet meer. Alsmaar minder plekken werden leefbaar. [127]De voede[128] [129]selketen kwam stil te vallen. [130]De econo~~no~~[131] [132]mie. [133]Het bruto nationaal product van elk~~k~~[134] [135]land kelderde. [136]De beurs~~[137]~~ [138]zen crashten. [139]Plunderingen. ~~[140]Het was~~[141] [142]Een pretje was het allemaal niet. [149]Ik verschanste me onder bruggen, ~~leegst~~[150] [151]verlaten panden, ~~i~~[152][153]doodde als ik moest doden. [143]Ik weet niet waaraan ik het verdiend heb om het te ~~vo~~[144] [145]overleven. <u>Ik weet niet hoeveel levens ik heb, maar dit is wellicht het laatste.</u>[147] Men probeerde zijn heil elders te zoeken. [69]Natuurlijk kwam daar handgemeen van. [126]Men ging om het minst met elkaar op de vuist. [71]Men moorde elkaar uit. [70]Beschaafd is men nooit geweest. Men zocht koelte. Die was haast niet te vinden. Er sneuvelden er velen, van uitputting en uitdroging. [72]Men sneed elke[73] [74]aar de keel door en at elkaar vervolgens op. [75]Tot op het bot. Men viel bij bosjes. Men vocht om drinkbaar water. Om een plek in de schaduw. Ook ik. (Session 7)

The effects of climate change forced people to seek refuge elsewhere, and they began to fight, kill, and even eat each other. Here we can see Troch's interaction with the text produced so far: because he had first described how the world changed, he could later add how this also affected human beings. In expanding this paragraph, Troch also focuses on the behaviour of the narrator, who also *killed* when he could. And this last addition, that the narrator also had to kill, leads to the addition elsewhere in the text

---

[219] Translation: "And a man alone in an apartment above an empty shop in a deserted mountain village called Mondini, which could be Italian. He looks out in fear. There is no more life, it seems. How did he end up there, what does he say about the past?"

that the narrator keeps hearing the death rattle of his partner: "en toch blijk ik maar haar doodsreutel horen" (n154-156).[220] Because of this connection in time between the two revisions, we can ask what the connection is between the narrator's having to kill and his partner's death. Something in this description must have triggered an idea for Evelina's death.

In Session 12 – still on 12 August – Troch moves this description of the world to a place near the end of the document. It is no longer part of the text but can later be consulted for similar descriptions. Now it is by means of the references to the world in the fictive present that the extent to which the world has changed becomes clear. These references are constantly elaborated on and altered. They emphasise the opposition between inside (the apartment), which is safe, and outside, which is unsafe.

The backstory is addressed again in Session 23 (27 August) as Troch inserts a passage that both describes how the consequences of the climate change had an impact on the environment and the interaction between people as well as how this led to the loss of the narrator's partner and daughter. In "Mondini", people's behaviour is geared towards self-protection, no matter the sacrifice. During the writing process for this paragraph, Troch shifts between whether it was this change in behaviour – the violence – or the climate change itself that caused most of the deaths.

Troch begins with the first-person narrator describing how there was little left of the North and South Pole when they tried to escape. Rising sea levels forced everyone in low-lying areas to pack their bags. This changed people's behaviour: they became more violent towards each other.

> Het probleem van veel volk op een beperkte ruimte loste de mens zelf op met menig handgemeen (After n71)

> veel te veel volk op een beperkte ruimte; problemen lost de mens graag op met menig handgemeen en lynchpartijen. (After n102)

> veel te veel volk op een beperkte ruimte; de mens die een probleem graag oplost met een handgemeen of een lynchpartij. (After n116)

The above versions show how Troch is constantly exacerbating the impact of climate change on human behaviour. At first it just causes a scuffle, then a lynching. Next, he adds that this led to a thinning of the population, which was made worse by the extreme temperatures:

> veel te veel volk op een beperkte ruimte; de mens die een probleem graag oplost met een handgemeen of een lynchpartij; en als dat onvoldoende was

---

[220] Translation: "and yet I keep hearing her death rattle"

> om de bevolking uit te dunnen zorgde de onmenselijke temperaturen er wel voor dat er slachtoffers bij bosjes vielen. (After n140)
>
> de mens die een probleem graag oplost met een handgemeen of een lynchpartij. De onmenselijke temperaturen werkten de uitdunning van de wereldbevolking alleen maar in de hand. (After n159)
>
> de mens die een probleem graag oplost met een handgemeen of een lynchpartij. Maar de onmenselijke temperaturen zorgden voor de echte genocide. (After 188)

In the end, it is not man but climate change that causes most deaths. During this implementation of the backstory, Troch also starts to focus on Evelina and Lily, the partner and daughter of the first-person narrator, and what happened to them. The versions below indicate that Troch started with the idea that it was the severe heat that became fatal for them:

> We wisten uit handen van bendes te blijven, maar de hitte is onverbiddelijk. (After n175)
>
> We wisten uit handen van bendes te blijven, maar al volgden we zoveel mogelijk waterwegen, de hitte is onverbiddelijk. (After n180)
>
> De onmenselijke temperaturen zorgden voor de echte genocide. Ook Evelina en Lily zijn een van de slachtoffers. (After n196)
>
> De onmenselijke temperaturen zorgden voor de echte genocide. Ik vond het wijs, het was wijs om langs zoveel mogelijk waterwegen te verplaatsen en toch bleek de hitte ook voor Evelina en Lily onverbiddelijk. (After n202)
>
> Het leek mij het verstandigst om langs zoveel mogelijk waterwegen te verplaatsen, om te drinken, om te verfrissen en toch bleek de hitte ook voor Evelina en Lily onverbiddelijk. (After n207)

The insertion of "om te drinken" leads to the insertion of new text: everybody wanted to drink. The revision that included the drinking of water in the sentence, thus prompted the idea to bring back the possibility of human involvement in the death of Evelina and Lily:

> Iedereen zocht verfrissing, iedereen wilde drinken, drinken, drinken. Evelina, Lily en ik waren niet de enigsten die zich zoveel mogelijk langs waterwegen verplaatsten. Ze hebben het niet gehaald hoe ik ook mijn best heb gedaan, ze hebben het niet gehaald. (After n219)

At the end of the session, the paragraph does not yet provide a clear answer for Evelina and Lily's cause of death, although the large number of people moving along the waterways seems to be related to it. The writing actions of the following session make clear that Troch had not resolved this question either. First, he added that the I-narrator tried to quench their thirst. This shows that the climate (the heat and drought) was eventually fatal for them. But Troch then deletes this and instead adds that they screamed for mercy, describing how the I-narrator tried to defend them too. As such, it is more likely that a human action killed them, rather than the heat:

> Zij, mijn twee oogappels, hebben het niet gehaald, hoe ik ook mijn best heb gedaan, ~~hoe ik ook gepoogd heb hun dorst te lessen,~~ [235][236] hoe hard ze ook om genade hebben [238] ge[239]gild~~en[240]~~, [237]hoe ik hen ~~ok[233][232]~~ok met hand en tand heb verdedigd, [234]zij hebben het niet gehaald. (Session 24)

In Session 25 (27 August) Troch expands the paragraph even further, going into details of the fatal event. The I-narrator remembers how they were traveling with about seven people, but that these people suddenly started pulling, pushing, and shouting. Before they knew it, the family had ended up in a brawl. The I-narrator tried to defend them, but before he could get hold of everyone it was too late; they had already got hold of Evelina and Lily. In the end, it was not the climate but the way that the climate made people behave, that took the lives of Evelina and Lily.

By means of the genetic material, we can trace the development of the setting; firstly, through Troch's quest for how much the world changed – the environment as well as people's behaviour – due to climate change and secondly, in his decision-making process over whether it was the heat, or this change in behaviour – the violence – that led the narrator to survive alone. In making these decisions, Troch interacted with the text produced so far, which helped him to further develop his ideas. Moreover, the keystroke logging also proved that Troch made a connection between human behaviour (that people were forced to kill) and the death of Evelina at an early stage in the writing process. Even though he was not working on a description of Evelina, text production regarding the changes in the world prompted the addition that the narrator still hears her death rattle.

6.2 NARRATIVE

The second level of structuralist narratology is narrative, which is the "the concrete way in which events are presented to the reader" (Herman and Vervaeck 2019, 64). Within the level of the narrative, Herman and Vervaeck make a distinction between time, characterisation, and focalisation. Within this section, I address how the keystroke logging data can enrich the analysis of these aspects.

6.2.1 Time

Bogaert: Using the fictional present to generate ideas about the past

As Herman and Vervaeck point out, time is analysed within structuralism "by studying the relation between the time of the story and the time of the narrative" (2019, 65). In order to achieve this, Genette for example, used three criteria to systematise the different aspects of time: duration, order, and frequency (Herman and Vervaeck 2019, 65). For duration, "the time necessary to read the account of an event" is compared "to the time an event takes on the level of the story" (65), and with frequency, "the relation between the number of times an event occurs in the story and the number of times it occurs in the narrative" is studied (65). Here, however, I focus on the order, which is "determined on the basis of the relation between the linear chronology in the story and the order of events in the narrative" (69). Again, Genette distinguishes three categories to specify order: direction, distance, and reach. These categories are analyses concerning the primary narrative, which functions as a norm. With this primary narrative as a basis, two directions are possible. When there is a backward direction, such as a memory, it is termed an analepsis or flashback. And when there is a forward direction, such as an anticipation, it is termed a prolepsis. Distance, then, concerns the "temporal gap between primary narrative on the one hand and prolepsis or analepsis on the other" (70); reach denotes "the stretch of time covered by the analepsis or prolepsis" (70).

The genetic material of *Roosevelt*, specifically the sections narrated by the character Faraaz, presents an example of how writing the primary narrative can invoke ideas in the analepsis. In the second section narrated by Faraaz, Fran's mouth reminds Faraaz of Naheed, who could draw with her mouth. Seeing Fran's mouth therefore invokes a prolepsis. But what triggered Bogaert to make this connection in the first place?

In Session 227 (16 August 2014), Bogaert mentions Gaard and Naheed for the first time in the Word document:

> [6]Gaard met[10]en[11] Halisse op Gustav[13]terras[12] gei[7]zi[8]en[14] 9.[15]Afrekenin[16]g[17]/afgerekend? [18]Over N[19]n[20]aheed[21]. (Session 227)

These sentences are roughly based on the notes in the Atoma notebook, where he planned that Faraaz would encounter Gaard and Fran (here still called Halisse), and later also added that this section would mention Naheed as well.
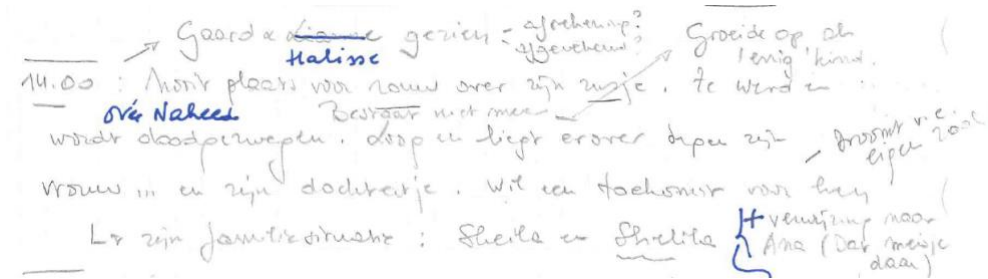
Figure 6.2: Detail from the Atoma notebook: rough outline of the section narrated by Faraaz at two o'clock.

By looking at the nanogenesis, we can investigate the decision-making process and the process of creative invention in establishing the link between Fran's mouth and Naheed drawing with her mouth. In Session 243 (26 September 2014), Bogaert starts to work on the paragraph describing a conversation between Gaard and Fran, the latter is then still called Halisse. At a certain moment he starts to describe their appearances. He writes, including some typos which he almost immediately corrects: "De man kan haar vader niet zijn. Ze lijken helemaal niet op elkaar. Misschien hebben ze een relatie" (n81-91).[221] Then he makes some changes around these new sentences (n92-94), after which he writes: "Ik weet: leeftijd is geen punt" (n95).[222] This already establishes a connection with Faraaz's secret: age was not a problem for Naheed's forced marriage. After some other modifications in the paragraph above, Bogaert adds to the paragraph that following: "Niemand weet dat ik geen enig kind ben. Behalve mijn ouders" (n102–103).[223] This shows that he simultaneously worked on the paragraph describing Gaard and Halisse, and on the one introducing Naheed. He adds two asterisks before these sentences to remind himself to elaborate this (n106). Then he deletes "helemaal" ["at all"] from the sentence "Ze lijken helemaal niet op elkaar" (n107-108),[224] and starts describing how their appearances differ: "Hij is groot en , zij is van vreemde afkomst" (n109).[225] He inserts the sentence "Misschien slapen ze met elkaar" (n110)[226] and modifies the sentence "Ik weet: leeftijd is geen punt" into "Leeftijd is geen punt" (n111-112).[227] Then he adds "slank" ["slim"] and "scherp van gezicht" ["sharp-faced"] to the description of Gaard (n113) and makes from the descriptions of Gaard and Halisse two separate sentences (n114-117). He adds "ook hier", making it "Leeftijd is ook hier geen

---

[221] Translation: "The man cannot be her father. They don't look alike at all. Maybe they have a relationship".

[222] Translation: "I know: age is not an issue".

[223] Translation: "No one knows I am not an only child. Except for my parents".

[224] Translation: "They do not look alike at all".

[225] Translation: "He is tall and , she is of foreign origin".

[226] Translation: "Maybe they sleep with each other".

[227] Translation: "Age is not an issue".

punt" (n118).[228] Then he makes some changes in the previous paragraph (n119-120) and modifies the descriptions of Gaard and Halisse into: "Hij is groot en slank, met een scherp gezicht. Zij is vreemd. Misschien oosters. Of ge… Afrikaans" (n121-130).[229] Then he writes "* Naheed was/had zo'n haar?"[230] after this paragraph (n132-139). This indicates that Bogaert now came up with the idea to make a connection between the appearance of Halisse and that of Naheed: the similarity between Halisse and Naheed was thus activated. Therefore, seeing Halisse could trigger memories of Naheed.

After a couple of changes in other paragraphs, Bogaert changes the sentences of the appearance again into: "Hij is groot en slank. Hij heeft een scherp gezicht met een hoge neus en bijna groene ogen. Zij is donker en niet groot. Ze heeft ronde bruine ogen. Ze is vreemd, misschien Anatolisch of gemengd Aziatisch" (n142-185).[231] In the process, Bogaert doubted about the colour of Gaards eyes, whether they were blue or green. A couple of modifications later, he inserts "een mooie, sensuele mond"[232] into the description of Halisse, which replaces the description of her eyes (n189-192). He searches the word "mond" online, starts writing "ro" – probably intending to write "ronde" ["round"] – but immediately deletes it and continues writing "volle" ["full"], which replaces "mooie" (n193-195). He substitutes "Aziatisch" with "oosters" ["eastern"] (n196-199). After making some modifications in the first paragraph about Gaard and Halisse, Bogaert continues writing the introduction of Naheed. This can be summarised into the following stages, in which the reference to her hair is replaced by a reference to her mouth:

> * Naheed was/had zo'n haar? Mijn zus(je). (n205).

> * Ook Naheed had zo'n haar? Mijn zus(je). (n206-209).

> * Naheed kon iets met haar mond? zo'nhaar? Mijn zus(je). (n210-214).

> * Naheed kon … met haar mond. We speelden Mijn zus(je). (n215-220).

Then he continues with writing the paragraph describing how nobody knows that he is not an only child (n221-278). In writing this paragraph, he probably came up with the idea of what Naheed could do with her mouth: drawing.

---

[228] Translation: "Age is not an issue here either".

[229] Translation: "He is tall and slim, with a sharp face. She is foreign. Oriental, perhaps. Or ge… African".

[230] Translation: "* Naheed was/had such hair?"

[231] Translation: "He is tall and slim. He has a sharp face with a high nose and almost green eyes. She is dark and not tall. She has round brown eyes. She is foreign, perhaps Anatolian or mixed Asian".

[232] Translation: "a beautiful, sensual mouth".

> * Naheed kon tekenen met haar mond. Potloodstolmpje. Handen achter haar rug. We speelden Mijn zus(je). Lachen.++(n279-288).

The nanogenesis thus shows that the link between Halisse and Naheed was established during writing – it was not planned in advance. Because Fran, as Halisse, was "Oosters" and "vreemd" in this version, Bogaert was able to make this connection.[233] Initially it was the similarity their hair that triggered the memory of Naheed. Later, when Bogaert changed the description of Halisse's hair into a description of her mouth, the idea of giving Naheed a property linked to her mouth was invoked.

Troch: The distance of the past

Just as Troch had to establish the specifics of his dystopian setting, he also had to decide how long it had been since climate change led to the catastrophe. The distance between one of the analepses – the narrator's memory of fleeing with his partner, and his daughter – and the primary narrative is between six and seven years, but this distance varied during the genesis of the story.

In the second writing session Troch worked on the story (Session 4; 10 August 2020), he already added several time indications, but these did not specifically address the time since the family fled from the heat and the drought. One of the time indications is the moment that the narrator saw his partner's smile for the last time, which was 'a long time ago':

> [128]De laatste keer dat ik naar haar (naam?)[137] glimlach keek, is op[138] de dag dat de [131]bate[129] [130]terij van mijn smartphone heeft [132]het allang [133]begeven[134]af[135]. [136]En dat is lang geleden. (Session 4)

It does not concern the last time he had seen her in person, but the last time he could see her picture on his smartphone. Somewhat later that day, in Session 5, this changes, as Troch substitutes "keek" ["looked"] with "streelde" ["stroked]:

> De laatste keer dat ik naar [19]haar (naam?) glimlach keek[20] streelde[21], is op de dag dat de batterij van mijn smartphone het begaf. (Session 5)

The narrator now stroked the smile of his partner for the last time on the day his phone battery died, which was a long time ago. And in Session 7, on 12 August, Troch adds to this that it happened long before the narrator came to Mondini: "Lang voor ik hier kwam" (n167).[234] And importantly, the narrator now also has a child that his partner carried on her back in that particular moment. In the next session (Session 12), a couple

---

[233] Translations: "Eastern" and "foreign".
[234] Translation: "Long before I came here".

of hours later, he expands this paragraph by adding that the narrator took the picture of his partner and child on the morning of their flight, and that the start of their flight was 'long ago':

> ~~Wannee[60][61], wanneer was d[62]~~D[63]e laatste keer[86] Hoelang is het geleden[87] dat ik die ene weerbarstige lok voor haar (naam?) rechteroog probeerde weg te vegen en in een moeite door~~streelde ¹~~over het hoofd van ons kind (naam?) [81] streelde ²dat ze op haar rug droeg?[66]~~, is op de dag dat de batterij van mijn telefoon het begaf.[67]~~ [88]Lang. [73]Ik nam de foto op de ochtend[76] ~~in van onze[77]~~ dat we[78] vlu~~h[74]~~ [75]chtten, stelde hem in als achtergrond, putte er energie uit telkens ik ~~hem zag[84][83]~~mijn telefoon ontgrendelde[85][79]. ~~En dat is l[68]L[69][70]~~ Het begin van onze ~~lvucht[90][89]~~vlucht, het [91]is l[71]ang geleden. (Session 12)

These time indications are altered a bit in the next session (Session 13; 19 August). Troch adds that it is 'inhumanly long ago' (n2) that the I-narrator stroked his partner's hair but deletes the specification that the start of their flight was so long ago. Later that day (Session 15), the time frame between their flight and the primary narrative is made explicit since Troch includes the age of the child in the story. The still nameless child, although he was briefly called Pjotr during this session, was three at that time and would now be six or seven. From abstract, general time indications, Troch now clarifies the distance of the analepsis: around three or four years ago.

> ~~Hoelang is het geleden dat ik d[44]~~ D[45]ie ene weerbarstige lok voor Evelina [28]~~haar (naam?) [29]~~rechteroog ~~probeerde[46]~~ zal ik nooit weer[47] ~~proberen[49]~~ weg ~~te[48]~~ te [50]vegen, nooit meer zal ik[51] ~~en[52]~~ in een moeite door over het hoofd van ons kind ~~(naam?) [30]~~stree~~[53]ld[54]~~[55]en[56] dat ze in een draagzak op haar rug droeg.[57][58] ~~De foto nam ik[71][140] [68]~~,[70] ~~toen[141]~~ ~~Pjotr[72][73]~~ (naam?)[74] was toen [142]nog maar drie.[143] Hij zou nu zes zijn, zeven misschien.[144] ~~was[145].[76]~~ ~~Lang. Onmenselijk lang.[59]~~ ~~Ik nam d[60]~~ D[61]e foto ~~nam ik[62][69]~~ ~~op de ochtend dat we vluchtten[41]~~er eindelijk in ~~gelukt waren de grens naar Italië over te steken[42][75]~~, stelde hem in als achtergrond, putte er energie uit telkens ik mijn telefoon ontgrendelde.[146] (Session 15)

In the following writing session, on August 21, Troch changes the age of the child, who now also became a daughter called Lily. She was three months old when they fled, and she would still be six to seven years old by now. This change in age also changes the distance, the analepsis now happened six to seven years ago.

> Die ene weerbarstige lok voor Evelina haar rechteroog zal ik nooit meer proberen weg te vegen, nooit meer zal ik in een moeite door over het hoofd van ons kind strelen dat ze in een draagzak op haar ~~rug[321]~~buik[322] droeg. ~~(naam?)[5]~~ Lily[6] was toen ~~nog maar[317]~~ drie maanden oud[318]. ~~H[7]~~Z[8]ij zou nu zes zijn, zeven misschien. [323]Jaa~~[324]ren[325].[326]~~ar.[327] [328]Geen manden. (Session 16)

This explicit designation of years is deleted in Session 22, Troch changes it to a more implicit indication of how old his daughter would be by now. She would be able to write, read, and manage to count much more than one plus one:

> Ze was toen drie maanden oud. Nu zou ze zes zijn, zeven misschien. ~~Jaren. Geen maanden.~~[4] [5]Ze zou kunnen schrijven, lezen~~,~~[19]~~rekenen~~[6] en veel mee~~r~~[21][20]er dan [22] [7]een bij een ~~ku~~[8] [10]~~wet~~[9] [11]~~kunnen o~~[12] [13]weten op te tellen. (Session 22)

The distance changes one more time, in Session 29 (17 September), just before the final version is completed. Troch reduces the distance with two months by making Lily five months old instead of three. This change to five months was actually triggered by the insertion of a new sentence, that adds a completely new aspect to the narrator's family. The narrator's partner, Evelina, appeared to be pregnant on the day she was killed by some people they met on the road:

> Evelina gilde dat ze op moesten houden, maar ze hielden niet op~~.~~[41] [39]~~Z~~[40], z[42]e ~~negeerden~~[43]gaven geen sikkepit om[44] haar bolle buik. Lily krijste oorverdovend. (Session 29)

A second pregnancy only three months after giving birth to the first child is possible, although not recommended, but already having a 'rounded belly' at that time seems highly implausible. This possibly led Troch to immediately change the age of Lily from three to five months, which makes it a bit more likely that Evelina's second pregnancy was already visible. The eventual distance is shortened by two months, but still remains generally between six and seven years. The genetic material of Troch's story thus shows how the distance was a variable aspect of the narrative during the writing process.

6.2.2 Character

Van Rijswijk: 'My weird brain'

Within narratology, characterisation "concerns the way in which a character is present and represented in fiction" (Herman and Vervaeck 2019, 73). Rimmon-Kenan distinguishes three ways of characterisation. The first is direct characterisation, in which a character is introduced by describing various character traits, such as outward appearance and psychological states (Herman and Vervaeck 2019, 73). An important question in this regard is who provides the characterization since it could present wrong or subjective information. The second is indirect characterization, which is based on metonymy: "it works with elements that are contiguous with the character" (74). These include actions, discourse, specific style, and words that the character uses, their norms and their values, and their physical appearance and environment (74). The third way is using analogy, "which leads to metaphor instead of metonymy" (74). During the writing

process, the author thus has several options at hand to present information about the character to the reader. This often entails a search for the right balance between showing or telling character traits. As described above, the text-genetic material of *Roosevelt* demonstrates the dynamic between making Franklin's 'square-ness' explicit and implicit in the first chapter. Instead of explicitly referring to squares, Bogaert opted to include more subtle, implicit references to the Franklin Roosevelt square, but he could only arrive there by making the references to squares very explicit at first.

Herman and Vervaeck state that "the decision concerning (un)reliability largely lies with the reader" (Herman and Vervaeck 2019, 96). In the writing process of "Zorgvlied" we can see through a nanogenetic analysis the modifications Van Rijswijk makes that enable the reader to make their own decision about the reliability of the first-person narrator. Van Rijswijk allows the reader to decide whether to believe the narrator's story that she can perceive ghosts, and that – at least in this fictional world – ghosts do exist, or to consider the narrator unreliable, that she only imagines the ghosts.

There are some surreal aspects of the story, such as the Asian man whom the first-person narrator and her mother meet twice in the graveyard: once with a pink rose in his mouth and a life jacket in each hand – the life jackets were added in the final writing session and replaced watering cans – and the other time with a wheelbarrow of uncooked macaroni. And some of the narrator's descriptions of the surroundings also express her imagination, such as some of the watering cans, which are described as 'seeming to only pretend to be watering cans', but that 'they were still tiptoeing around the grounds on spider legs' before the narrator and her mother arrived. These elements can be interpreted both as evidence of the first-person narrator's over-imaginative mind and as elements of the story world that open up the possibility of ghosts.

During the writing process, however, Van Rijswijk often wrote sentences containing direct characterisations of the first-person narrator that refer to her mental state – or, more literally, to the way her brain works. The genetic material thus suggests that Van Rijswijk, while writing, had to constantly suppress a tendency to attribute the ability to perceive ghosts to a malfunction of the narrator's brain. Some of these references remain, but most of them are immediately deleted.

In Session 1 (7 December 2020), for example, Van Rijswijk started to write a sentence that would explain why the I-narrator forgets things so often. It has something to do with her brain:

> [192]~~Dat ik zo veel vergeet komt mda~~[193][194]~~omdat~~[195][196]~~doodr~~[197][198]~~rdat mijn brein zich van~~ [199] (Session 1).

Yet, before this sentence was finished, Van Rijswijk deleted it. Something similar takes place in Session 3 (8 December 2020), as she is writing the paragraph of the I-narrator's

first perception of her grandmother's ghost. The I-narrator thinks that she is made for this (perceiving ghosts), and so finally her brain has enough to do:

> [70]Hier ben ik voor gemaakt, dacht ik terwijl ik moeiteloos [75]twee gesprekken tegelijk [76]ga~~nd~~[7172]ande hield. [73]~~Eindelijk heeft mijn brein genoeg te doen.~~[74] (Session 3)

As such, it appears to be her brain enabling her to perceive what others cannot. But again, this sentence is immediately deleted after it was written. However, this idea returns in Session 6 (9 December 2020), when Van Rijswijk inserts a clause that states that this is what the brain of the narrator is made for:

> Hier ben ik voor gemaakt, dacht ik terwijl ik moeiteloos twee gesprekken tegelijk gaande hield~~.~~[543],[544] [542]~~D~~[545]d[546]it is waar mijn rare brein voor bedoeld is. (Session 6)

This clause does make it into the published version and is one of the indications that the narrator considers her brain to be weird.

Session 3 contains another reference to the mental health of the I-narrator, as Van Rijswijk ends the writing session with pinning down some ideas she intends to include in the session to come:

> [116][.....]/haar leven/dat ze mee m~~oe~~[117][118]ag lopen/neemt ~~ineens~~[119][120]steeds meer mensen mee/niet als ik m~~id~~[121][122]edicijnen slik. (Session 3)

The last idea suggests that the I-narrator is not able to see the ghosts when she had her medication, which implies that it may indeed be a 'malfunction' of her brain. And indeed, in the following session (session 5, 9 December 2020), Van Rijswijk starts to include this in the story:

> [79]~~Later mertke k~~[8081]~~kte ik dat ik als ik mijn medicatie nam minder vaak bezoke~~ [8283]~~ek kreeg van de overledenen, en omdat~~[84] (Session 5)

During the formulation of this sentence, Van Rijswijk seems to change her mind about this explicit reference to the effect that the medication has on perceiving ghosts and deletes the idea mid-sentence. Lastly, Van Rijswijk wrote in Session 6 two other explicit 'judgements' about the brain of the I-narrator. First, the I-narrator tells the ghost of her grandmother that they can drink a glass of wine, although it is not even 10 o'clock in the morning, since she is currently not working. For a very brief moment, the I-narrator was not working because of a minor malfunction in her brain:

> Ik schonk ons, ook al was het nog geen tien uur 's ochtends, een glas wijn in. [289]Ik zei dat ik even niet aan het we~~k[290][291]~~rk ~~was door een kleine storing~~ [295] was[296]~~aan mijn[292][293]~~in mijn brein[294], dus dat dat best kon. (Session 6)

Just like the other statements about the brain, Van Rijswijk removed it shortly after she had written it. This happens yet again when Van Rijswijk lists several moments on which the I-narrator and her mother could have died but did not. Again, the brain is mentioned, it 'abandoned' her:

> [45]De keer ~~dat iemand[606]~~ mijn moeder ~~voor een rijdende tram wegtrok[607]~~ met onverklaarbare koorts in het ziekenhuis werd opgenomen[608], de keer dat ik als kind ineens stopte met ademen, ~~haar hart dat het begaf[609]~~die bezorgwagen[610], ~~mijn brein dat me in de steek liet waardoor ik [4647]~~ ~~en[68]~~ wie weet welke macabere dansjes we allemaal nog meer [69]ontsprongen zijn[55]. (Session 6)

This moment was immediately deleted as well. Almost all the explicit references to the (dysfunctional) brain of the I-narrator were immediately deleted, either by means of pre-contextual deletions – before the sentence was even finished – or as immediate contextual deletions – a deletion of a contextual addition to an already finished sentence. This means that these revisions were only visible because the writing process was logged, otherwise they would not have been easily recoverable. The nanogenetic analysis of Van Rijswijk's writing process shows that she had the tendency to characterise the I-narrator as someone who has 'a weird brain', and that she constantly corrected herself. While the peculiarities of the narrator's brain remain present in the published story, it is not as explicitly mentioned as during the writing process, leaving room for the interpretation of the reader.

Posthuma: What would the character think or do?

As mentioned in Chapter 5, Galbraith (1999) argues that the writer's disposition plays an important role in the writing process. As Heeks clarifies, this means that "thinking and writing are filtered through the writer's own self, engaging a writer's own individualistic thinking habits and thought processes" (Heeks 2012, 252). In his study of discovery during writing in different genres, including fiction writing, Heeks found that fiction writers strongly associate discovery with the writing of characters and dialogue (230). He states that in order to get an understanding of a character, the writer can try to "inhabit" that character:

> This process of inhabiting a character, of seeing the world through their eyes, seems to suggest an experience that is curiously different from writing within the writer's own disposition. There is a sense that the writer is either using a different disposition, or is, more simply, adapting or adjusting their own

disposition slightly. Writers might, in this sense, either inhabit a character heavily, and think as them, or perhaps less dramatically, adjust their own disposition to think, in effect, what would I do in this character's position? (Heeks 2012, 252)

This also corresponds with the process of simulation discussed by Hogan (2013). Finding textual evidence for this in either analogue or keystroke logged genetic material would easily lead to speculative readings. Still, a nanogenetic analysis against this backdrop could lead to hypotheses that explain how the writing process *may* have proceeded, what *may* have led to some aspects of the story. The added value of keystroke logging data, in this case, is the data about the pauses the author makes during writing. This can indicate moments of simulation, in which the author stops writing to ask themself the question: what would the character do/think in this situation? As such, it may prove valuable for the re-enactment of "the author's cognitive process during the writing of their work" (Van Hulle 2022, 149).

This re-enactment in terms of the writer's disposition – or the process of inhabiting a character – can be illustrated with an example from Posthuma's writing process, in which she seems to have put herself in the character's shoes. Posthuma, in Session 17 (22 September 2020), writes a passage in which the I-narrator becomes irritated because technology is getting in the way of her making a good Instagram story. Angrily, she throws away her phone: "Flikker op" – 'flikker' is a Dutch swear word for homosexual men. The I-narrator then questions whether the word 'flikker' is offensive in this context:

> [135]Flikker op, zeg ik, en vraag me af of ~~ik in die con~~[136] [137]het woord flikker in deze context <u>kwetsend is.</u>[138]~~wel kan.~~[139] [140]~~Ik denk van niet~~[141] [142]Als ik het wil opzoeken moet ik opstaan en mijn telefoon[145] <u>gaan halen</u>~~.~~[149]<u>, maar om te kalmeren kan ik beter blijven zitten.</u>[150] (Session 17)

The process of inhabiting a character can *hypothetically* be made visible in a nanogenetic analysis. Before Posthuma started writing the sentence that contains the swear word, she paused for over 38 seconds. During this time, she could have planned this reaction to the annoyance of the I-narrator, asking herself the question 'What would the character say after tossing away her phone?'. In Posthuma's notes, the narrator is characterised as 'being angry all the time', therefore a swear word is a logical option. By using 'flikker', she created to opportunity for the politically correct reflection whether it is offensive in this context. Trying to be politically correct is something that characterises the I-narrator, although she is sometimes a bit inept at it. The swear word and the reflection therefore fit the image of the character Posthuma is trying to create, and she could well have put herself in the shoes of the first-person narrator to come up with this utterance.

Another possibility is that Posthuma merely 'discovered' the reflection of the word 'flikker' once she had written this word. Heeks point out that reflection can come "in the brief moment or pause after writing a word or sentence" (2012, 245). She paused for 3,2 seconds after writing "Flikker op, zeg ik", and after the second "ik" (3,6 seconds). She also paused for 9 seconds before "wel kan", perhaps doubting about how to formulate the reflection. Eventually, she is not fully content with the expression and substitutes "wel kan" with "kwetsend is", one minute after finishing the sentence. First, she writes "kwetsend is" in between "context" and "wel", after 7 seconds she relocates the cursor and deletes the redundant part of the sentence. She then immediately writes "Ik denk het niet".[235]

This second sentence could also be interpreted as being written from the character's disposition. Posthuma had earlier described how the I-narrator hated herself because she said something naive about the Dutch slavery history: she said, "I think" ("Volgens mij") after she said that slave owners saw their enslaved people as animals. 16 seconds after writing "Ik denk het niet", Posthuma deletes it and replaces it with a sentence that expresses the narrator's intention to fact-check it on her phone. From the perspective of 'what would the character do in this situation?', the revision seems logical. Since the I-narrator was annoyed by her insecure reaction and her overall awareness to being politically correct, making sure that she has the right answer to her question, instead of waving it off with "I don't think so", seems adequate. Although such a nanogenetic analysis does not immediately prove that authors indeed adjust their own disposition in writing characters, such a framework provides guidance for hypotheses about the writing process. The theory about disposition gives us guidance on how to interpret writing actions.

6.2.3 Focalisation

Van Rijswijk: Emphasising the child's perspective

Focalisation "refers to the relation between that which is focalized – the characters, actions, and objects offered to the reader – and the focalizer, the agent who perceives and who therefore determines what is presented to the reader" (Herman and Vervaeck 2019, 77). One question that should be asked with regard to the type of focalisation is what the position is of the focaliser in relation to the fictional universe: "If the focalizers belong to it, they are internal; if they remain outside of it, they are external" (Herman and Vervaeck 2019, 78). All the writing processes in this study have to do with writing a literary text in first-person narration, but as Herman and Vervaeck point out, "[e]ven if character and narrator coincide in a first-person text, there still exists a difference between internal and external focalization" (2019, 80). This is the case when a narrating

---

[235] Translation: "I don't think so".

I (the I-now) looks back to something the experiencing I (the I-then) did: when the scene is perceived by the narrating I, then it is a matter of external focalisation, and while it is internal focalisation when it is perceived by the experiencing I (Herman and Vervaeck 2019, 80).

An example of this can be found in a passage that Van Rijswijk wrote at the start of the writing process of "Zorgvlied" (Session 0; 3 December 2020), but that was deleted along the way. Here the narrating I describes one of her most vivid memories, in which she lies in a white (hospital) bed and a doctor in protective clothing walks around her.

> [75]Een van mijn helderste herinneringen was oo~~t~~ [76][77]it die van mijn ouders achter een ruit. [78]Ze zijn jong, zo'n tien jaar jonge~~n~~ [79][80]r dan ik nu ben, en ze zwaaien. [81]Ik lig in een groot wit bed en er loopt een volledig ingepakte arts om me heen. [82]Er mag niemand in mi~~n bu~~ [83][84]jn buurt komen en alle deuren worden dichtgedaan. [85]Omdat mijn moeder heel vrolijk zwaait wordt de angst niet ondraaglijk. (Session 0)

In what can be considered as a first version of this memory, the first two sentences are externally focalised with regard to the childhood memory, as the experiencing I would – as a child – not have described her parents as "jong, zo'n tien jaar jonger dan ik nu ben".[236] But with "en ze zwaaien", the focalisation starts to shift to internal focalisation.[237] It is the experiencing I who perceives that she is lying in a large white bed with a doctor walking around her and that nobody is allowed to come near her as all doors are shut. In the last sentence, the focalisation gradually shifts to external focalisation again, evaluating the experience now, the narrating I concludes that the fear did not become unbearable because her mother was cheerfully waving. This was most likely also the experience of the experiencing I, but it is formulated by the narrating I – "ondraaglijk" is a word a child would not use so quickly.

As the description of the parents already suggests, textual indications can help with establishing the types of focalisation (Herman and Vervaeck 2019, 85). Together with, among others, descriptions of focalised objects or people, linguistic features, words of modality and time indications, style can provide signs regarding the focaliser: "Childhood memories with many complicated and technical observations are probably externally focalized because a child would not achieve such intricacy" (Herman and Vervaeck 2009, 86). The first version already contained some of these indications of external focalisation. But 30 seconds after Van Rijswijk had written this first version – the pause time could indicate the process of re-reading, of interacting with the text

---

[236] Translation: "young, about ten years younger than I am now".
[237] Translation: "and they wave".

produced so far – she began to elaborate the passage with sentences that are internally focalised.

> [75]Een van mijn helderste herinneringen was oot [76][77]it die van mijn ouders achter een ruit. [78]Ze zijn jong, zo'n tien jaar jongen[79][80]r dan ik nu ben, en ze zwaaien. [81]Ik lig in een groot wit bed en er loopt een volledig ingepakte arts om me heen. [82]Er mag niemand in mi[86]n bu[83][84]jn buurt komen [87]en a[88]A[89]lle deuren worden[91] dichtgedaan[90]zin [93][92]jn geblokkeerd met zware balken[94]. [95]Sleutels zo groot als onderarmen worden omgedraaid. [96]Er worden ophaal[97][98]brugen [99][100]gen opgehald[101][102]ald en luiken naar beneden gerold*. [103]De kamer is heel donker maar ik word beschenen door een felle lamp.[104] en bepoteld door een [106][105]de tentakel van een machine.[107] [85]Omdat mijn moeder heel vrolijk zwaait wordt de angst niet ondraaglijk. [108]Daar houdt de herinnering op. (Session 0)

All the new sentences and revisions emphasise a child's perspective, and therefore an internal focalisation as well. The first revision is a substitution: changing "deuren worden dichtgedaan"[238] to "deuren zijn geblokkeerd met zware balken".[239] This revision marks the beginning of the dramatisation of the events. The sentences added subsequently state that keys the size of underarms are turned, then bridges are raised, and shutters lowered, in the dark she is illuminated by a bright lamp and pelted by the tentacle of a machine. The memory is now almost like a nightmare and underlines what hospitalisation would have been like for a child. Van Rijswijk's dramatic treatment of this passage makes the reader aware of how traumatic this event was for the character, which brings the possibility of extending the story to include the first-person narrator's fear of hospitals and doctors. Even though this passage did not make it into the published version, this example shows that the genetically informed reader can be made aware of those textual clues that provide information about focalisation, and that these clues may not have been there in the first place but were implemented during an interaction with the text produced so far. The temporal dimension of writing, in this case, brings the textual clues of focalisation to the fore.

Van Pelt: Empathic focalisation, or not?

Herman and Vervaeck further state that focalisers "can be specified with reference to a number of properties", which are spatiotemporal properties, emotional properties and ideological properties (2019, 82). On the emotional level, two types of focalisation can be distinguished: detached and empathic. The relation between the focalised object and the focaliser is important in establishing this emotional property of focalisation (Herman and Vervaeck 2019, 84). The focalisation is detached when "only the outside

---

[238] Translation: "doors are closed".
[239] Translation: "doors are barricaded with heavy beams".

of the focalized object is perceived"; it is empathic when "there is constant speculation about the thoughts and feelings of the focalized object" (84). In the writing process of "Dauphin" there is a clear example in which Van Pelt revised a passage from a clear empathic focalisation to a rather detached focalisation. During Session 15 (19 December 2020) Van Pelt elaborates a scene describing the lunch on Flamingo-Island. The son has just told Mustafa that his father has died, after which Mustafa puts his hand on the son's hair and leaves the hut: "Mustafa staat recht, legt zijn hand even op Finns haar en verdwijnt dan uit de hut". Van Pelt starts to revise this sentence by writing new sentences within it. First, she adds that Mustafa looks the I-narrator straight in the eyes. The I-narrator then states that she sees what he thinks ("Ik zie wat hij denkt"), which is an indication of empathic focalization. What she thinks he sees – a speculation about the thoughts of Mustafa – is that she and Finn are a ship, who have lost their anchor, who sail on without keeping course, and that they cannot dock anywhere. However, what is actually described here is that the I-narrator is projecting her own feelings on Mustafa, which serves to establish the characterization of the I-narrator. Then Van Pelt deletes this expression of emphatic focalization and replaces it with a phrase that states that the I-narrator wants to tell Mustafa this, but she does not.

> [283]~~Finn en ik, w~~[284] Mustafa kijkt me recht in de ogen.[285] [286]~~Ik zie wat hij denkt.~~ [300] [287]~~Dat~~ We [301][302] ~~we~~ [304][303] ~~Finn en ik een schip zijn, zou~~ [305][306] Ik zou hem willen vertellen dat [307]Finn en ik een schip zijn, ~~een schip z~~[288] [289]~~dat~~[290] dat we ons anker kwijt ~~kr~~[309][308] raakten.[312][310] [291]~~we raakten ons anker kwijt~~ ~~en varen nu~~ [202][311][293], dat [313]we blijven varen~~, houden koers,~~[294] zonder koers te ~~varen~~[296][295]houden,dat [314]we niet [315]kunnen ~~niet meer~~[316][297] ~~maar kunnen nergens aan~~[298]aan[299]meren, ook hier niet[317]. ~~staat recht,~~ [318] ~~Mu~~[320][319]stafa knikt naar me, [321]legt zijn hand even op Finns haar en verdwijnt dan uit de hut. (Session 15)

Although the sentence contributes to the characterisation of the narrator, the emphatic focalisation of Mustafa does not coincide with how the I-narrator perceived Mustafa in the rest of the text produced so far. Thus far, the focalisation was rather detached: descriptions of Mustafa were limited to his outer appearance, he wears, for example, dark clothing and sunglasses, and when he finally takes off his sunglasses, the I-narrator describes his eyes as 'dark and cold'. The sudden emphatic expression – a personal projection – does not seem to be in accordance with the context. From a text genetic perspective, this example highlights that focalization does not need to be stable during the writing process, authors try different expressions, and their decisions are constantly affected by the text produced so far.

6.3 NARRATION

The last, and least abstract, level of structuralist narratology is narration. As Herman and Vervaeck point out, narration "is concerned with formulations – the entire set of ways in which a story is actually told" (Herman and Vervaeck 2019, 87). In fact, it denotes the concrete words and sentences that are presented to the reader, and the wording of the events (87). Narration can, according to Herman and Vervaeck, be subdivided into narrating (including the narrating agents) and the way in which these agents present a character's consciousness. In this section, I show how the keystroke logging data can reveal the choices made with respect to these aspects.

6.3.1 Narrating

Bogaert: Unexpected narratee

Whereas focalization expresses the relationship between the object and subject of perception, narration expresses "the relationship between the narrator and that which is narrated" (Herman and Vervaeck 2019, 87). An extradiegetic narrator hovers, as it were, over the narrated world, while an intradiegetic narrator "belongs to the narrated world and is therefore below another narrating agency" (87). In *Roosevelt*, Franklin is an extradiegetic narrator: he knows all too well what is going on in the minds of the other characters, but in the final chapter of the novel, "Terminus", Franklin admits that all the life stories he has previously described have been invented by him. On page 12 in the Atoma notebook, where Bogaert started to address the structure of the novel, he wrote: "In de epiloog wordt alles op de helling gezet!", which means that in the epilogue everything is being revised or rectified. This indicates that the plan for this guiding principle of the novel – Franklin as an unreliable extradiegetic narrator – was established quite early in the writing process.

But this is not the only thing the final chapter 'reveals' about the narration. The last sentence in the published novel – "Ik heb zelfs jou verzonnen" (Bogaert 2016, 206)[240] – again calls everything into question. Apparently, it all took place in Franklin's head. This sentence was written into the Word document during Session 89:

> ₂Want zie ~~je~~: ik heb zelfs jo~~p~~u₄ ~~(w verhaal)~~ verzo~~i~~nnen₈ (Session 89)

With this sentence, Franklin reveals a specific truth about the novel: that Carla was actually the narratee. Herman and Vervaeck point out that intradiegetic narrators mostly speak to other characters, and extradiegetic narrators mostly to an extradiegetic narratee, but that "cross bonds are possible" (2019, 80). An extradiegetic narratee is not the same as the empirical reader, although it may appear as such, "but rather an agent who does

---

[240] Translation: "I even made you up".
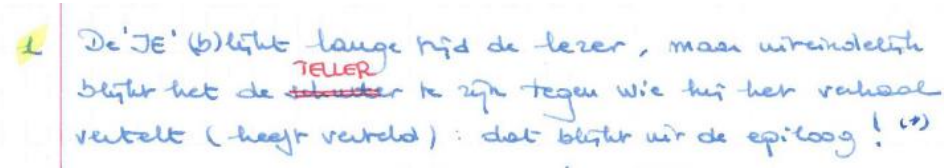
Figure 6.3: Detail of the Atoma notebook: idea about who is addressed with 'you' in the novel.

not appear in the story and yet functions as the narratee" (89). While reading the novel, the reader is led to believe that Franklin is addressing an extradiegetic narratee when he uses the second person narration: 'you'. During the start of the writing process, Bogaert gradually developed this idea. First, he considered an option where it seems that the 'you' is the reader, but it eventually appears to be the shooter: "De 'JE' (b)lijkt lange tijd de lezer, maar uiteindelijk blijkt het de schutter te zijn tegen wie hij het verhaal vertelt (heeft verteld): dat blijkt uit de epiloog!" (6).[241] Later, he changed the shooter into 'the counter', which is Carla.

The genetic material also shows a dead-end regarding Carla, the girl who counts the tram passengers, and what the last chapter reveals about her. Carla's main secret is that she was raped the day before, but she is afraid to tell anyone (e.g. Atoma notebook page 16). Gradually, after seeing Remy pass by on the tram a few times, she begins to believe that in was in fact Remy who raped her. When Bogaert wrote this idea in the Atoma notebook, he added that Franklin will suggest in the epilogue that she made up the rape: "(Op het eind wordt gesuggereerd ^EPILOOG: door Roosevelt^ dat ze de verkrachting heeft verzonnen om ^een keer^ begeerd te worden!) Want: ze heeft nog nooit de liefde gekend!" (88).[242] According to Franklin, she never knew love, so she made up the story so she could be 'desired' once.

This idea is included in the Word document in Session 264. During this session, Bogaert inserted a new version of one section narrated by Carla. In this section, she says that she has been raped. At the top of this section, Bogaert wrote a note: "In de epiloog suggereert Roosevelt dat ze de verkrachting heeft verzonnen, misschien om te voelen wat het is begeerd te worden".[243] In the same session, Bogaert decided to directly insert this idea into the epilogue:

> En jij? [31]S[32] [33]Voel je no[34]g wat[35] in je buik? [36]N~~B~~[37] [38]atuurlijk nie[39]t[40]. [41]Je wilde alleen maar [62]begeerd worden. [46]Niks meer of n[47]iks ~~l~~[48]minder~~f~~[49][50].

---

[241] Translation: "The 'JE' seems to be the reader for a long time, but in the end it turns out to be the shooter to whom he tells (has told) the story: this is revealed in the epilogue!"

[242] Translation: "(At the end, it is suggested ^EPILOGUE: by Roosevelt^ that she made up the rape to be desired ^one time^!) Because: she has never known love!"

[243] Translation: "In the epilogue, Roosevelt suggests that she made up the rape, perhaps to feel what it is like to be desired".

~~[42]Z[43]z[44]o was het~~[45]. [51]~~Dee~~[64]~~e~~[65][63]e rest is in je hoofd gebeurd.[66] [52]~~Je hebt jezelf~~[53]~~lf voo~~[54] [55]~~wat in het hoofd gehaald.~~[67] (Session 264)

The sentence "Voel jij nog wat in je buik"[244] refers to Carla saying that she has abdominal pain since yesterday: "Mijn buik doet pijn. Sinds gisteren" (Bogaert 2016, 15).[245] Over the course of reading the novel, we have come to know that this is because she has been raped. In Session 347, this paragraph is extended with two sentences. According to Franklin, she just needed a little love, just like everybody else. Franklin, with his 150 years of age, expresses a few sexist opinions throughout the novel, such as pointing at the short skirts of the women and how men watch them. He even says that it is no wonder people rape: "Tegenwoordig verhullen vrouwen nauwelijks nog wat. Bloot en spelen. Geen wonder dat er wordt verkracht" (Bogaert 2016, 72).[246] In this way, his paternalistic and sexist remark fits Franklin's character. Franklin also states that he knows that everyone only thinks of sex.

In one of Remy's sections, Remy's thoughts wander to the girl he saw standing at the tram stop (Carla) and he decides to wave to her, or maybe even talk to her for a moment. Then, Franklin makes himself present as an extradiegetic narrator, who can decide on the unfolding of the narrative. He corrects Remy, because according to Franklin, Remy does not *only* want to talk: "Hij wil niks anders dan seks. Alle mannen willen niks anders dan seks. Ze vinden het de beste remedie tegen hun eenzaamheid" (Bogaert 2016, 188).[247] Therefore, Franklin lets Remy start over: "Ik laat hem dat laatste stuk opnieuw vertellen" (187).[248] Now, Remy indeed narrates about his sexual fantasies. Bogaert planned this interruption in the notebook, and worked on the textualisation between Sessions 345 and 392 (with the primary focus in Sessions 387 to 392). In Session 366, Bogaert also elaborated the derogatory remarks about Carla. Franklin thinks that Carla probably does not understand it that well and that she has failed to tell a good lie:

> [51]Misschien begrijp je het zelf ale[52]lemaal wel niet zo goed. [37]Je hebt het in elk geval [48]~~alleen~~ [49]niet zo [50]goed verzonnen. [38]Het is zoals de ~~to~~[39] [40](oude) toneelauteu[41]r[42]r [43]/schrijver zei/schreef: je hebt talent nodig om te liegen. (Session 366)

---

[244] Translation: "Do you still feel something in your belly".

[245] Translation: "My belly hurts. Since yesterday"

[246] Translation: "These days, women barely conceal anything. Nude and play. No wonder people get raped".

[247] Translation: "He wants nothing but sex. All men want nothing but sex. They think it is the best cure for their loneliness".

[248] Translation: "I'll let him retell that last part".

In Session 405, Franklin's remark that Carla made it all up and only wanted to be desired is deleted from the document. Now that all sexual references are included in the text produced so far, it could be that Bogaert decided that this was too much, and that even a hundred- and fifty-year-old square cannot accuse a girl of making up a rape – even though she also sprung from Franklin's imagination. And if it really is the case that all men just want to have sex, as Franklin suggests, then some could do harm. Given the context, Carla's rape would not be so implausible, although the novel does not accuse Remy. In sum, the text produced so far did not seem to fit with the revelations in the last chapter, or in other words: Bogaert 'corrected' the statements of his extradiegetic narrator.

As Heeks indicated, the text produced so far can replace the original plan (Heeks 2012, 246). This is further underlined by Bogaert deleting the original plan that Carla made up that she was raped. Even Bogaert, who attached great value to his notes, and had no intention of deviating from them, may still come to the point of making an adaptation to the initial plan. Hayes and Nash have pointed out that writers "appear to use the text written so far to remind them of the constraints imposed by what has already been written" (Hayes and Nash 1996, 41). The text produced so far thus becomes a major point of reference in the task environment and "influences the planning of the next text segment" (41). This example about Carla, however, also proves that as the text produced so far is expanded over time, this can also lead to changes in the ideas of the initial plan that were already incorporated in the text produced so far.

Posthuma: Changing person

With regard to narration, Van Hulle observed that "[a] change that occurs quite often in the course of literary geneses is a shift from first-person to third-person narration, or vice versa" (Van Hulle 2022, 157). I would add that this also remains the case for the keystroke-logged genetic dossiers. As noted in Chapter 4.3, Posthuma introduced a change in narration from the third person to the first person. While Cohn could only hypothesise the moment in which Kafka introduced the shift from a first-person to a third person narration in *Das Schloss*, we now have all the data at hand to investigate the moment Posthuma changed her text to first-person narration, and what might have incentivised this change.

Posthuma began planning and writing her story in third-person narration, but she soon started experimenting with the first-person: the first time was in Session 3, and we can trace exactly *when* in the writing session this change occurred. Posthuma starts the third writing session, producing new text in multiple locations, all written in the third-person narration. At a certain point in this session, Posthuma writes about the female friend, who grew up by the sea and returned every year. The day before, she had made pasta with squid according to an old family recipe and there had been a

misunderstanding about something Posthuma still has to decide on or, at least, write down. But instead of finishing that sentence, she stops writing for over ten minutes, and the first thing she does when she starts writing again is changing the text to first-person narration from top to bottom. This means that she changes all the instances of "ze" ["she"] into "Ik" ["I"], or "haar" ["her"] into "mijn" or "me" as well as the conjugation of the verbs:

> ~~Ze~~[100] Ik[101] hou~~dt~~[102] van de zon […]. (Session 3)

In some cases, she only had to delete the introductory expression, like: "Ze denkt".[249] There is also a revision that is directly related to this global change from third to first person narration: "die een zoon hebben van dezelfde leeftijd als de hare"[250] is changed into "die een zoon van dezelfde leeftijd hebben" (n130);[251] because "als de hare" had to be deleted, the syntax of the sentences also needed to be changed. Then there are two revisions that were made during the change from third to first-person narration. She deletes the clause "schaduw is onder de overhangende bomen tussen het smalle strand en de weg" (n106)[252] because this was a left-over of a previous deletion, and she deletes "van hun vrienden" (n133)[253] about the son of the friends of the main character, because this was already mentioned in the previous sentence. These revisions were thus made during the same revision episode but were unrelated to the global change.

Not all the sentences are matched to the third-person narration yet, but Posthuma already continues finishing the unfinished sentence about the misunderstanding. She adds that the husband of the now first-person narrator and the friend disagree on whether you should defrost the squid before putting it in the pan. Then Posthuma ends the session.

This shift was probably related to the difficulty Posthuma experienced writing about multiple 'she's' in the story, especially now that she was writing about the female friend. Here, the first two 'she's' – "zij" – are clearly referring to the friend, but for the third – "ze" – it is less clear. It could either be the friend, or the main character. This ambiguity could easily be fixed, by replacing it with a name or with 'the friend', for example. But since she pauses for ten minutes and then opts for the first-person narration, we may also hypothesise that the problem with the multiple 'she's' could be the trigger for Posthuma to start experimenting with a first-person narration, since it also fixes the issue of the ambiguity.

---

[249] Translation: "She thinks"
[250] Translation: "who have a son the same age as hers".
[251] Translation: "die een zoon van dezelfde leeftijd hebben".
[252] Translation: "shade under the overhanging trees between the narrow beach and the road".
[253] Translation: "of their friends".

In the next writing session – Session 4 – that takes place over two weeks later, Posthuma starts with writing in the first-person. The husband is defrosting the squid, and Posthuma adds two new sentences to elaborate this scene: "Maar het moet bevroren de pan in, zei ik. Wil jij het doen? vroeg hij" (n2-7).[254]

One and a half minute later, she relocates the cursor to the top of the document, to change the text back to third-person narration, which again requires a change in conjugation. To make a clearer distinction between the 'she' referring to the main character, and 'she' referring to the friend, Posthuma changes a couple of 'she's to 'the friend'. Again, we can spot revisions that are performed during the same revision episode, but that are not related to the global change in narration, such as in this sentence:

> ~~Ik~~[16]Ze[17] fronst[18] naar de ~~bruine~~ [19][20]z~~p~~[21][22]onnende [23]vrouw in de kleine bikini, die ~~op haar rug in de zon ligt en~~ [24]zich net loom heeft opgericht om de kleur van haar buik te inspecteren. (Session 4)

Moreover, the change from first to third-person narration is also interrupted by new text production: Posthuma writes a new paragraph about the main character disliking the body of Miranda July. She also locks her computer for half an hour in the meantime. As pointed out earlier, Posthuma alternates during this session between new text production, revision thereof and the global revision of the change of person.

The text written prior to this change does not seem to point directly towards a trigger which made her change the text back to third-person, although it could be the writing of the dialogue. Posthuma could have disliked writing "zei ik". In general, it shows that she was not immediately convinced to write in the first person. Her initial idea was to write in the third person narration, her notes were also written from this perspective, and in terms to the story, the third-person narration perhaps emphasises a distant self-reflection, as if the main character experiences herself constantly through the eyes of others.

For a couple of sessions, Posthuma continues writing in third-person narration. But four sessions or seven days later, Posthuma changes the story again to the first-person narration. This time, she started the session with these changes, which makes it more difficult to deduce a hypothesis about what exactly triggered her to make this change yet again. But since she made this change to first-person narration earlier in the writing process, she might have kept this possibility in the back of her mind, allowing her to reflect on this while she continued writing. What may have steered Posthuma to eventually opt for first-person narration is that the story is about the self-reflection, or rather, self-critique of the characters. Written from this perspective, she could represent

---

[254] Translation: "But it has to be put in the pan frozen, I said. Do you want to do it? he asked".

the main character's worldview more naturally and write what the main character is thinking.

While we must admit that we cannot be certain of the real reason for the change in person, we do have answers to questions that we did not have for the shift in person in Kafka's *Das Schloss*. For the first shift from third to first person narration, we could see that the revisions – "she" substituted with "I" – were made before Posthuma wrote directly in the first-person. The change was thus not the result of a sudden insight during writing, as she did not write immediately in first-person and then adjust the other parts of the text. Moreover, we could also observe that she was pausing for over ten minutes before she made the change, which could indicate that she was re-reading her text or reflecting on it. Also, we could see revisions that are not directly related to the shifts in narration, but that were performed while making this change – these revisions were thus triggered by rereading the text.

6.3.2 Consciousness representation

Van Rijswijk: 'Real' and 'unreal' conversations

The representation of consciousness, which is "the way in which the narrator renders the consciousness of the characters" is as essential as the narrating agent (Herman and Vervaeck 2019, 99). Dorrit Cohn distinguishes three kinds of consciousness representation. The first is psycho-narration, which corresponds to indirect speech. This indirect diegetic representation is generally called "summary" or "telling" (Herman and Vervaeck 2019, 99). The second, quoted monologue, corresponds to direct speech. This direct mimetic representation is often referred to as "scene" or "showing" (99). These two represent the two poles in grammatical means, while the third can be seen as an intermediary form. This third from, narrated monologue, corresponds to free indirect speech (99). Yet, as Herman and Vervaeck point out, there are certainly sentences that "cannot unambiguously be put in a single category" (100). Therefore, Brian McHale has put the intermediary forms on a sliding scale. From the most diegetic to the most mimetic, this scale contains: the diegetic summary, the less purely diegetic summary, the indirect content paraphrase, semimimetic indirect discourse, free indirect discourse, direct quotation, and free direct discourse. Nonetheless, textual elements can also be ambiguous, which is why Herman and Vervaeck emphasise that consciousness representation is also "the work of the reader trying to imagine the original version of a represented thought or utterance" (101).

Regarding consciousness representation, two levels of quoted monologue – or direct quotation – seem to be present in "Zorgvlied", as the story offers two kinds of dialogue to the reader. The first dialogue is the one the first-person narrator has with her mother while they are walking in the cemetery. They are soon joined by the ghosts of their (female) ancestors, but the mother cannot perceive them. The second dialogue

is that between the I-narrator and the ghosts, mostly the ghost of the grandmother. The conversation only takes place between the I-narrator and the ghosts, though they sometimes react to something the mother says. While the first dialogue is given in quotation marks – which makes it seem reliable – the second is not. Therefore, the reader is made aware that the conversation between the I-narrator and the ghosts only takes place inside the head of the I-narrator. The first conversation is 'real', the second not so much.

Until the third writing session (Session 2; 8 December 2020), the dialogue between the I-narrator and her mother was also not given in quotation marks and the keystroke logging material indicates that the need to insert quotation marks was born in the moment the conversation partners – the mother and the ghosts – are simultaneously present in the fictive present. Van Rijswijk did not include the conversations with the ghosts until Session 2, or rather, the idea of the ghosts in its entirety is still absent. This changes at the start of Session 2, when Van Rijswijk includes a passage in which the I-narrator describes a conversation she once had with 'the dead'.

> [4]Ik heb aan de doden gevraagd of ze dat niet vervelend vonden ~~maar omdat ze in gestorven vorm~~ [6]waarop ik wat korzelig terugkreeg dat zij het ook heus niet de hele tijd over de levend[n][8]en hadden. ~~Waaro~~[10]
> [11]Wa[r][12][13]arover dan wel, vroeg ik.
> [14]Din[e][15][16]gen die jij niet begrijpt, zeiden ze. (Session 2)

The first sentence can be interpreted as belonging to the less purely diegetic summery, which shows "some of the content without representing it faithfully" (Herman and Vervaeck 2019, 100). But it seems more likely that it is a form of indirect content paraphrase, "which represents the thoughts or utterances faithfully as far as content is concerned but not in terms of style" (100). The next two sentences share some important characteristics of direct speech in a dialogue: each speaker gets a new paragraph; each paragraph is indented, and dialogue tags are used to indicate which character is speaking – just like the dialogue lines between the first-person narrator and her mother. The two dialogues therefore take place at the same level, and because the dialogue with the ghosts took place in the past, it does not (yet) bring along problems for the interpretation of this rather unusual conversation: it may well have happened and be 'real' in the story world.

Later in the writing session, however, Van Rijswijk writes that the first-person narrator hears footsteps and smells the perfume of her mother's mother.

> [189]Achter me ho[r][190][191]or ik ~~de kenmerkende tred van mijn oma, de moeder van mijn moeder,~~ [192][193]~~moeder~~[194][195]nieuwe voetstappen, rustig, en ik ruik het parfum van de moeder van mijn moeder. (Session 2)

Introducing the ghosts in the fictive present prompts Van Rijswijk, 17 seconds later, to insert the quotation marks around the utterances in the dialogue between the I-narrator and her mother. For this moment, the conversation between the I-narrator and 'the dead' could be interpreted to belong more or the less purely diegetic summary or the indirect content paraphrase: the I-narrator recaps some parts of the conversation more or less faithfully. Yet, the addition of the quotation marks seems to be Van Rijswijk's anticipation of a 'problem' she would face as she continued writing. She introduced the ghosts into the fictive present, but her initial idea was that only the I-narrator can perceive them. After she had inserted the quotation marks, Van Rijswijk writes that the mother of the I-narrator does not know that her mother is there. As a result, Van Rijswijk writes a new conversation between the I-narrator and the ghost of her grandmother without the quotation marks. This is directly followed by the re-writing of a dialogue between the I-narrator and her mother, which, again, have quotation marks. The ghost of the grandmother then reacts to this conversation, and her utterance is not placed inside quotation marks. As such, it becomes clear to the reader that there is one conversation, but that only the narrator can hear the words of both interlocutors. With the mother she speaks out loud, with her grandmother's ghost only in her head.

The genetic material thus makes clear how decisions, and solutions to problems of consciousness representation are made and found during the writing process. At first, Van Rijswijk most probably did not anticipate the issue that arises when writing conversations in which not every interlocutor can perceive everyone, but at the moment the 'problem' arose, she quickly found the simple solution by adding the quotation marks around the 'real' dialogue lines.

Van Pelt: The difference between saying and thinking

The representation of consciousness influences the interpretation of the characters as well. During the writing process of "Dauphin", Van Pelt made subtle references to the grief of the I-narrator – and the decisions about the representation of consciousness played a role in this process as well. The final version of "Dauphins" contains a passage about collecting seashells.

> 'Mogen ze in de kast van papa?' vraagt Finn.
> 'Ja, we geven ze een mooi plekje.'
> Thuis zal ik de schelpen schoonmaken en insmeren met olie om ze te doen glanzen. Finn zal me daarmee helpen. Op een identificatie kaartje zal ik naam

> en vindplaats noteren, de schelpen dan zorgvuldig opbergen. Ze zullen de lege plekken in de vitrinekast vullen. (Final version)[255]

In writing this short dialogue, Van Pelt first wrote "Thuis zullen we ze in houten bakjes leggen, met de juiste naam erbij" (n246-248), as a thought and plan of the I-narrator. Then she decides that it is actually the son who initially came up with this, as she inserts a question in direct quotation: "'Mogen ze in de bakjes van papa?'" (n250).[256] This question makes clear that collecting shells (and/or other collectibles) was something the father used to do, which gives the activity an additional layer of meaning. The answer of the I-narrator is then also given in direct quotation: "'Ja, er zijn nog lege plekken waar deze inpassen'" (n252-256).[257] Yet, Van Pelt then decides to remove this direct quotation of a rather practical answer and replaces it with "Ik knik" (n260-262).[258] This is followed by a description of what they will do with the shells when they are back home. For this, Van Pelt carefully researches online how best to clean, protect and store shells. The final sentence of this description then is: "Ze zullen de lege plekken in de vitrinekast vullen" (n312-333).[259] This sentence addresses again the absence of the father; he will not be able to fill the empty spaces in the cabinet, so now they will take on that task themselves. Since the I-narrator does not verbally answer her son, the description of what they will do with the shells can easily be seen as an indirect content paraphrase: it sums up what the narrator is mentally planning in that moment. This non-verbal description, especially the last sentence, contributes to the characterisation of the mother, as she tries to hide her grief from her son. The son, on the other hand, is more openly referring to his father. For him, collecting the shells and storing them in his father's cabinet is a way to cope with his grief. The deletion of the direct quotation of the mother, in sum, emphasises the internal struggle of the mother as the indirect content paraphrase represent her internal thoughts, not a conversation with her son.

In Session 12, Van Pelt – incidentally or not – does not open the document with the version of the previous session, meaning that all the text production and revisions from Session 11 are undone. The passage returns in Session 16, as Van Pelt searches for the previous version and copies the passage and pastes it into the current working document. The following session then actually introduces a change with regard to the consciousness representation in this passage: the non-verbal nodding is replaced by an

---

[255] Translation: "'Can they go in Daddy's cupboard?' asks Finn. 'Yes, we'll give them a nice place.' At home, I will clean the shells and rub them with oil to make them shine. Finn will help me with that. On an identification card, I will write down name and find location, then carefully store the shells. They will fill the empty spaces in the display case".

[256] Translation: "'Can they go in Daddy's trays?'"

[257] Translation: "'Yes, there are empty spaces where these fit in'".

[258] Translation: "I nod".

[259] Translation: "They will fill the empty spaces in the display case".

actual, verbal, reaction of the mother. The I-narrator now answers that they will give the shells a nice spot. Compared to the direct quotation that was written and removed in Session 11, this expression is still practical, but more caring as well. By referring to the 'nice spot', the I-narrator acknowledges the importance these shells have for her son, as it will bring him closer to his father.

> 'Ja, we ~~vinden vast~~ [170] [171]er ~~va[172][174]ee[173][175]zullen~~ [181]geven [182]ze een ~~mooi~~ [176] [177]mooi plekje~~s~~[178] [179] ~~geven~~[183].'
> ~~Ik knik.~~[180] Thuis ~~zullen we~~ [184] zal ik [185]de schelpen schoonmaken en insmeren met olie om ze te doen glanzen. [186]Finn zal me ~~helpen,~~[187] [188]daarmee helpen~~, hij~~[189] [190]. Op een identificatie kaartje zal ik de naam en vindplaats noteren, ~~hen~~[191] de sche~~ple~~[193][192]pen[194] dan zorgvuldig opbergen. ~~Deze schelpen~~ [195] Ze [196]zullen de lege plekken in de vitrinekast vullen. (Session 17)

Bringing back the direct quotation in the answer to the son's question makes it more ambiguous whether the indirect content paraphrase is a representation of thoughts or utterances. This therefore triggers revisions in the description that follows. By adding the sentence "Finn zal me daarmee helpen" (n186-190),[260] Van Pelt reduces the ambiguity. Naming the son in this sentence makes it less likely that this is an indirect content paraphrase of the utterances of the I-narrator to her son; rather, it is a representation of her thoughts. In this way, the interpretation of this passage remains the same. The formulations, and more specifically different consciousness representations, fluctuate during the writing process; a nanogenetic reading of this process emphasises this repeatedly.

6.4 CONCLUDING REMARKS

In César Aira's *The Literary Conference* (2010), the narrator – a famous author – discovers the secret of the "Macuto Line", one of the world's wonders: a braided cord on the coast of Venezuela that leads to a pirates' treasure. Nobody was ever able to figure out the secret of the rope that would retrieve the treasure, which would be attached to the rope, but located somewhere under water. That is, until the narrator touches the knot with the tip of his fingers and turns it over "without attempting to untie anything", which makes the treasure chest burst out of the sea, into the sky, to land just three feet from where he was standing (Aira 2010, 12). César's theory is not that he is a genius or exceptionally gifted, but that it was his unique configuration of experiences, memories, and skills, that made him the one required to solve the problem of the Macuto Line:

> It is not that I am a genius or exceptionally gifted, not by any means. Quite the contrary. What happened (I shall try to explain it) is that every mind is shaped

---

[260] Translation: "Finn will help me with that".

by its own experiences and memories and knowledge, and what makes it unique is the grand total and extremely personal nature of the collection of all the data that have made it what it is. Each person possesses a mind with powers that are, whether great or small, always unique, powers that belong to them and to them alone. This renders them capable of carrying out a feat, whether grandiose or banal, that only they could have carried out. (Aira 2010, 8)

In César's theory, every person has a unique mental configuration through the books they have read, the music they listened to, the films and paintings they have seen, their activities during the day and night, the things they have experienced, and so forth. This leads them to achieve things only they can do. This theory for solving the Macuto Line can also be read metaphorically to describe the literary writing process: the unique configuration of the author's minds – with all their experiences – lead them to write stories only they could have written; only they can solve the secret of the line that leads them to the treasure of ideas. And with the nanogenesis, we can possibly uncover some units in their mental configuration that were activated during writing.

Through a (nano)genetic narratological analysis we can reconstruct the decisions that were made about details of the story, narrative and narration, and the relations between these details on a very fine-grained level. These decisions can be made suddenly, or through progressive insight. Thus, whereas with analogue manuscripts we have to hypothesise even about the order in which the writing took place, we can now take it a step further and hypothesise about what might have triggered certain revisions and the invention of new directions in the story. In a sense, nanogenetic narratology reveals how every author comes to write a unique text, with connections only they could have made, because the story sprung from their unique 'mental configuration'.

With a nanogenetic analysis of the five writing processes through a narratological lens, several observations can be made regarding the story (events, actants, and setting), narrative (time, character, and focalisation), and narration (narration and consciousness representation). It should be noted, however, that these observations are specific to the writing of these texts.

For the events, the keystroke logging data allowed an analysis of how Bogaert searched for the cause of the effect that compelled Faraaz's family to flee Iran. Through the nanogenesis, we could see that the changes that focused on the mother's role may have also triggered the idea of making the mother's family the reason they had to flee. Within Van Pelt's writing process, on the other hand, we were able to reconstruct how a pure index – the reading of a specific book – was incorporated into the text and how it became more prominent in the story. In terms of actants in "Dauphin", nanogenesis provides information about how Mustafa's role as an actant gradually crystallised during the writing process. In particular, the immediate revisions show how Van Pelt was more explicit about Mustafa's role in the initial composition of these ideas, but that she

reduced it as she wrote. Posthuma's writing process revealed how a change in the role of an actor can also stimulate ideas in the production of new text: reducing the role of the first-person narrator in her own goal of achieving body positivity created the opportunity to implement reflection on why people are reliant on positive affirmation from others to accept themselves. Bogaert had the difficult task of conveying the message that the setting was also the narrator. The nanogenesis and session versions showed how Bogaert gradually reduced the explicit references to Franklin's 'square-ness', opting instead for more subtle references to the actual Franklin Roosevelt square. The keystroke logging data also showed the evolution of the setting in Troch's "Mondini". By interacting with the text produced so far, he was able to find a way to balance the impact of climate change on the world and people's behaviour and worldview. After a process of immediate revision, Troch finally decided the human violence was responsible for the deaths of the narrator's partner and daughter.

With regard to time, the example from the writing process of Bogaert showed how an idea for the connection between the fictive past and present was actually triggered during the act of writing. As such it shows how interaction with the text produced so far can lead to the discovery of new connections and ideas. The genesis of Troch's "Mondini" showed how the distance between analepses and the fictive present was actually fluid during the writing process. In Van Rijswijk's writing process, we could see a repeated shift between 'telling' and 'showing'. The nanogenetic analysis showed how Van Rijswijk's initial formulations were more explicit about the 'malfunctions' of the first-person narrator's brain, but that she deleted these to leave room for the interpretation of the reader about the question of whether the narrator can actually perceive ghosts. Also, with respect to character, an example from Posthuma's writing process illustrated how the theory of the author's disposition can lead to hypotheses about the author's cognitive process during writing. Posthuma wrote a passage in which the narrator wonders about whether the word "flikker" can still be used in specific contexts. Through a nanogenetic analysis that also takes the pauses into consideration, I hypothesise that in writing this passage Posthuma inhabited the character's disposition. In this example, her own disposition and that of the character may not be that far apart, but based on the text, a reconstruction of the character's disposition is somewhat less speculative. Focusing on focalisation, we could see how Van Rijswijk made revisions to emphasise the internal focalisation in a childhood memory by adding textual indications that signalled a child's perspective. And an example from Van Pelt's writing process indicated that the focalisation is not always stable during the writing process, as she changed a sentence from empathic to detached focalisation.

In *Roosevelt*, Franklin is an extradiegetic narrator and in the last chapter it is revealed that Franklin has made up all of the stories that were told in the preceding chapters. This was already part of the initial plan, but the genetic material showed that this initial

plan included even more 'revelations' in the final chapter. Franklin would state that Carla made up that she was raped. However, after writing the sections narrated by Remy, in which Franklin also leaves his mark as an extradiegetic narrator, Bogaert decides to delete this 'accusation' from the final chapter. The genetic dossier thus shows that even though Bogaert is an extensive planner, he also deviated from the initial plan. The nanogenesis provides us with more information about the shift in narration in "En daarom haten ze zichzelf". Posthuma started writing her story in third person but decided to experiment with the first person after she had to write about multiple 'she's' in the story. Although she changed the narration back to the third person once again, she eventually opted for the first person. This choice may be motivated by the fact that writing from the first person would allow her to write about the main character's thoughts more naturally. With a focus on consciousness representation, we could see that Van Rijswijk used two levels of quoted monologue or direct quotation in "Zorgvlied". One with quotation marks to indicate the 'real' dialogue the first-person narrator has with her mother, and one without quotation marks to indicate her conversation with the ghosts. The nanogenetic analysis pointed to the moment in which Van Rijswijk found the need to include quotation marks: the moment she started to write a conversation in which not every interlocutor could perceive everyone. The representation of consciousness also influences the interpretation of the characters. During the writing process of "Dauphin", Van Pelt made subtle references to the grief the first-person narrator experiences, which also becomes visible in the decisions about the representation of consciousness. The example indicated how the deletion of the direct quotation of the narrator emphasised her internal struggle. More generally, the examples show how formulations, and more specifically different consciousness representations, fluctuate during the writing process, a nanogenetic reading of this process emphasises this over and over again.

Although it is difficult to make generalisations about the implementation of these narratological aspects in the text, one commonality can be observed. In some of the examples discussed above, we can notice the tendency of the author to make the narratological aspects less explicit as the writing process proceeds. Van Pelt was initially more explicit about Mustafa's role as saviour (she even tended to describe him as a prince on a white horse); Bogaert reduced the explicit references to features of squares in his characterisation of Franklin; and Van Rijswijk referred less explicitly to the 'malfunctions' of the narrator's brain. This may be related to the 'show don't tell' recommendation in creative writing, but what these writing processes show, is that writers sometimes need to *tell* and be explicit, however briefly, for their own understanding of the text. In sum, the nanogenesis thus provides clues to understanding the choices made regarding these narratological aspects. That is what we are able to

uncover through nanogenetic narratology: all the dozens of (nano-)choices that the author has to make during the writing process.

CHAPTER 7. CONCLUSION

Y2K FOR GENETIC CRITICISM

"Pay no attention to the neatly formatted and deceptively typo-free surfaces of the average Microsoft Word file. […] The story of writing in the digital age is every bit as messy as the ink-stained rags that would have littered Gutenberg's print shop or the hot molten lead of the Linotype machine" (Kirschenbaum quoted in Schuessler 2011)

Recall the anecdote from the introduction, in which Maggie Nelson is being assisted by two machines while writing *The Argonauts*: "words piled into one machine, milk siphoned out by another" (Nelson 2016, 100). The child Nelson was breastfeeding, Iggy, was born in 2012. Born and raised in the Digital Age, Iggy will most likely be amazed by the 'Y2K scare' around the turn of the millennium. Although this anecdote seems comical now, for many, anxiety caused by the computer formatting and storage of calendar data for dates in and after the year 2000 overshadowed the New Year celebrations as the calendar moved from 31 December 1999 to 1 January 2000. The Y2K bug, or the Millennium Bug, refers to the potential computer errors caused by the fact that many computer programmes represented four-digit years with only the last two digits. This would mean that the year 2000 would be impossible to distinguish from 1900, as both would be represented as 00. The inability of computer systems to differentiate correctly between dates had the potential to disrupt global infrastructures for many important areas, including banking, air travel and government records, thereby leading to widespread chaos. Luckily, very few major errors occurred in 2000. On the one hand, this led to accusations of that the entire situation had been exaggerated. At the same time, those who had worked to make computers and application programmes Y2K-compliant argued that the threat had indeed been real, but that the absence of major problems was precisely the result of the collective effort to update the systems.

For me, the Y2K bug provides a useful analogy for the 'challenges' genetic criticism is facing regarding born-digital archives. For example, consider the subtitle of Mathijsen's essay about the consequences of digital files for genetic criticism: 'The End

of an Era'. Mathijsen feared that the consequences were so rigorous that the field was facing a new millennium – one *without* genetic criticism. As in the debate regarding the Y2K bug, the debate about the future of genetic criticism was also divided between those who feared catastrophe and those who did not seem to worry about it at all. Either way, the majority of scholars in the field would concur that the traditional methodology needs an update in order to be able to deal with the files to be encountered in the Digital Age. In this dissertation, I have tested keystroke logging as a potential software update that could be implemented to secure the future of genetic criticism. For the field of genetic criticism, it is still 1999. We are still alarmed by the upcoming millennium. It is therefore time to evaluate this software update and see what it can offer as we move into the new millennium.

7.1 DEBUGGING

Before discussing the benefits of keystroke logging as 'software update', I would like to begin this evaluation with some final debugging by raising several important questions about the method. The first concerns whether it will ever be possible to find enough authors who are willing to log their writing with a keystroke logger. Although keystroke logging is obviously not a method with which every author will feel comfortable, some authors are motivated to challenge themselves during the writing process. In February 2023, the Flemish author Saskia De Coster locked herself in a small room behind a transparent wall in the *Heldenzaal* (Heroes Gallery) at the KMSKA (the Royal Museum of Fine Arts Antwerp). During this Marina Abramovic-inspired performance, which is unapologetically entitled 'The Author is Present', De Coster would live and sleep in the museum and, most of all, write her new book. She would refrain from every communication, and instead of observing as an author, she would now be observed by the visitor. This performance shows that authors are not afraid to experiment with their writing processes, even if it means that their writing behaviour is under constant observation. The same applies to the C.M. Taylor. One reason that he recorded his writing process for his novel *Staying On* (2018) was to avoid the solitude that he experienced during writing: "emotionally I was daunted by the long-haul loneliness of novel writing, a process I considered in my most despairing moments as like wallpapering a dungeon" (Taylor 2018). And let us not forget the authors who have participated in our Track Changes project, who have been eager to learn about their writing habits and who have found it relevant to the teaching of creative writing. They have set an excellent example, and I hope they can be an inspiration to other writers. Moreover, authors will never be required to log their writing processes, and I am not advocating this.

The performance by De Coster also touches upon the frequently asked question (e.g. by Linkès 2017) concerning what value can be placed on writing under surveillance.

In other words, how does being 'observed' affect the writing process? With keystroke logging, it is indeed important to consider whether an author's writing process has been affected by the awareness of being logged. Although the answer is likely to be 'probably', it remains open to debate whether these possible influences are actually undesirable and unwanted, and whether they would interfere with a proper analysis of the writing process. At present, my view on this issue is that I believe that writing processes cannot be completely 'directed' and 'staged', as changing writing habits, most of which are unconscious behaviours, would require considerable effort. The writing processes studied here did not seem unnatural: they involved a variety of typographical errors, personal communication through social media, hesitation over formulations and overt revision. I am therefore not under the impression that the authors had staged their writing processes. Of course, it could be that the author deliberately chose not to record certain moments of the writing process. The question is whether this is really problematic, especially given the wealth of data we get from the parts that were recorded. And with paper manuscripts there are always gaps; genetic criticism is used to deal with uncertainties. In the one case I know of where the author made some adjustments to her writing process as a result of being 'observed', the impact on the writing process – and on a genetic analysis – is quite minimal. Posthuma explained at the Track Changes symposium that she did her downtime browsing on a different device. For my research into how the story was written, the absence of downtime in the keystroke logging data does not alter my interpretations of her writing process.

Nevertheless, the intricacies – or sometimes complications – of using Inputlog can obviously influence the ways in which authors work. For example, most of the authors did not work on their personal laptops, but on laptops borrowed for this particular project. In the keystroke logging data, this may have shown up as more typographical errors and searches (e.g. 'how to type an umlaut in Windows'). Troch's keystroke logging data also showed a clear example of how Inputlog had actually affected one of his specific typing habits. In Session 12, after several crashes of Inputlog, Troch wrote in an email to my colleague Floor Buschenhenke that he likes to work with large chunks of text – with all the text in one big paragraph. Inputlog did not seem to be able to handle this well, however, as it kept crashing. He therefore decided to split his text into several paragraphs. Within the context of textual genetic analysis, I do not think that this problem is likely to have caused any major interpretive problems, although it remains unclear whether this adjustment to the author's working method also affected his general writing process. As such, this example is a reminder of the impact that a particular software application can have on an author's writing process.

A related observation is obviously that it will never be possible to know exactly how authors have experienced the writing process and what their thoughts were during the process – unless this is made explicit in a think-aloud protocol or an interview. For

this reason, I have tried to emphasise throughout this study, and I will do so again here, that my analyses of the writing processes remain hypotheses. In terms of the model of the writing process (Leijten et al. 2014), it is not possible to know which ideas the 'proposer' suggested, how the 'translator' transformed this into language or how the 'evaluator' reviewed the ultimate output, unless the 'transcriber' created written text out of it. It is not possible to look into the mind of the author.

A very clear example of this is when the text is a work of autobiographical literature or autofiction, or is partly based on autobiographical aspects. In these cases, the keystroke logging data alone cannot shed light on the autobiographical aspects of the text. In other words, it is impossible to tell whether an aspect of the text is the product of the author's imagination, or based on personal experience, or a combination of both. This became clear in the analysis of Van Pelt's story. The keystroke logging data, which includes many online searches describing the setting of the story, makes it seem as if Van Pelt did extensive research to find the right setting for the story. However, she clarified that she had actually taken the boat trip to Flamingo Island herself. Other stories may also have been based (in part) on personal experience, but as it was beyond the scope of this research to investigate the personal experiences behind the texts, this remains an open question.

The traces of the writing process that remain are highly dependent on the ways in which authors work. Therefore, just as genetic criticism applied to analogue manuscripts, it remains difficult to 'go beyond the singularity' of each particular writing (Plane, Alamargot, and Lebrave 2010, 9). In addition, given that every textual genetic research is prompted by the material, I do not intend to argue that keystroke logging is the only route that should be taken – or that it is the only software update that should be made – for genetic criticism applied to born-digital works of literature. Digital forensic research (e.g. as conducted by Thorsten Ries or Veijo Pulkkinen) is just as necessary to secure the future of genetic criticism.

For authors for whom using a keystroke logger to record their writing processes may seem like a techno-horror, it is always possible to return to the comforting words of Annie Dillard, given that it is not an obligation to track one's changes:

> *Process is nothing; erase your tracks. I hope your tracks have grown over, I hope birds ate the crumbs; I hope you will toss it all and not look back. (Dillard 1990, 4)*

But…

This is obviously not to say that there is no value in exploring the use of keystroke logging for genetic criticism applied to born-digital literary works. On the contrary, to stay within the Y2K analogy, I would argue that the use of keystroke logging is actually one of the software updates that will prepare genetic criticism for the new millennium, the digital age – and not least because it paves the way to improving the study of the cognitive processes involved in literary writing in future research. In what follows, I synthesise the findings of the previous chapters, with a focus on the initial research questions, in order to outline how keystroke logging is one of the steps that can be taken to usher in the new millennium for genetic criticism.

7.2 APPLYING EXISTING METHODS AND THEORIES

The first central question of my dissertation focuses on how the methodology and theories of genetic criticism could be applied to keystroke logging data, and how this might enrich these existing methods and theories:

> *How can existing methods and theories of textual scholarship be applied to reconstruct and analyse digital writing processes, and in which ways will the analysis of digital writing processes logged with keystroke logging software enrich the existing methods and theories?*

In my study, the genetic dossiers include Word documents, as well as keystroke logging material. While the existing methods and theories still seem to be applicable to this specific material, the study encourages a rapprochement between genetic criticism and writing studies, and specifically research on the cognitive writing process. The two objectives of genetic criticism, as distinguished by De Biasi (2004) – establishing an analysable manuscript, mostly in the form of a genetic edition, and the critical aspect of analysing the writing process from a specific point of view – can both benefit from the insights of cognitive writing process research.

7.2.1 Render the ~~handwriting~~ `keystrokes` readable

As the first step in the methodology, the constitution of the genetic dossier remains important with regard to born-digital files. Although the chronology is more of a given in keystroke logging data than it is in analogue documents, geneticists must still examine the files to select the sessions that are relevant to the genetic analysis and to verify whether the keystroke logging files do not contain errors (e.g. incorrect timestamps).

With regard to the question of what constitutes a version, the keystroke logging material provides session versions: states of the text after each writing session. Taking to heart Van Hulle's plea to explicate the unit of the version, I believe that the unit of the session version is the most appropriate starting point for a text-genetic analysis. As a unit, the session version may not fully conform to traditional notions of version, which

is used primarily to denote a particular state of the whole text that represents an author's "overall intention" (Shillingsburg 1996, 44). It nevertheless does provide insight into the "intentions to do" (Shillingsburg 1996, 33), and the aspects of "textual identity" (Vauthier 2019, 40) and "interpretative importance" (Lüdecke 2013 cited in Bleeker 2017, 81) of a version are in fact still applicable to the "session versions" – especially when the aim is to investigate the genesis of the text. Taking session versions as a starting point, geneticists can indeed examine the growth of the text after each session and the revisions that had already been made while working towards the first complete version of the text. Proceeding from the session versions, I was also able to distinguish general writing phases. It is clear is that all kinds of phases can be observed in a single document, and the functions of the text fragment have now become even more important in distinguishing phases.

As with analogue manuscripts, keystroke logging data must be made suitable for genetic analysis. The ability to encode the keystrokes in TEI-XML is one of the ways in which the 'traditional' methodology of genetic editing still proves valuable for keystroke logging data. To this end, I have chosen TEI-XML encoding, as it is familiar to colleagues, while also allowing the reconstruction of the writing process to be shared as a digital edition. For such a genetic edition of keystroke-logged writing processes, the ways in which revisions have been analysed in cognitive writing process research have proven valuable for annotating revisions in TEI-conformant XML. The types of revisions that I have used in the encoding – pre-contextual and contextual revisions – emphasise the writing process, thereby leaving room for further interpretation at a later stage. Although it might seem counterproductive to try to fit a dynamic writing process into the fixed data structure of XML, the simplification of keystroke logging data allows a focus on textual development while including the temporal dimension. Without visualising the reconstructions of the writing process in TEI-XML, the analyses discussed above would not have been possible.

The manual encoding of keystroke logging data is time-consuming, but the transcription of the data has the advantage of making the scholar aware of the peculiarities of the 'writing style' of each individual author. The ability to semi-automate the coding process would be a welcome next step, however, in order to encourage the study of born-digital literary writing processes based on keystroke logging material. Based on these visualisations, revisions could be studied in more detail. Revision episodes and triggers could lead towards the new aspect of textual genesis: nanogenesis.

7.2.2 Nanogenetic endogenesis and exogenesis

The use of a word processor during the production of literary works acts as a mediator between genetic criticism and cognitive writing process research. As demonstrated previously, two important levels of genetic analysis are endogenesis (the composition

of the text) and exogenesis (the use of sources). Both endogenesis and exogenesis can be enriched by an even more detailed temporal dimension.

The temporal aspect of writing, which can now be closely observed with keystroke logging, provides access to aspects of the writing process that were previously more difficult to grasp with analogue material. This led me to introduce the term 'nanogenesis', which focuses on the order of writing actions. Nanogenesis makes it possible to draw connections between different instances of text production and between revisions. As demonstrated by Troch's example, an author can be aware of the need to implement coherence in the story even before the textual references that would ensure this coherence have been properly incorporated into the text. In addition, Van Rijswijk's example reveals textual awareness during composition: from the very first formulation of a sentence, the author tried to avoid repetition of the same phrase.

With regard to exogenesis, the keystroke logging data primarily provide information about the sources that the author consulted on the internet during the writing process. Second, it allows for a reconstruction of the process of appropriation of the sources by revealing how and when the source text has been transformed. Third, in cases of invisible intertextuality, keystroke logging data provide more evidence on whether the author actually used the source. Finally, when the actual source is not given in the genetic material, keystroke logging data may also assist in tracing the source. This is because such data provide us access to stages of the text that can reveal more details on the sources than the published text does, especially when the parts of the text based on the source ultimately proved to be a 'dead end'.

7.3 NEW KNOWLEDGE ABOUT THE GENESIS OF TEXTS

The second central question focuses on what scholars could potentially learn from the reconstruction of the writing process based on keystroke logging material about textual genesis and the creative process:

> *How can digital and narratological text analysis be applied to the reconstructions of the writing process and intermediate text versions to create new knowledge about the genesis of texts and the creative process?*

This question is apparently aimed at general knowledge about the creative process and the genesis of texts. As highlighted by the close examination of the writing processes of Bogaert, Posthuma, Van Rijswijk, Troch and Van Pelt, however, each writing process is unique. The authors differed considerably from each other at the level of writing phases, revision habits and typing styles. While keystroke logging yields the benefits associated with the level of detail, it also makes it necessary to learn how to deal with it.

7.3.1 Cognitive processes and discovery

For the critical dimension (the reconstruction of the logic of the genesis), the models of the writing process established within cognitive writing process research may be helpful in addressing the cognitive processes involved in the creation of literary texts. The reconstruction of the writing process based on the keystroke logging data also provides a way to examine intentions-to-do, as well as the triggers of revision (e.g. digital writing technologies). The investigation of revision episodes can reveal shifts in the intention-to-do, and the act of re-reading is a common trigger for revision, as well as for new text production.

Examination of reconstructions of keystroke logging data can lead to hypotheses about the possibility of discerning the internal mental processes of authors during the composition process within these reconstructions (e.g. when the 'proposer' suggested an idea for the first time or how the 'evaluator' assessed the text produced so far). This subsequently makes it possible to obtain a better account of the aspect of discovery in writing. With respect to discovery, I used nanogenetic analysis to demonstrate how Troch 'discovered' the contours of his fictional world as he wrote, how Van Rijswijk associated during writing and how Van Pelt tried to inhabit her character in order to find an appropriate response in a fictional situation.

7.3.2 Nanogenetic narratology

By examining hesitations in the writing process, as represented by deletions and additions in analogue documents, genetic criticism can hypothesise about questions concerning why the writing proceeded as it did (Van Hulle 2004). This is also very much the case for the genetic analysis of keystroke logging data – even at an unprecedented level of detail in some cases. For example, as demonstrated in the present analysis, Van Rijswijk was not inspired to insert quotation marks around dialogue between the I-narrator and the mother until after she had decided to include a dialogue with the ghost of the grandmother as well. Moreover, if Van Rijswijk had given in to her first impulses to write explicit statements about the 'malfunctions' of the I-narrator's brain in 'Zorgvlied', the reader would have read the story differently: there would be less space for the suspension of disbelief to believe that the I-narrator could actually perceive ghosts in this fictional world.

What can genetic and nanogenetic narratology reveal about narratological aspects, the text and the writing process? It is important to note that few generalisations are possible. Each text, each author and each writing process is far too singular. It nevertheless does show that small details in the text – which may seem insignificant at first glance – can affect larger, sometimes abstract aspects (e.g. story, narrative and narration) when they are changed in the writing process. Moreover, we now have the opportunity to witness the creation of these aspects of literary text right before our eyes.

7.4 USHERING IN THE NEW MILLENIUM

For the Y2K bug of genetic criticism, the new millennium seems full of possibilities. As systems are updated to meet the needs of the digital age (only one of which I have 'tested'), scholars can continue to build on traditional methods and theories. This does not mean, however, that scholars can sit back and relax. As with any software update, continuing improvements are needed. For example, keystroke loggers are not yet optimised for logging long-term (literary) writing processes. Inputlog has the advantage of logging keystroke in an MS Word document – and, soon, in Google Docs and Libre Office. One issue with this, however, is that the positions of keystroke are not always recorded accurately. While other loggers (e.g. GGXLog) offer the advantage of accurate logging, their writing applications itself does not have all the functionalities an author would like to have at hand. If logging eventually becomes more accurate, the automatic reconstruction of the writing process will become a feasible option. There is still work to be done, including further analysing keystroke logging data within the framework of genetic criticism.

The International Y2K Cooperation Center (IYKCC) was set up to promote collaboration aimed at minimising the effects of the Y2K problem. For genetic criticism, and for all other areas of research interested in digital literature and writing processes, I hope that it will be possible to build a community to promote sustainable digital archiving, with keystroke logging as one recommendation. There is much to be learned by combining the analyses of a diverse group of authors. As long as we keep making updates, genetic criticism will continue to be relevant in the digital age. For now, let us raise our glasses to usher in the new millennium for genetic criticism.

SUMMARY

'Behind the computer screens'
The use of keystroke logging for genetic criticism applied to born-digital works of literature

This dissertation explores the ways in which a keystroke logging tool can be used for genetic criticism applied to born-digital works of literature. Genetic criticism is a field of research concerned with the dynamics of literary writing processes. The digital environment in which present-day literature is composed, significantly changes the materiality of the 'traces' of the writing process. Present-day literary authors write their texts mainly in a word processor. However, common word processors tend to hide the writing operations: additions are always visualised on the screen as inline text production, and deleted text 'disappears' from the screen. In addition, how structured authors are in saving different versions under different names is highly variable, let alone that they document the writing process working towards the first version. These changes first led to fears that genetic criticism would not survive the transition to the digital age, since the genesis of the text is predominantly digital.

The dissertation shows that this will not be the case as long as genetic criticism adapts to working with digital files, even though the author's working method often remains the leading factor in analysis. One possibility is digital forensics, where forensic methods can be used to recover and examine files from old computers. For example, a hex editor can be used to explore the binary structure of a document, which may contain a hidden genetic layer of text. The work of Matthew K. Kirschenbaum, Doug Reside and Thorsten Ries are prime examples of this kind of research. But, as Bénédicte Vauthier has shown, all the remaining files, folder structure and metadata can also provide enough information for a thorough text genetic analysis. However, the above-mentioned research methods do not always allow us to analyse the genesis of the text at a very detailed level; large parts of the genesis of the text are often still missing, especially at the level of the immediate revision. This gap can be filled by keystroke logging.

This study explores the possibilities of keystroke logging by analysing the writing processes of Gie Bogaert, Jente Posthuma, Roos van Rijswijk, David Troch, and Ellen

Van Pelt. Each writer recorded their writing process using the keystroke logging programme Inputlog. This software, when activated by the author, records every keystroke and mouse movement in Microsoft Word, together with a timestamp, and also offers the possibility of recording all other opened 'windows'. It also automatically saves the Word file in a separate folder at the beginning and end of a writing session. Using the keystroke logging data and text versions, the study answers the question of how we can apply existing methods and theories of textual scholarship and genetic criticism to reconstruct and analyse logged writing processes, and how keystroke logging data enriches these methods and theories.

A key concept in textual scholarship is the version, and it is therefore important to define what a version is at the outset of the research. Dirk Van Hulle advocates making explicit the unit of the text being studied and linking it to the unit of the version, such as versions of a sentence. In the approach to the material that can be obtained by keystroke logging, the size of the unit to be studied also plays a crucial role. The genetic dossier, which in this case consists of the keystroke logging data and the Word files saved at the beginning and end of the writing session (possibly supplemented by other material supplied by the authors, such as notebooks and hand-corrected printed versions), lends itself ideally to centralise the unit of the 'session' – in itself also a unique opportunity for genetic criticism.

Although the session version is not intentionally saved as a version by the author, these versions provide insights into certain 'intentions to do', the motivational force underlying a particular writing action (Shillingsburg, Bushell). While it was previously thought that phases could no longer be distinguished in the working process of a digitally working writer, the session versions actually show that the digital writing process still consists of different phases. The session versions also provide insight into the phases of the writing process. In the case of Gie Bogaert, the session versions show clear cycles in which he goes through the stages of initialising, inserting notes, textualising, revising and finalising each overarching chapter. Examining the stages in the different session versions therefore creates new possibilities for studying macrogenesis (the genesis of the text as a whole). In Posthuma's, Van Rijswijk's, Troch's and Van Pelt's material, more idiosyncratic phases can also be observed, such as starting with notes, which often remain at the end of the document, as well as unused text fragments, revision phases at the beginning of a session (which also indicates the rereading of the text produced so far), or a general revision phase at the end of the writing process. What also becomes visible in the session versions are highly temporal phases, such as reminders of ideas that have yet to be incorporated or doubts about a particular aspect of the text or word choice. In short, different phases of the writing process can be visible in one Word document.

Where traditional methods of textual scholarship also prove valuable in working with keystroke logging data is in transcribing the writing process into TEI-XML. Although the TEI (Text Encoding Initiative) does not provide guidelines for encoding keystroke logging data, it does offer ways of making this data useful for textual scholarship. Indeed, the various types of output provided by Inputlog are not entirely suited to textual genetic analysis, as there are few visualisations that focus on textual development. However, a reconstruction of the writing process in TEI-XML offers a solution here, whereby the method from textual scholarship can be combined with theories from (cognitive) writing process research. Indeed, there is common ground between the way revisions are described in both fields, even if the terminology is different. By using the types of revision used in writing process research, in particular the distinction between pre-contextual revisions (revisions that take place immediately, i.e., before a sentence is completed) and contextual revisions (revisions that take place in a text that has already been written), the writing process can be reconstructed, leaving room for interpretation at a later stage of the research. TEI-XML-based visualisation allows researchers to see the changes in the text at a glance and to replay the writing session. The dynamics of the writing process remain visible, allowing interpretation of the writer's path through the text.

When writing processes are recorded using a keystroke logger, the word processor is not an obstacle to genetic criticism, but rather a mediator between genetic criticism on the one hand and cognitive writing process research on the other. This opens up new possibilities for the analysis of the creative process. The second question central to this study is therefore how we can apply digital and narratological text analysis to the reconstructions of logged writing processes and session versions in order to gain new insights into the genesis of the text and the creative process.

In analysing revisions, both keystroke logging data and session versions show that the affordances of the word processor, such as the well-known freedom of the word processor to start typing at any point in the text, can stimulate revision. The keystroke logging data also provides a better understanding of revision episodes, where revisions are related to each other (in time and/or content). Whereas with analogue manuscripts these links between revisions can only be established to a limited extent, the data in this case provide more certainty – it is actually possible to see how revisions follow on from each other, and thus to establish that one revision was a trigger for another. On this basis, the revisions can be better interpreted and understood. Furthermore, the reconstructions of the writing processes show that rereading the already produced text can also lead to the emergence of new ideas and thus to revisions.

The keystroke logging data also enriches genetic criticism by providing an even more detailed insight into the temporal dimension of the writing process. First, it provides insight into the way writers work; by recording the exact moment of writing,

we can see not only when writing sessions took place and how long writers wrote, but also *how* they wrote. The study shows that writers differ in the way they perform certain writing actions. Van Rijswijk and Van Pelt, for example, made more immediate revisions while writing a sentence, whereas Bogaert, Posthuma and Troch more often revised an already written sentence.

The granularity with which the genesis of the text can be examined is referred to in this study as the nanogenesis. The nanogenesis involves the ability to trace the movement of authors through the text. It shows, for example, how linearly the authors proceeded, where in the text they wrote, how new text production and revision alternated, how characters were removed – someone typing on a computer also develops their own typing style in the process. In short, the exact sequence of writing the characters that make up the final text. Nanogenetic analysis can provide insight into how Troch made his text coherent, and how Van Rijswijk, in order to avoid repetition, worked on previously written formulations while writing. Nanogenesis also offers the possibility of hypothesising about the process of discovery during writing. In the data studied, this 'discovery' seems to relate to several parts of the text, providing insight into, among other things, the associations made during writing. In addition to the possibility of analysing endogenesis (text-internal developments) from a nanogenetic perspective, keystroke logging also offers the possibility of studying exogenesis (extratextual impressions) at the nanogenetic level. Inputlog can record all windows opened during the writing session, allowing a quantitative mapping of all online resources accessed, including the moment when the author decides to gather additional information on the Internet. In addition, keystroke logging data shows how a resource (analogue or digital) is transformed during the writing process.

Finally, (nano)genetic narratology provides a well-founded framework for bringing together the aforementioned advantages of keystroke logging data in an analysis of how details of the narrative text – events, actants, setting, time, characterisation, focalisation, narrative perspective and consciousness representation – are integrated into the text and how they change during the writing process. Here, the nanogenesis can contain clues for understanding the choices made in relation to these narratological aspects. For all of the above, however, they remain the researcher's hypotheses. Although the keystroke logging data provide a great deal of information about the course of the writing process, the researcher obviously cannot see inside the writer's head.

All in all, this dissertation shows how keystroke logging is a way of continuing text genetic research in the digital age. We can only hope that the writers who made this research possible will also inspire other writers to record their writing process.

SAMENVATTING

'Achter de computerschermen'
Het gebruik van een toetsaanslagregistratieprogramma voor
tekstgenetisch onderzoek naar digitaal geschreven literatuur

Dit proefschrift onderzoekt op welke manier een toetsaanslagregistratieprogramma kan worden ingezet voor tekstgenetisch onderzoek naar literaire teksten geschreven in het digitale tijdperk. De tekstgenetica is een onderzoeksveld dat zich bezighoudt met de dynamiek van (literaire) schrijfprocessen. De digitale omgeving waarin hedendaagse literaire teksten worden geschreven, verandert echter de materialiteit van de 'sporen' van het schrijfproces. Hedendaagse auteurs schrijven hun teksten voornamelijk in een tekstverwerker, waar de *backspace*- of *delete*-toets de geschrapte tekst daadwerkelijk van het scherm doet verdwijnen, en waar toevoegingen 'vlekkeloos' in de zin worden opgenomen. En daarnaast is het de vraag hoe gestructureerd auteurs zijn in het bewaren van verschillende versies onder een andere naam en of het toewerken naar een eerste versie in de eerste plaats wel wordt gedocumenteerd. Deze veranderingen leidde eerst tot de vrees dat de tekstgenetica de transitie naar het digitale tijdperk, waarin de tekstgenese met name digitaal plaatsvindt, niet zou overleven.

Het proefschrift toont aan dat het zo ver niet zal komen zolang de tekstgenetica zich aanpast aan het werken met digitale bestanden, al blijft de werkmethode van de schrijver wel vaak leidend. Een van de mogelijkheden hierbij is digitaal forensisch onderzoek, waarbij forensische methodes kunnen worden toegepast om bestanden van oude computers te herstellen en te onderzoeken. Met een *hex-editor*, bijvoorbeeld, kan inzicht worden verkregen in de binaire structuur van een document, waar zich mogelijk een verborgen tekstgenetische laag kan bevinden. Het onderzoek van Matthew K. Kirschenbaum, Doug Reside, en Thorsten Ries zijn hier uitgelezen voorbeelden van. Maar ook alle overgebleven bestanden, de mappenstructuur, en de metadata kan, zoals Bénédicte Vauthier aantoonde, genoeg opleveren voor een gedegen tekstgenetische analyse. Toch kan door bovenstaande onderzoeksmethodes de tekstgenese niet altijd op een zeer gedetailleerd niveau worden geanalyseerd; er ontbreken vaak nog grote

delen van de ontstaansgeschiedenis van de tekst, met name op het niveau van onmiddellijke revisie. Het is deze lacune die door *keystroke logging* kan worden opgevuld.

Deze studie onderzoekt de mogelijkheden die *keystroke logging* biedt door de schrijfprocessen van Gie Bogaert, Jente Posthuma, Roos van Rijswijk, David Troch, en Ellen Van Pelt te analyseren. Zij legden elk hun schrijfproces vast met het toetsaanslagregistratieprogramma Inputlog. Deze software, indien door de auteur geactiveerd, registreert elke toetsaanslag en muisbeweging in Microsoft Word in combinatie met een tijdstempel, en biedt daarnaast de mogelijkheid om ook alle andere geopende 'vensters' vast te leggen. Verder slaat het automatisch het Word bestand op in een aparte map, zowel aan het begin als aan het eind van een schrijfsessie. Met deze toetsaanslag-data en tekstversies als uitgangspunt, beantwoordt de studie de vraag hoe we, enerzijds, bestaande methoden en theorieën van *textual scholarship* en de tekstgenetica kunnen toepassen om de gelogde schrijfprocessen te reconstrueren en te analyseren, en hoe, anderzijds, de toetsaanslag data deze methoden en theorieën verrijkt.

Een belangrijk begrip in *textual scholarship* is de versie, en het is daarom van belang om bij aanvang van de tekstgenetische studie te definiëren wat er on een versie. Dirk Van Hulle pleit voor het expliciet maken van de te onderzoeken eenheid van de tekst, en daar de eenheid van de versie aan te koppelen, zoals versies van een zin. In de benadering van het materiaal dat door *keystroke logging* verkregen kan worden, speelt de grootte van de te onderzoeken eenheid ook een cruciale rol. Het tekstgenetische dossier, dat in dit geval bestaat uit de toetsaanslag-data en de Word bestanden opgeslagen aan het begin en eind van de schrijfsessie (aangevuld door mogelijk ander materiaal aangeleverd door de auteurs, zoals notitieschriften en met de hand gecorrigeerde geprinte versies), leent zich bij uitstek om de eenheid van de 'sessie' – op zichzelf ook een unieke opportuniteit voor de tekstgenetica – centraal te stellen.

Hoewel de sessie-versie niet doelbewust als versie door de auteur is opgeslagen, geven deze versies wel inzichten in bepaalde 'intenties om te doen', de motiverende kracht die ten grondslag ligt aan een bepaalde schrijfactie (Shillingsburg, Bushell). Hoewel eerder werd gedacht dat er geen fases meer zouden kunnen worden onderscheiden in het werkproces van een digitaal werkende schrijver, tonen de sessie-versies juist aan dat het digitale schrijfproces nog wel degelijk uit verschillende fases bestaat. De sessie-versies geven ook inzicht in de fases van het schrijfproces. Bij Gie Bogaert zijn in de sessie-versies duidelijke cycli aan te wijzen, waarin hij per overkoepelend hoofdstuk de fases van initialiseren, invoegen van notities, tekstualiseren, reviseren, en finaliseren doorloopt. Het onderzoek naar de fases in de verschillende sessie-versie biedt daarom nieuwe mogelijkheden voor het bestuderen van de macrogenese (de tekstgenese van de tekst als geheel). In het materiaal van Posthuma, Van Rijswijk, Troch, en Van Pelt zijn ook, maar meer idiosyncratische fases waar te nemen, zoals het starten met aantekeningen, die vaak aan het eind van het document

blijven staan, evenals ongebruikte tekstfragmenten, revisie-fases aan het begin van een sessie (dat ook het nalezen van de al geproduceerde tekst aanduidt), of een algehele revisie-fase aan het eind van het schrijfproces. Wat tevens zichtbaar wordt in de sessie-versies zijn de zeer temporele fases, zoals herinneringen aan ideeën die nog geïncorporeerd moeten worden, of twijfels over een bepaald aspect van de tekst, of woordkeuze. Kortom, verschillende fases in het schrijfproces kunnen zichtbaar zijn in één Word document.

Waar de traditionele methoden van *textual scholarship* ook van waarde blijken voor het werken met toetsaanslag-data is voor het transcriberen van het schrijfproces in TEI-XML. Hoewel de TEI (Text Encoding Initiative) geen richtlijnen heeft voor het encoderen van toetsaanslag-data, biedt het wel mogelijkheden om deze data inzichtelijk te maken voor tekstgenetisch onderzoek. De verschillende soorten output die Inputlog aanbiedt, blijken namelijk niet geheel geschikt voor een tekstgenetische analyse, omdat er weinig visualisaties zijn die focussen op de tekstuele ontwikkelingen. Een reconstructie van het schrijfproces in TEI-XML biedt hierbij de uitkomst, waarbij de methode uit *textual scholarship* kan worden gecombineerd met de theorieën uit het (cognitief) schrijfproces onderzoek. Er blijken namelijk raakvlakken te bestaan tussen de manier waarop revisies worden beschreven in beide vakgebieden, hoewel de terminologie verschilt. Door de revisie-typen te gebruiken die worden gehanteerd in het schrijfprocesonderzoek, voornamelijk het onderscheid tussen pre-contextuele revisies – revisies die onmiddellijk plaatsvinden, en dus voordat een zin is afgeschreven – en contextuele revisies – revisies die plaatsvinden in een al geschreven tekst kan het schrijfproces worden gereconstrueerd, waarbij ruimte wordt gelaten voor interpretatie in een later stadium van het onderzoek. De op de TEI-XML gebaseerde visualisatie stellen de onderzoekers in staat om de wijzigingen in de tekst in één oogopslag te zien, en ook om de schrijfsessie af te laten spelen. De dynamiek van het schrijfproces blijft zo zichtbaar, waardoor het pad dat de schrijver door de tekst aflegt kan worden geïnterpreteerd.

Wanneer de schrijfprocessen worden vastgelegd met een toetsaanslagregistratieprogramma blijkt de tekstverwerker geen obstakel te vormen voor tekstgenetisch onderzoek, maar wordt het juist een bemiddelaar tussen de tekstgenetica enerzijds en het cognitief schrijfproces onderzoek anderzijds. Dit opent nieuwe mogelijkheden voor de analyse van het creatieve proces. De tweede vraag die derhalve in deze studie centraal staat is op welke manier we digitale en narratologische tekstanalyse kunnen toepassen op de reconstructies van de gelogde schrijfprocessen en sessie-versies om nieuwe kennis te vergaren over de tekstgenese en het creatieve proces.

Voor de analyse van revisies, toont de toetsaanslag-data alsmede de sessie-versies, dat de mogelijkheden die de tekstverwerker de schrijver biedt, zoals de alom bekende vrijheid die de tekstverwerker biedt om op allerlei plaatsen in de tekst te kunnen gaan

typen, revisie kunnen stimuleren. De toetsaanslag-data geeft ook een beter inzicht in revisie-episodes, waarbij revisies aan elkaar gerelateerd zijn (in tijd en/of in inhoud). Waar bij analoge manuscripten deze verbanden tussen revisies maar tot op een zekere hoogte kan worden vastgesteld, geeft de data hierin meer zekerheid – er kan daadwerkelijk worden gezien hoe revisies elkaar opvolgen en daarom worden vastgesteld dat een revisie een *trigger* is geweest voor een andere revisie. Op basis hiervan kunnen de revisies beter worden geïnterpreteerd en begrepen. Verder tonen de reconstructies van de schrijfprocessen aan dat ook het herlezen van de al geproduceerde tekst kan leiden tot het ontstaan van nieuwe ideeën, en dus tot revisie.

De toetsaanslag-data verrijkt de tekstgenetica tevens doordat het een nog gedetailleerder inzicht geeft in de temporele dimensie van het schrijfproces. Allereerst geeft het inzicht in de werkmethode van de schrijvers; doordat het precieze moment van schrijven is vastgelegd, kunnen we niet alleen zien wanneer de schrijfsessies plaatsvonden en hoe lang de schrijvers schreven, maar ook *hoe* ze schreven. Uit de studie blijkt dat de auteurs verschillen in de manier waarop ze bepaalde schrijfacties uitvoeren. Van Rijswijk en Van Pelt, bijvoorbeeld, maakten meer onmiddellijke revisies *tijdens* het schrijven van een zin, terwijl Bogaert, Posthuma, en Troch, vaker al een geschreven zin reviseerden.

De fijnmazigheid waarmee de ontwikkeling van de tekst kan worden onderzocht wordt in deze studie nanogenetisch onderzoek genoemd. De nanogenese omvat onder andere de mogelijkheid om te beweging van de auteur door de tekst heen te volgen. Dit laat bijvoorbeeld zien hoe lineair de schrijvers te werk ging, op welke locaties ze in de tekst schreven, hoe nieuwe tekstproductie en revisie elkaar afwisselde, de manier waarop karakters worden verwijderd – iemand die typt op de computer ontwikkelt daarbij ook een eigen typstijl. Kortom, de exacte volgorde van het schrijven van de karakters die de uiteindelijke tekst vormen. Door middel van een nanogenetische analyse inzichtelijk worden gemaakt hoe Troch zijn tekst coherent maakte, een hoe Van Rijswijk, om herhaling te voorkomen, al tijdens het schrijven bezig was met eerder geschreven formuleringen. De nanogenese biedt ook de mogelijkheid om hypotheses op te stellen over het proces van het ontdekken (*discovery*) tijdens het schrijven. In de onderzochte data blijkt dit 'ontdekken' betrekking te hebben op verscheidene onderdelen van de tekst, en geeft het onder andere inzicht in de associaties die tijdens het schrijven worden gemaakt. Naast dat de endogenese (tekstinterne ontwikkelingen) op een nanogenetisch perspectief kan worden genanalysereerd, biedt *keystroke logging* ook de mogelijkheid om de exogenese (extratekstuele impressies) op een nanogenetisch niveau te bestuderen. Inputlog kan alle tijdens de schrijfsessie geopende vensters vastleggen, waardoor alle geraadpleegde online-bronnen kwantitatief in kaart kunnen worden gebracht, inclusief het moment waarop de auteur besloot om extra informatie te vergaren op het Internet.

Daarnaast laat de toetsaanslag-data ook zien hoe een bron (analoog of digitaal) tijdens het schrijfproces wordt getransformeerd.

Tekstgenetische narratologie biedt ten slotte een gedegen kader om de hierboven beschreven profijten van de toetsaanslag-data samen te brengen in een analyse van hoe details van de verhalende tekst – de gebeurtenissen, de actants, de setting, de tijd, de karakterisering, de focalisatie, het vertelpersperspectief, en de bewustzijnsvoorstelling – in de tekst worden geïntegreerd en hoe ze veranderen tijdens het schrijfproces. De nanogenese kan hierbij aanwijzingen geven voor het begrijpen van de keuzes die met betrekking tot deze narratologische aspecten zijn gemaakt. Voor al het bovenstaande geldt echter wel dat het hypotheses van de onderzoeker blijven. Hoewel de toetsaanslag-data veel informatie biedt over het verloop van het schrijfproces, kan de onderzoeker natuurlijk niet in het hoofd van de schrijver kijken.

Al met al toont dit proefschrift aan hoe het registreren van toetsaanslagen een manier is om het tekstgenetisch onderzoek ook in het digitale tijdperk kan worden voortgezet. We kunnen nu alleen hopen dat de schrijvers die dit onderzoek mogelijk hebben gemaakt ook andere schrijvers inspireren om hun schrijfproces vast te leggen.

LIST OF TABLES

LIST OF FIGURES

BIBLIOGRAPHY

Afman, Huib. 2016. *Where the Magic Happens: Schrijverskamer = Writers' Rooms*. Tielt: Lannoo.

Aira, César. 2010. *The Literary Conference*. Translated by Katherine Silver. New York: New Directions Pub.

Alamargot, Denis, and Jean-Louis Lebrave. 2010. 'The Study of Professional Writing: A Joint Contribution from Cognitive Psychology and Genetic Criticism'. *European Psychologist* 15 (1): 12–22. https://doi.org/10.1027/1016-9040/a000001.

Alexander, Benjamin. 2015. 'The Salman Rushdie Archive and the Re-Imagining of a Philological E-Volution'. In *Texts, Transmissions, Receptions*, edited by André Lardinois, Sophie Levie, Hans Hoeken, and Christoph Lüthy, 71–95. Leiden: Brill. https://doi.org/10.1163/9789004270848_006.

Andrews, Tara L., and Joris J. van Zundert. 2018. 'What Are You Trying to Say? The Interface as an Integral Element of Argument'. In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 12:3–33. Schriften Des Instituts Für Dokumentologie Und Editorik. Norderstedt: Books on Demand (BoD).

Baaijen, Veerle M., and David Galbraith. 2018. 'Discovery Through Writing: Relationships with Writing Processes and Text Quality'. *Cognition and Instruction* 36 (3): 199–223. https://doi.org/10.1080/07370008.2018.1456431.

Barney, Brett. 2020. 'TEI, the Walt Whitman Archive, and the Test of Time'. *Journal of the Text Encoding Initiative*, 13. https://doi.org/10.4000/jtei.3249.

Bartscherer, Thomas. 2003. 'Ecce HyperNietzsche: It's Not Just the Philology of the Future Anymore'. *Journal of New Media & Culture* 2 (2). https://www.ibiblio.org/nmediac/fall2003/ecce.htm.

Barwell, Graham. 2005. 'Original, Authentic, Copy: Conceptual Issues in Digital Texts'. *Digital Scholarship in the Humanities* 20 (4): 415–24. https://doi.org/10.1093/llc/fqi046.

Baughman, Melvin J. 2019. "How Trees and Woodlands Grow." In *Woodland Stewardship: A Practical Guide for Midwestern Landowners*, 3rd edition., 33–41. Minneapolis, MN: University of Minnesota Libraries Publishing.

Beal, Peter. 2011. *A Dictionary of English Manuscript Terminology: 1450-2000*. Oxford: Oxford University Press.

Bekius, Lamyk. 2021a. 'The Reconstruction of the Author's Movement Through the Text, or How to Encode Keystroke Logged Writing Processes in TEI-XML'. *Variants*, 15–16: 3–43. https://doi.org/10.4000/variants.1245.

———. 2021b. 'De geboorte van een plein: De analoge en digitale tekstgenese van het incipit in Gie Bogaerts roman Roosevelt'. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 137 (2): 122–49. https://doi.org/10.5117/TNTL2021.2.002.BEKI.

Bereiter, C., and M. Scardamalia. 1987. *The Psychology of Written Composition*. Hillsdale, N.J: Erlbaum.

Bernaerts, Lars, and Dirk Van Hulle. 2013. 'Narrative across Versions: Narratology Meets Genetic Criticism'. *Poetics Today* 34 (3): 281–326. https://doi.org/10.1215/03335372-2325232.

Bernaerts, Lars, Dirk Van Hulle, and Gunther Martens. 2011. 'Narratologie En Tekstgenese: Een Terreinverkenning'. *Spiegel Der Letteren*, 3: 281–309. https://doi.org/10.2143/SDL.53.3.2131161.

Bernlef, J. 2012. *Help me herinneren: verhalen*. 2e druk. Amsterdam: Querido.

Bleeker, Elli. 2015. 'The Future of the Digital Scholarly Editor. Interpretation, Subjectivity and Presence?' *Manuscrítica. Revista de Crítica Genética* 28: 112–22.

———. 2017. 'Mapping Invention in Writing: Digital Infrastructure and the Role of the Genetic Editor'. Doctor of Philosophy in Literature, Antwerpen: University of Antwerp.

Bleeker, Elli, and Aodhán Kelly. 2018. 'Interfacing Literary Genesis'. In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 12:193–218. Schriften Des Instituts Für Dokumentologie Und Editorik. Norderstedt: Books on Demand (BoD).

Bleier, Roman, and Helmut W. Klug. 2018. 'Discussing Interfaces in Digital Scholarly Editing'. In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 12:V–XV. Schriften Des Instituts Für Dokumentologie Und Editorik. Norderstedt: Books on Demand (BoD).

Bogaert, Gie. 2016. *Roosevelt: Roman*. Amsterdam: De Bezige Bij.

Bolter, J. David. 1991. *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, N.J: L. Erlbaum Associates.

Boomen, Marianne van den. 2014. *Transcoding the Digital: How Metaphors Matter in New Media*. Amsterdam: Institute of Network Cultures.

Boomen, Marianne van den, Sybille Lammes, Ann-Sophie Lehmann, Joost Raessens, and Mirko Tobias Schäfer, eds. 2009. *Digital Material: Tracing New Media in*

Everyday Life and Technology. MediaMatters. Amsterdam: Amsterdam University Press.

Bowen, Neil, and Luuk Van Waes. 2020. 'Exploring Revisions in Academic Text: Closing the Gap Between Process and Product Approaches in Digital Writing'. *Written Communication* 37 (3): 322–64.
https://doi.org/10.1177/0741088320916508.

Bryant, John. 2002. *The Fluid Text: A Theory of Revision and Editing for Book and Screen / John Bryant*. Editorial Theory and Literary Criticism. Ann Arbor: University of Michigan Press.

Buch Leander, Niels. 2018. *The Sense of a Beginning: Theory of the Literary Opening*. Copenhagen: Museum Tusculanum Press.

Burnard, Lou. 2014. What Is the Text Encoding Initiative?: How to Add Intelligent Markup to Digital Resources. OpenEdition Press.
https://doi.org/10.4000/books.oep.426.

Buschenhenke, Floor. 2016. 'Het Literaire Werk Anno 2016. Digitale Schrijfprocessen Vastleggen En Analyseren.' *Vooys* 34 (4): 8–20.

———. forthcoming. 'Inspiration Is for Amateurs: A Process Approach to Contemporary Fiction Writing'. Doctor of Philosophy in Literature, Amsterdam ; Antwerpen: University of Amsterdam & University of Antwerp.

Buschenhenke, Floor, Rianne Conijn, and Luuk Van Waes. 2023. "Measuring non-linearity of multi-session writing processes. *Reading and Writing*. https://doi.org/10.1007/s11145-023-10449-9

Bushell, Sally. 2005. 'Intention Revisited: Towards an Anglo-American "Genetic Criticism"'. *Text* 17: 55–91.

———. 2009. *Text as Process: Creative Composition in Wordsworth, Tennyson, and Dickinson*. Charlottesville: University of Virginia Press.

Calonne, David Stephen. 2006. 'Creative Writers and Revision'. In *Revision: History, Theory, and Practice*, edited by Alice S. Horning and Anne Becker, 142–76. Reference Guides to Rhetoric and Composition. West Lafayette, Ind: Parlor Press.

Chandler, Daniel. 1992. 'The Phenomenology of Writing by Hand'. *Intelligent Tutoring Media* 3 (2–3): 65–74. https://doi.org/10.1080/14626269209408310.

———. 1995. The Act of Writing: A Media Theory Approach. Aberystwyth: Prifysgol Cymru.

Cislaru, Georgeta. 2015. 'Writing(s) at the Crossroads: The Process-Product Interface'. In *Writing(s) at the Crossroads: The Process-Product Interface*, edited by Georgeta Cislaru, 1–17. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Cohn, Dorrit. 1968. 'K Enters The Castle: On the Change of Person in Kafka's Manuscript'. *Euphorion* 62 (1): 28–45.

Conijn, Joost, dir. 2012. 'A.L. Snijders - Een Handige Dromer'. *Het Uur van de Wolf.*

Conijn, Rianne, Emily Dux 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/10.1177/07410883211052104.

Conijn, Rianne, Menno van Zaanen, Mariëlle Leijten, and Luuk Van Waes. 2019. 'How to Typo? Building a Process-Based Model of Typographic Error Revisions'. *The Journal of Writing Analytics* 3: 69–95.

Crasson, Aurèle, Jean-Louis Lebrave, and Jérémy Pedrazzi. 2019. 'Le « siliscrit » de Jacques Derrida. Exploration d'une Archive Nativement Numérique'. *Genesis*, 49 (December): 7–12. https://doi.org/10.4000/genesis.4316.

Crasson, Aurèle, Jean-Louis Lebrave, Jérémy Pedrazzi, and Laurent Alonso. 2022. 'Le Toucher Touch/to Touch Him1: Étude Forensique de Données Numériques de Jacques DerridaL'archive Numérique et Les Modalités de Son Appréhension Pour La Recherche'. *Genesis*, 55 (December): 1–200. https://doi.org/10.4000/genesis.7747.

Creutzburg, Reiner. 2017. 'The Strange World of Keyloggers - an Overview, Part I'. *Electronic Imaging* 29 (6): 139–48. https://doi.org/10.2352/ISSN.2470-1173.2017.6.MOBMU-313.

Crombez, Thomas, and Edith Cassiers. 2015. 'Postdramatic Methods of Adaptation in the Age of Digital Collaborative Writing'. *Digital Scholarship in the Humanities*, October, fqv054. https://doi.org/10.1093/llc/fqv054.

Crowley, Ronan, and Dirk Van Hulle, eds. 2016. *New Quotatoes: Joycean Exogenesis in the Digital Age.* European Joyce Studies, volume 25. Leiden: Brill.

Dahlström, Mats. 2000. 'Drowning by Versions'. *Human IT* 4 (4). https://humanit.hb.se/article/view/174/187.

Davidson, Andrea H. 2022. 'Writing Talk: Constructions of Adolescence in the Early Genesis of Aidan Chambers' Dance Sequence Novels'. Doctor of Philosophy in Literature, Antwerpen: University of Antwerp.

De Biasi, Pierre-Marc. 2004. 'Toward a Science of Literature: Manuscript Analysis and the Genesis of the Work'. In *Genetic Criticism: Texts and Avant-Textes*, edited by Jed Deppman, Daniel Ferrer, and Michael Groden, 36–68. Material Texts. Philadelphia: University of Pennsylvania Press.

———. 2005. *La génétique des textes.* 128. Paris: A. Colin.

———. 2010. 'Pour Une Génétique Généralisée : L'approche Des Processus à l'âge Numérique'. *Genesis*, 30 (June): 163–75. https://doi.org/10.4000/genesis.133.

De Biasi, Pierre-Marc, and Ingrid Wassenaar. 1996. 'What Is a Literary Draft? Toward a Functional Typology of Genetic Documentation'. *Yale French Studies*, 89: 26. https://doi.org/10.2307/2930337.

Debray-Genette, Raymonde. 1977. 'Génétique et poétique : esquisse de méthode'. *Littérature* 28 (4): 19–39. https://doi.org/10.3406/litt.1977.2073.

Del Lungo, Andrea. 2003. *L' incipit romanesque*. Poétique. Paris: Éd. du Seuil.

Derrida, Jacques. 2005. *Paper Machine*. Cultural Memory in the Present. Stanford, Calif: Stanford University Press.

Dillard, Annie. 1990. *The Writing Life*. New York: HarperPerennial.

Dillen, Wout. 2015. 'Digital Scholarly Editing for the Genetic Orientation: The Making of a Genetic Edition of Samuel Beckett's Works'. Doctor of Philosophy in Literature, Antwerpen: University of Antwerp.

———. 2018. 'The Editor in the Interface: Guiding the User through Texts and Images'. In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 12:35–59. Schriften Des Instituts Für Dokumentologie Und Editorik. Norderstedt: Books on Demand (BoD).

D'Iorio, Paolo. 2010. 'Qu'est-Ce Qu'une Édition Génétique Numérique ?' *Genesis*, 30 (June): 49–53. https://doi.org/10.4000/genesis.116.

Donk, Toine, and Daniël van der Meer, eds. 2016. *Gekrenkt & hongerig: ongepubliceerde debuten van de grootste schrijvers*. Amsterdam: Das Mag.

Doquet, Claire. 2014. 'Pour une approche linguistique de l'écriture enregistrée'. In *Temps de l'écriture: enregistrements et représentations*, edited by Christophe Leblay and Gilles Caporossi, 21–42. Sciences du langage, carrefours et points de vue, n° 12. Louvain-la-Neuve (Belgique) [Paris]: Academia l'Harmattan.

Dorrestein, Renate. 2017. *Onder de motorkap van het schrijverschap*. Amsterdam: Em. Querido's Uitgeverij BV.

Doyle, Charlotte L. 1998. 'The Writer Tells: The Creative Process in the Writing of Literary Fiction'. *Creativity Research Journal* 11 (1): 29–37.

Drucker, Johanna. 2013. 'Performative Materiality and Theoretical Approaches to Interface'. *Digital Humanities Quarterly* 7 (1). http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html.

Ehrensberger-Dow, Maureen, and Daniel Perrin. 2009. 'Capturing Translation Processes to Access Metalinguistic Awareness'. *Across Languages and Cultures* 10 (2): 275–88. https://doi.org/10.1556/Acr.10.2009.2.6.

Elbow, P. 1973. *Writing without Teachers*. London: Oxford University Press.

Faigley, Lester, and Stephen Witte. 1981. 'Analyzing Revision'. *College Composition and Communication* 32 (4): 400–414.

Fenoglio, Irène. 2009. 'Textual Genetics and Manuscript in Word Processing. A New Definition of the Text?: Essay on the "Avant-Texte" of a Short Story by Pascal Quignard'. *Belgian Journal of Linguistics* 23 (1): 45–61.

———. 2015. 'From Writing under Production to the Finished Product: A Processual Threshold'. In *Writing(s) at the Crossroads*, edited by Georgeta Cislaru, 125–50. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.194.07fen.

Ferrer, Daniel. 2022. 'Models for Genetic Criticism'. Paper given at the Genesis 2022 Conference. Oxford, March 15.

Fitzgerald, Jill. 1987. 'Research on Revision in Writing'. *Review of Educational Research* 57 (4): 481–506. https://doi.org/10.3102/00346543057004481.

Flower, Linda S., and John R. Hayes. 1980. 'The Cognition of Discovery: Defining a Rhetorical Problem'. *College Composition and Communication* 31 (1): 21. https://doi.org/10.2307/356630.

Flower, Linda S., John R. Hayes, Karen S. Scriver, James F. Stratman, and Linda Carey. 1987. 'Cognitive Processes in Revision'. In *Advances in Applied Psycholinguistics*, edited by Sheldon Rosenberg, 176–240. New York: Cambridge University Press.

Fordham, Finn. 2010. I Do, I Undo, I Redo: The Textual Genesis of Modernist Selves in Hopkins, Yeats, Conrad, Forster, Joyce, and Woolf. Oxford ; New York: Oxford University Press.

Fürer, Mathias. 2017. 'Modeling, Scaling and Sequencing Writing Phases of Swiss Television Journalists'. Doctor of Philosophy, Bern: University of Bern.

Galbraith, David. 1999. 'Writing as a Knowledge-Constituting Process'. In *Knowing What to Write: Conceptual Processes in Text Production*, edited by Mark Torrance and David Galbraith, 139–60. Amsterdam: Amsterdam University Press.

———. 2009. 'Writing as Discovery'. *British Journal of Educational Psychology* 2 (6): 5–26. https://doi.org/10.1348/978185409X421129.

———. 2022. 'The Dual Process Model of the Development of Content during Writing'. Paper given at the Track Changes Symposium. Amsterdam, April 28.

Galbraith, David, and Veerle Baaijen. 2015. 'Conflict in Writing: Actions and Objects'. In *Writing(s) at the Crossroads*, edited by Georgeta Cislaru, 253–76. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.194.13gal.

———. 2018. 'The Work of Writing: Raiding the Inarticulate'. *Educational Psychologist* 53 (4): 238–57. https://doi.org/10.1080/00461520.2018.1505515.

Ganascia, Jean-Gabriel, Irène Fenoglio, and Jean-Louis Lebrave. 2004. 'Manuscrits, Genèse et Documents Numérisés. EDITE : Une Étude Informatisée Du Travail de l'écrivain'. *Document Numérique* 8 (4): 91–110. https://doi.org/10.3166/dn.8.4.91-110.

Gelman, Annelyse. n.d. 'Midst'. Midst. n.d. http://www.midst.press.

Gooding, Paul, Jos Smith, and Justine Mann. 2019. 'The Forensic Imagination: Interdisciplinary Approaches to Tracing Creativity in Writers' Born-Digital Archives'. *Archives and Manuscripts* 47 (3): 374–90. https://doi.org/10.1080/01576895.2019.1608837.

Grésillon, Almuth. 1994. *Éléments de Critique Génétique: Lire Les Manuscrits Modernes*. Paris: Presses universitaires de France.

———. 1997. 'Slow: Work in Progress'. *Word & Image* 13 (2): 106–23. https://doi.org/10.1080/02666286.1997.10434276.

Grésillon, Almuth, and Daniel Perrin. 2015. 'Methodology: Investigating Real-Life Writing Processes'. In *Writing(s) at the Crossroads*, edited by Georgeta Cislaru, 33–54. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.194.03gre.

Haas, Christina. 1996. *Writing Technology: Studies on the Materiality of Literacy*. Mahwah, N.J: L. Erlbaum Associates.

Hammond, Adam. 2016. *Literature in the Digital Age: An Introduction*. Cambridge: Cambridge University Press.

Hammond, Ray. 1984. *The Writer and the Word Processor: A Guide for Authors, Journalists, Poets and Playwrights*. London: Coronet.

Hasselt, Hester van, and Bianca Sistermans. 2021. *Een Mogelijk Begin van Veel: 29 Dichters Aan Het Werk*. Amsterdam: Em. Querido's Uitgeverij BV.

Hay, Louis, and Jean-Louis Lebrave. 2010. 'Lettres Ouvertes: Une Génétique sans Rivages ? (Par Louis Hay) Post-Scriptum – L'ordinateur, Olympe de l'écriture ? (Par Jean-Louis Lebrave)'. *Genesis*, 31 (October): 151–55. https://doi.org/10.4000/genesis.386.

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*, edited by C. Michael Levy and S. Randall, 1–27. Mahwah, N.J: Erlbaum.

———. 2004. 'What Triggers Revision?' In *Revision Cognitive and Instructional Processes*, edited by Linda Allal, Lucile Chanquoy, and Pierre Largy, 13:9–20. Studies in Writing. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-1048-1_2.

———. 2012. 'Modeling and Remodeling Writing'. *Written Communication* 29 (3): 369–88. https://doi.org/10.1177/0741088312451260.

Hayes, John R., and Linda Flower. 1980. 'Identifying the Organization of Writing Processes'. In *Cognitive Processes in Writing*, edited by Lee W. Gregg and Erwin R. Steinberg, 3–30. Hillsdale, N.J: Erlbaum.

Hayes, John R., and Jane Gradwohl Nash. 1996. 'On the Nature of Planning in Writing'. In *The Science of Writing: Theories, Methods, Individual Differences, and Applications*, edited by C. Michael Levy and S. Randall, 29–55. Mahwah, N.J: Erlbaum.

Hayles, N. Katherine. 2003. 'Translating Media: Why We Should Rethink Textuality'. *The Yale Journal of Criticism* 16 (2): 263–90. https://doi.org/10.1353/yale.2003.0018.

———. 2008. *Electronic Literature: New Horizons for the Literary*. The Ward-Phillips Lectures in English Language and Literature. Notre Dame, Ind: University of Notre Dame.

Heeks, Richard James. 2012. 'Discovery Writing and Genre'. Doctor of Philosophy in Education, Exester: University of Exester.

Heerma van Voss, Thomas. 2020. *Condities*. Eerste druk. Amsterdam: Das Mag Uitgevers.

Hentea, Marius. 2010. 'The Forms and Functions of Back Story in the Novel'. *Narrative* 18 (3): 347–66. https://doi.org/10.1353/nar.2010.0001.

Herrmann, Berenike J., Karina van Dalen-Oskam, and Christof Schöch. 2015. 'Revisiting Style, a Key Concept in Literary Studies'. *Journal of Literary Theory* 9 (1). https://doi.org/10.1515/jlt-2015-0003.

Hoffmann, Hein, dir. 2003. 'Opkomst Computer'. *Andere Tijden*. NTR/VPRO. https://www.youtube.com/watch?v=SIJfmffun4c&list=WL&index=1&t=1s.

Hogan, Patrick Colm. 2013. *How Authors' Minds Make Stories*. New York (N.Y.): Cambridge university press.

Hoover, David L. 2021. *Modes of Composition and the Durability of Style in Literature*. New York, NY Lonon: Routledge, Taylor & Francis Group.

Iser, Wolfgang. 1974. *The Implied Reader: Patterns of Communication in Prose Fiction from Bunyan to Beckett*. Baltimore: Johns Hopkins University Press.

Johansson, Victoria. 2022. 'The Poetic Writing Process: A Case Study of Text Composing by Professional Poets'. Paper given at the SIG Writing 2022 Conference. Umeå, June 20-22.

Johncock, Ben. 2011a. 'Tom McCarthy: My Desktop'. *The Guardian*, 24 November 2011, sec. Books. https://www.theguardian.com/books/2011/nov/24/tom-mccarthy-desktop.

———. 2011b. 'Julie Myerson: My Desktop'. *The Guardian*, 8 December 2011, sec. Books. https://www.theguardian.com/books/2011/dec/08/julie-myerson-my-desktop.

———. 2012a. 'Louise Doughty: My Desktop'. *The Guardian*, 11 January 2012, sec. Books. https://www.theguardian.com/books/2012/jan/11/louise-doughty-my-desktop.

———. 2012b. 'Joanne Harris: My Desktop'. *The Guardian*, 27 January 2012, sec. Books. https://www.theguardian.com/books/2012/jan/27/joanne-harris-my-desktop-twitter-computer.

———. 2012c. 'Jon McGregor: My Desktop'. *The Guardian*, 20 June 2012, sec. Books. https://www.theguardian.com/books/2012/jun/20/jon-mcgregor-my-desktop-photos.

———. 2013. 'Nikesh Shukla: My Desktop'. *The Guardian*, 5 June 2013, sec. Books. https://www.theguardian.com/books/2013/jun/05/nikesh-shukla-my-desktop.

Kellogg, Ronald T. 1996. 'A Model of Working Memory in Writing'. In *The Science of Writing: Theories, Methods, Individual Differences, and Applications*, edited by C. Michael Levy and Sarah E. Ransdell, 57–71. Mahwah, N.J.: L. Erlbaum.

———. 2008. 'Training Writing Skills: A Cognitive Developmental Perspective'. *Journal of Writing Research* 1 (1): 1–26. https://doi.org/10.17239/jowr-2008.01.01.1.

Kirschenbaum, Matthew G. 2008. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, Mass: MIT Press.

———. 2012. *Mechanisms: New Media and the Forensic Imagination*. 1. paperback ed. Cambridge, Mass: MIT Press.

———. 2016. *Track Changes: A Literary History of Word Processing*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

———. 2021. *Bitstreams: The Future of Digital Literary Heritage*. Material Texts. Philadelphia: University of Pennsylvania Press.

Kirschenbaum, Matthew G., Erika L. Farr, Kari M. Kraus, Naomi L. Nelson, Catherine Stollar Peters, Gabriela Redwine, and Doug Reside. 2009. 'Digital Materiality: Preserving Access to Computers as Complete Environments'. *UC Office of the President: California Digital Library*, iPress 2009. The sixth international conference on the preservation of digital objects, 105–11.

Kirschenbaum, Matthew G., and Doug Reside. 2013. 'Tracking the Changes: Textual Scholarship and the Challenge of the Born Digital'. In *The Cambridge Companion to Textual Scholarship*, edited by Neil Fraistat and Julia Flanders, 257–88. Cambridge: Cambridge University Press.

Knopse, Yvonne, Kirk P.H. Sullivan, Anita Malmqvist, and Ingela Valfridsson. 2019. 'Observing Writing and Website Browsing: Swedish Students Write L3 German'. In *Observing Writing*, edited by Eva Lindgren and Kirk Sullivan, 38:258–84. Studies in Writing. Leiden: Brill. https://doi.org/10.1163/9789004392526_013.

Koenders, Edith, and Paula Stevens. 2017. 'Het vertalen met fictieve kunst en plattegronden van Johan Harstads Max, Mischa & het Tet-offensief'. Athenaeum Boekhandel. 12 July 2017.

https://www.athenaeum.nl/nieuws/vertalers/2017/het-vertalen-van-johan-harstads-max-mischa-het-tet-offensief.

Kollberg, Py. 1998. 'S-Notation – A Computer Based Method for Studying and Representing Text Composition'. Lic. thesis, Stockholm: University of Stockholm.

Kollberg, Py, and Kerstin S. Eklundh. 2002. 'Studying Writers' Revising Patterns with S-Notation Analysis'. In *Contemporary Tools and Techniques for Studying Writing*, edited by Thierry Olive and C. Michael Levy, 10:89–104. Studies in Writing. Dordrecht: Springer.

Krementz, Jill. 1996. *The Writer's Desk*. 1st ed. New York: Random House.

Leblay, Christophe. 2011. 'Le Temps de l'écriture. Genèse, Durée, Représentations'. Doctor of Philosophy, Jyväskylä: University of Jyväskylä.

Leblay, Christophe, and Gilles Caporossi. 2014. 'Introduction aux données temporelles de l'écriture'. In *Temps de l'écriture: enregistrements et représentations*, edited by Christophe Leblay and Gilles Caporossi, 5–15. Sciences du langage, carrefours et points de vue, n° 12. Louvain-la-Neuve (Belgique) [Paris]: Academia l'Harmattan.

Lebrave, Jean-Louis. 2011. 'Computer Forensics : La Critique Génétique et l'écriture Numérique'. *Genesis*, 33 (October): 137–47. https://doi.org/10.4000/genesis.633.

Leijten, Mariëlle, and Luuk Van Waes. 2013. 'Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes'. *Written Communication* 30 (3): 358–92. https://doi.org/10.1177/0741088313491692.

Leijten, Mariëlle, Luuk Van Waes, and Sarah Ransdell. 2010. 'Correcting Text Production Errors: Isolating the Effects of Writing Mode From Error Span, Input Mode, and Lexicality'. *Written Communication* 27 (2): 189–227. https://doi.org/10.1177/0741088309359139.

Leijten, Mariëlle, Luuk Van Waes, Iris Schrijver, Sarah Bernolet, and Lieve Vangehuchten. 2019. 'Mapping Master's Students' Use of External Sources in Source-Based Writing in L1 and L2'. *Studies in Second Language Acquisition* 41 (3): 555–82. https://doi.org/10.1017/S0272263119000251.

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): 285–337. https://doi.org/10.17239/jowr-2014.05.03.3.

Leonardi, Paul M. 2010. 'Digital Materiality? How Artifacts without Matter, Matter'. *First Monday* 15 (6). https://doi.org/10.5210/fm.v15i6.3036.

Levy, Deborah. 2019. *The Cost of Living*. London: Penguin Books.

Light, Michelle. 2014. 'Managing Risk with a Virtual Reading Room: Two Born Digital Projects'. In *Reference and Access Innovative Practices for Archives and Special Collections*, edited by Kate Theimer, 17–35. Lanham, MD: Rowman & Littlefield Publishers. https://digitalscholarship.unlv.edu/lib_articles/463.

Lindgren, Eva, Yvonne Knopse, and Kirk P.H. Sullivan. 2019. 'Researching Writing with Observational Logging Tools from 2006 to the Present'. In *Observing Writing*, edited by Eva Lindgren and Kirk P.H. Sullivan, 1–29. Studies in Writing 38. Leiden: Brill. https://doi.org/10.1163/9789004392526_002.

Lindgren, Eva, and Kirk P.H. Sullivan. 2006a. 'Analysing On-Line Revision'. In *Computer Keystroke Logging: Methods and Applications*, edited by Kirk P.H. Sullivan and Eva Lindgren, 157–88. Studies in Writing 18. Oxford: Elsevier.

———. 2006b. 'Writing and the Analysis of Revision: An Overview'. In *Computer Keystroke Logging and Writing: Methods and Applications*, edited by Kirk P.H. Sullivan and Eva Lindgren, 18:31–44. Studies in Writing. Leiden: Brill. https://doi.org/10.1163/9780080460932_004.

———, eds. 2019. *Observing Writing: Insights from Keystroke Logging and Handwriting*. Vol. 38. Studies in Writing. Leiden ; Boston: Brill.

Lindgren, Eva, Asbjørg Westum, Hanna Outakoski, and Kirk P.H. Sullivan. 2019. 'Revising at the Leading Edge: Shaping Ideas or Clearing up Noise'. In *Observing Writing*, edited by Eva Lindgren and Kirk P.H. Sullivan, 346–65. Leiden: Brill. https://doi.org/10.1163/9789004392526_017.

Linkès, Serge. 2017. 'Du Manuscrit Moderne Au Brouillon Nmérique: Un Nouveau Défi Pour La Critique Génétique?' In *L'Œuvre Comme Processus*, edited by Pierre-Marc de Biasi and Anne Herschberg-Pierrot, 471–79. Collection 'Génétique'. Paris: CNRS éditions.

Linnemann, Markus. 2019. 'Anticipation of Audience during Writing'. In *Observing Writing*, edited by Eva Lindgren and Kirk Sullivan, 38:326–45. Studies in Writing. Leiden: Brill. https://doi.org/10.1163/9789004392526_016.

MacArthur, Charles A., and Steve Graham. 2017. 'Writing Research from a Cognitive Perspective'. In *Handbook of Writing Research*, edited by Charles A. MacArthur, Steve Graham, and Jill Fitzgerald, Second edition, 24–40. New York London: The Guilford Press.

Mahlow, Cerstin. 2015. 'A Definition of "Version" for Text Production Data and Natural Language Document Drafts'. In *Proceedings of the 3rd International Workshop on (Document) Changes: Modeling, Detection, Storage and Visualization*, 27–32. Lausanne Switzerland: ACM. https://doi.org/10.1145/2881631.2881638.

Mahlow, Cerstin, Malgorzata Anna Ulasik, and Don Tuggener. 2022. 'Extraction of Transforming Sequences and Sentence Histories from Writing Process Data:

A First Step towards Linguistic Modeling of Writing'. *Reading and Writing*, January. https://doi.org/10.1007/s11145-021-10234-6.

Mahrer, Rudolf, Rossana De Angelis, Andrea Del Lungo, Almuth Grésillon, Jean-Louis Lebrave, Valentine Nicollier Saraillon, Thierry Poibeau, Frédérique Mélanie-Becquet, and Bénédicte Vauthier. 2015. 'Editorial Genesis: From Comparing Texts (Product) to Interpreting Rewritings (Process)'. In *Writing(s) at the Crossroads*, edited by Georgeta Cislaru, 151–70. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.194.08mah.

Malvik, Anders Skare, and Sarah J. Paulson. 2016. 'Introduction: Technology - Subjectivity - Aesthetics: Three Perspectives on Contemporary Media Culture'. In *Literature in Contemporary Media Culture: Technology - Subjectivity - Aesthetics*, edited by Sarah J. Paulson and Anders Skare Malvik, 1–20. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Mangen, Anne, and Jean-Luc Velay. 2010. 'Digitizing Literacy: Reflections on the Haptics of Writing'. In *Advances in Haptics*, edited by Mehrdad Hosseini. InTech. https://doi.org/10.5772/8710.

Manovich, Lev. 2001. *The Language of New Media*. 8. print. A Leonardo Book. Cambridge, Mass.: MIT Press.

Mara, Miriam O'Kane. 2013. 'Nuala O'Faolain: New Departures in Textual and Genetic Criticism'. *Irish Studies Review* 21 (3): 342–52. https://doi.org/10.1080/09670882.2013.808873.

Mathijsen, Marita. 2009. 'Genetic Textual Editing: The End of an Era'. In *Was Ist Textkritik? Zur Geschichte Und Relevanz Eines Zentralbegriffs Der Editionswissenschaft*, edited by Gertraud Mitterauer, Werner Maria Bauer, and Sabine Hofer. Beihefte zu Editio, Bd. 28. Tübingen: Niemeyer.

———. 2010. *Naar de letter: handboek editiewetenschap*. 4e ongew. opl. Amsterdam: KNAW Press.

Matsuhashi, Ann. 1981. 'Pausing and Planning: The Tempo of Written Discourse Production'. *Research in the Teaching of English* 15 (2): 113–34.

———, ed. 1987. 'Revising the Plan and Altering the Text'. In *Writing in Real Time: Modelling Production Processes*, 197–223. Writing Research. Norwood, N.J: Ablex Pub. Corp.

Meines, Myrthe. 2022. '21 vragen aan... Roos van Rijswijk'. De Groene Amsterdammer. 2022. https://www.groene.nl/artikel/21-vragen-aan-roos-van-rijswijk.

Micunovic, Milijana, Hana Marčetić, and Maja Krtalić. 2016. 'Literature and Writers in the Digital Age: A Small-Scale Survey of Contemporary Croatian Writers' Organization and Preservation Practices'. *Preservation, Digital Technology & Culture* 45 (1): 2–16. https://doi.org/10.1515/pdtc-2015-0028.

Molloy, Kevin. 2019. 'Literary Archives in the Digital Age: Issues and Encounters with Australian Writers'. *Archives and Manuscripts* 47 (3): 327–42. https://doi.org/10.1080/01576895.2019.1631863.

Nelson, Maggie. 2016. *The Argonauts*. First Graywolf paperback. Minneapolis, Minnesota: Graywolf Press.

NNI. n.d. 'Just How Small Is "Nano"? | National Nanotechnology Initiative'. Accessed 12 September 2022. https://www.nano.gov/about-nanotechnology/just-how-small-is-nano.

Page, Tom. 2014. 'Skeuomorphism or Flat Design: Future Directions in Mobile Device User Interface (UI) Design Education'. *Int. J. Mobile Learning and Organisation* 8 (2): 130–42.

Peeters, Koen. 2007. *Grote Europese Roman*. Antwerpen: Meulenhoff/Manteau.

Pereira, Elsa. 2021. 'The Concept of Version in Genetically Oriented Scholarly Editing'. *Studia Neophilologica* 93 (1): 107–19. https://doi.org/10.1080/00393274.2021.1885306.

Perrin, Daniel. 2003. 'Progression Analysis (PA): Investigating Writing Strategies at the Workplace'. *Journal of Pragmatics* 35 (6): 907–21. https://doi.org/10.1016/S0378-2166(02)00125-X.

Pierazzo, Elena. 2009. 'Digital Genetic Editions: The Encoding of Time in Manuscript Transcription'. In *Text Editing, Print and the Digital World*, edited by Marilyn Deegan and Kathryn Sutherland, 169–86. Farnham: Ashgate.

———. 2022. 'A Time for Writing, a Time for Encoding'. Paper given at the Track Changes Symposium. Amsterdam, April 28.

Pierazzo, Elena, and Peter A. Stokes. 2011. 'Putting the Text Back into Context: A Codicological Approach to Manuscript Transcription'. In *Codicology and Palaeography in the Digital Age 2*, edited by Franz Fischer, Christiane Fritze, and Georg Vogeler, 3:397–429. Schriften Des Instituts Für Dokumentologie Und Editorik. Norderstedt: Books on Demand (BoD).

Plane, Sylvie, Denis Alamargot, and Jean-Louis Lebrave. 2010. 'Temporalité de l'écriture et Rôle Du Texte Produit Dans l'activité Rédactionnelle:' *Langages* 177 (1): 7–28. https://doi.org/10.3917/lang.177.0007.

Posthuma, Jente. 2021. 'En Daarom Haten Ze Zichzelf'. *De Gids*, 2021.

Quinlan, Thomas, Maaike Loncke, Mariëlle Leijten, and Luuk Van Waes. 2012. 'Coordinating the Cognitive Processes of Writing: The Role of the Monitor'. *Written Communication* 29 (3): 345–68. https://doi.org/10.1177/0741088312451112.

Reside, Doug. 2014. 'File Not Found: Rarity in an Age of Digital Plenty'. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 15 (1): 68–74. https://doi.org/10.5860/rbm.15.1.416.

Ries, Thorsten. 2010. '„die Geräte Klüger Als Ihre Besitzer": Philologische Durchblicke Hinter Die Schreibszene Des Graphical User Interface'. *Editio* 24 (2010): 149–99. https://doi.org/10.1515/9783110223163.0.149.

———. 2017. 'Philology and the Digital Writing Process'. Edited by Reindert Dhondt and David Martens. *Genrehybriditeit in de Literatuur*, Cahier voor literatuurwetenschap, 9: 129–58.

———. 2018. 'The Rationale of the Born-Digital Dossier Génétique: Digital Forensics and the Writing Process: With Examples from the Thomas Kling Archive'. *Digital Scholarship in the Humanities* 33 (2): 391–424. https://doi.org/10.1093/llc/fqx049.

Rijswijk, Roos van. 2021. *De Dwaler*. Amsterdam ; Antwerpen: Em. Querido's Uitgeverij BV.

Saki. 1910. Reginald in Russia and Other Sketches. London: Methuen & Co.

———. 1983. *De Onrustkuur. Alle Verhalen*. Translated by Stella Bromet. Amsterdam: Bert Bakker.

Samyn, Laurent. 2011. 'Piraterij: Het Jack Sparrow Complex'. Master Thesis, Ghent: University of Ghent.

Schäuble, Joshua, and Hans Walter Gabler. 2018. 'The Editor in the Interface: Guiding the User through Texts and Images'. In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 12:165–91. Schriften Des Instituts Für Dokumentologie Und Editorik. Norderstedt: Books on Demand (BoD).

Scheibe, Siegfried. 1995. 'Theoretical Problems of the Authorization and Constitution of Texts'. In *Contemporary German Editorial Theory*, edited by Hans Walter Gabler, George Bornstein, and Gillian Borland Pierce, 171–92. Ann Arbor: The University of Michigan Press.

Schiller, Nicholas, and Dene Grigar. 2019. 'Born Digital Preservation of E-Lit: A Live Internet Traversal of Sarah Smith's King of Space'. *International Journal of Digital Humanities* 1 (1): 47–57. https://doi.org/10.1007/s42803-019-00004-w.

Schmidt, Desmond. 2019. 'A Model of Versions and Layers'. *Digital Humanities Quarterly* 13 (3).

Schmitz, Dawn. 2010. 'The Born-Digital Manuscript as Cultural Form and Intellectual Record'. In *UC Irvine: Libraries*. Irvine, CA. https://escholarship.org/uc/item/0609q36r.

Schrijver, Iris, and Luuk Van Waes. 2022. 'Als Manuscripten van de Wereld Verdwijnen. Een Digitale Blik in de Keuken van de Literaire Vertaler'. *WebFilter*, no. Literair vertalen en technologie. https://www.tijdschrift-filter.nl/webfilter/dossier/literair-vertalen-en-technologie/januari-2021/als-manuscripten-van-de-wereld-verdwijnen/.

Schuessler, Jennifer. 2011. 'The Muses of Insert, Delete and Execute'. *The New York Times*, 25 December 2011, sec. Books. https://www.nytimes.com/2011/12/26/books/a-literary-history-of-word-processing.html.

Shillingsburg, Peter L. 1996. *Scholarly Editing in the Computer Age: Theory and Practice*. 3rd ed. Ann Arbor: The University of Michigan Press.

Sommers, Nancy. 1980. 'Revision Strategies of Student Writers and Experienced Adult Writers'. *College Composition and Communication* 31 (4): 378. https://doi.org/10.2307/356588.

Spelman Miller, Kristyan, and Kirk P.H. Sullivan. 2006. 'Keystroke Logging: An Introduction'. In *Computer Keystroke Logging and Writing: Methods and Applications*, edited by Kirk P. H. Sullivan and Eva Lindgren, 18:1–9. Studies in Writing. Leiden: Brill. https://doi.org/10.1163/9780080460932_002.

Stevenson, Marie, Rob Schoonen, and Kees de Glopper. 2006. 'Revising in Two Languages: A Multi-Dimensional Comparison of Online Writing Revisions in L1 and FL'. *Journal of Second Language Writing* 15 (3): 201–33. https://doi.org/10.1016/j.jslw.2006.06.002.

Steyaert, Katrien. 2014. 'Sabien Clement'. *De Standaard*, 28 March 2014, sec. Boeken. https://www.standaard.be/cnt/dmf20140327_01043872.

Stiller, Louis. 2014. *eSchrijven*. Antwerpen: Atlas Contact.

Sullivan, Hannah. 2013. *The Work of Revision*. Cambridge, Massachusetts: Harvard University Press.

Sullivan, Kirk P. H., and Eva Lindgren, eds. 2006. *Computer Keystroke Logging and Writing: Methods and Applications*. Vol. 18. Studies in Writing. Amsterdam: Elsevier.

Swaya, Matthew E., and Stacey R. Eisenstein. 2005. 'Emerging Technology in the Workplace'. *Labor Lawyer* 21 (1): 1–18.

Taylor, Craig. 2018. 'C M Taylor on "Keystroke Logging Project" with British Library'. British Library - English and Drama Blog. 9 November 2018. https://blogs.bl.uk/english-and-drama/2018/11/c-m-taylor-on-keystroke-logging-project-with-british-library.html.

The TEI Consortium. 2020. 'TEI P5: Guidelines for Electronic Text Encoding and Interchange'. Text Encoding Initiative Consortium. https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf.

Tillema, Marion, Huub van den Bergh, Gert Rijlaarsdam, and Ted Sanders. 2011. 'Relating Self Reports of Writing Behaviour and Online Task Execution Using a Temporal Model'. *Metacognition and Learning* 6 (3): 229–53. https://doi.org/10.1007/s11409-011-9072-x.

Usoof, Hakim, Christophe Leblay, and Gilles Caporossi. 2020. 'GenoGraphiX-Log Version 2.0 User Guide'. Technical Report G–2020–68. Les Cahiers Du GERAD G–2020–68. Montréal, Canada: GERAD. https://www.gerad.ca/en/papers/G-2020-68.

Usoof, Hakim, and Eva Lindgren. 2008. 'Who Is Who and Doing What in Distance Education?: Authentication and Keystroke Dynamics'. *Tidskrift För Lärarutbildning Och Forskning* 15 (3–4): 175–87.

Van Hulle, Dirk. 2004. *Textual Awareness: A Genetic Study of Late Manuscripts by Joyce, Proust, and Mann*. Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.17739.

———. 2006. 'Textual Descent with Intentional Modification. Genetic Criticism and Genetic Modification'. In The Evolution of Texts: Confronting Stemmatological and Genetical Methods ; Proceedings of the International Workshop Held in Louvain-La Neuve on September 1-2, 2004, edited by Caroline Macé, Philippe Baret, Andrea Bozzi, and Laura Cignoni, 215–28. Linguistica Computazionale 24/25. Pisa: Istituti editoriali e poligrafici internazionali.

———. 2007. *De Kladbewaarders*. Nijmegen: Vantilt.

———. 2013. 'The Stuff of Fiction: Digital Editing, Multiple Drafts and the Extended Mind'. *Textual Cultures* 8 (1): 23–37. https://doi.org/10.14434/TCv8i1.5048.

———. 2014. Modern Manuscripts: The Extended Mind and Creative Undoing from Darwin to Beckett and Beyond. London: Bloomsbury Academic.

———. 2016. 'Modelling a Digital Scholarly Edition for Genetic Criticism: A Rapprochement'. *Variants* 12–13: 34–56.

———. 2017. 'Essai de Critique Épigénétique'. In *L'Œuvre Comme Processus*, edited by Pierre-Marc de Biasi and Anne Herschberg-Pierrot, 415–30. Collection 'Génétique'. Paris: CNRS éditions.

———. 2019. 'De Logica van de Tekstversie in Digitaal Geschreven Literatuur'. *Tijdschrift Voor Nederlandse Taal En Letterkunde/Journal of Dutch Linguistics and Literature* 135 (4): 465–77. https://doi.org/10.5117/TNTL2019.4.013.HULL.

———. 2021. 'Dynamic Facsimiles: Note on the Transcription of Born-Digital Works for Genetic Criticism'. *Variants*, 15–16: 231–41. https://doi.org/10.4000/variants.1450.

———. 2022. *Genetic Criticism: Tracing Creativity in Literature*. Oxford, United Kingdom: Oxford University Press.

Van Hulle, Dirk, and Peter Shillingsburg. 2015. 'Orientations to Text, Revisited'. *Studies in Bibliography* 59 (1): 27–44. https://doi.org/10.1353/sib.2015.0004.

Van Mierlo, Wim. 2013. 'The Archaeology of the Manuscript: Towards Modern Palaeography'. In *The Boundaries of the Literary Archive: Reclamation and*

Representation, edited by Carrie Smith and Lisa Stead, 15–29. Farnham, Surrey, England ; Burlington, VT: Ashgate.

———. 2020. 'Vision and Revision in the Manuscripts of William Wordsworth and W. B. Yeats'. In *Genesis and Revision in Modern British and Irish Writers*, edited by Jonathan Bloom and Catherine Rovera, 17–36. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-50277-5_2.

Van Waes, Luuk, and Mariëlle Leijten. 2020. 'Measuring and Assessing Typing Skills in Writing Research'. *Chinese Journal of Second Language Writing* 1 (1): 18–39.

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): 829–53. https://doi.org/10.1016/S0378-2166(02)00121-2.

Van Wing, Sage. 2022. 'Author Richard Powers on His Novel "Bewilderment"'. OPB. 2022. https://www.opb.org/article/2022/04/20/author-richard-powers-on-his-novel-bewilderment/.

Vandermeulen, Nina, Brenda van den Broek, Elke Van Steendam, and Gert Rijlaarsdam. 2020. 'In Search of an Effective Source Use Pattern for Writing Argumentative and Informative Synthesis Texts'. *Reading and Writing* 33 (2): 239–66. https://doi.org/10.1007/s11145-019-09958-3.

Vásári, Melinda. 2019. 'Securing the Literary Evidence. Some Perspectives on Digital Forensics'. In *Philology in the Making: Analog/Digital Cultures of Scholarly Writing and Reading*, edited by Pál Kelemen and Nicolas Pethes, 287–309. Digital Humanities, volume 1. Bielefeld: Transcript.

Vauthier, Bénédicte. 2016. 'Genetic Criticism Put to the Test by Digital Technology: Sounding out the (Mainly) Digital Genetic File of El Dorado'. *Variants*, 12–13: 163–86. https://doi.org/10.4000/variants.353.

———. 2019. 'The .Txtual Condition, .Txtual Criticism and .Txtual Scholarly Editing in Spanish Philology'. *International Journal of Digital Humanities* 1 (1): 29–46. https://doi.org/10.1007/s42803-019-00003-x.

Verso Books, dir. 2019. *Andrea Long Chu on Desire*. https://www.youtube.com/watch?v=6CFHRw79CrY.

Wason, P.C. 1980. 'Specific Thoughts on the Writing Process'. In *Cognitive Processes in Writing*, edited by Lee W. Gregg and Erwin R. Steinberg, 129–37. Hillsdale, N.J: Lawrence Erlbaum Associates.

Wechgelaer, Jikke, dir. 2021. 'Rob van Essen, Kiza Magendane, Ronald Snijders En Niña Weijers'. *Brommer Op Zee*. NPO.

Weisbrod, Dirk. 2015. 'Die Präkustodiale Intervention Als Baustein Der Langzeitarchivierung Digitaler Schriftstellernachlässe'. Doctor of Philosophy, Berlin: Humboldt-Universität Berlin.

———. 2016. 'Cloud-Supported Preservation of Digital Papers: A Solution for Special Collections?' *Liber Quarterly* 25 (3): 136–51. https://doi.org/10.18352/lq.10114.

Wengelin, Åsa, Johan Frid, Roger Johansson, and Victoria Johansson. 2019. 'Combining Keystroke Logging with Other Methods: Towards an Experimental Environment for Writing Process Research'. In *Observing Writing*, edited by Eva Lindgren and Kirk Sullivan, 38:30–49. Studies in Writing. Leiden: Brill. https://doi.org/10.1163/9789004392526_003.

Wengelin, Åsa, Mariëlle Leijten, and Luuk Van Waes. 2010. 'Studying Reading during Writing: New Perspectives in Research'. *Reading and Writing* 23 (7): 735–42. https://doi.org/10.1007/s11145-009-9187-5.

Wengelin, Åsa, Mark Torrance, Kenneth Holmqvist, Sol Simpson, David Galbraith, Victoria Johansson, and Roger Johansson. 2009. 'Combined Eyetracking and Keystroke-Logging Methods for Studying Cognitive Processes in Text Production'. *Behavior Research Methods* 41 (2): 337–51. https://doi.org/10.3758/BRM.41.2.337.

Wix, Gabriele. 2016. 'Stratigraphic Soundings: A Genetic Approach to the German Poet Thomas Kling'. *Variants*, 12–13: 125–47. https://doi.org/10.4000/variants.334.

Workgroup on Genetic Editions. 2010. 'An Encoding Model for Genetic Editions'. https://tei-c.org/Vault/TC/tcw19.html.

Zimmerman, Barry J., and Rafael Risemberg. 1997. 'Becoming a Self-Regulated Writer: A Social Cognitive Perspective'. *Contemporary Educational Psychology* 22 (1): 73–101. https://doi.org/10.1006/ceps.1997.0919.

The digital environment in which present-day literature is composed, significantly changes the materiality of the sources available for genetic criticism, since common word processors tend to hide the writing operations: additions are visualised on our screen as inline text production, and deleted text 'disappears' from the screen. This makes it a difficult endeavour to reconstruct the writing process, but does it herald the feared end of genetic criticism??

Behind the computer screens argues that this will not be the case as long as genetic criticism adapts to working with digital files. As one of the ways in which genetic criticism can adapt to the Digital Age the study examines the use of one method in particular: keystroke logging. To explore the possibilities of keystroke logging, the study analyses the writing processes of Gie Bogaert, Jente Posthuma, Roos van Rijswijk, David Troch, and Ellen Van Pelt. Each of them recorded their writing process using the keystroke logging software Inputlog.

When examining all this keystroke logging data, we can study the writing process at an unprecedented level of granularity, including non-linear behaviour and the sequentiality of text production and revision. The fine-grained data of keystroke logging therefore allows for a new type of what this study calls 'nanogenetic' research.

Nanogenesis allows, among other things, the study of the author's working methods and typing habits of the author, the triggers of text production and revision, aspects of discovery in writing, and how sources are transformed during the writing process. The study also explores how genetic narratology may be enriched by keystroke logging data, as this allows the temporal dimension to be included in the analysis. It examines how and when details of aspects of the narrative text are integrated into the text and how they change during the writing process. When authors are willing to record their writing processes with a keystroke logging tool, the future of genetic criticism seems full of possibilities.