

# Narrative Recomposition in the Context of Digital Reading

THÈSE N° 7592 (2017)

PRÉSENTÉE LE 4 AVRIL 2017  
AU COLLÈGE DES HUMANITÉS  
LABORATOIRE D'HUMANITÉS DIGITALES  
PROGRAMME DOCTORAL EN MANAGEMENT DE LA TECHNOLOGIE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Cyril Antoine Michel BORNET

acceptée sur proposition du jury:

Prof. D. Foray, président du jury  
Prof. F. Kaplan, directeur de thèse  
Dr L. Allard, rapporteuse  
Prof. R. Baroni, rapporteur  
Prof. M. Finger, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2017



Au secours, cher romancier,  
faites-moi sortir de votre monde.  
Je veux retourner au paradis du présent.  
— Kumo vom Pokk

To writing and digitization, arguably the greatest breakthroughs to knowledge...



# Acknowledgements

This research was supported by the *Fondation Jan Michalski pour l'Écriture et la Littérature*, and made possible with a very precious and close collaboration with **Daniel de Roulet**, Swiss architect, computer scientist, author, and master of the art of digital recomposition.

I would also like to thank:

My advisor, Prof. **Frédéric Kaplan**, for his help and guidance,

3×B & co. – **Mom, dad**, and my sister **Camille** for their unlimited support during all that time. Unlike fictional characters, I didn't get to choose you – but had it been the case, I would have wanted you not the least bit different!

**Maude Auderset** and **Sophia Ng**, for our numerous talks related to my work, **Vincent Barbay**, for his help to run the show while I was most busy, and **Trần Thị Thùy Linh**, **Imran Quaine**, **Thierry Coppey** and **Fang Liu**, for being incredible friends that were always there when I needed it the most,

**Antoinette Rychner**, **Maxime Maillard** and **Vincent Yersin**, for braving a snow storm to be the first volunteers to try out *Saga+*, and for their precious suggestions, as well as **Pierre Lukaszewski**, **José Redard** and the whole team of the Fondation Jan Michalski.

**EPFL**, in particular the very welcoming staff of Management of Technology Doctoral School (**EDMT**), and, least but not least, all my colleagues from the Digital Humanities Laboratory (**DHLAB**), for being a great team and for their continuous valuable inputs.

*Lausanne, December 2016*

C. B.



# Preface

Between those who write novels and those who write computer programs, the gap continues to widen. Three and a half years ago, when I met Cyril Bornet for the first time, I made the bet that, despite this gap and the half century between us, we would succeed in getting along. And that this agreement would be something original even if I did not know exactly what.

Every time I came to the EPFL campus, his enthusiasm conquered me. At each work session, I learned something. I took advantage of his knowledge, we rubbed our different cultures, that of the screen against that of the pen.

I remember that at first I was quite reluctant when he treated my literary corpus just as a database, containing names, verbs, syntactic indexes. I too had to admit that a sentence could be labeled, that a chapter could be passed to the statistics mill. For his part, the young doctoral student sometimes had to hesitate about his method when he discovered, in one of my chapters, a construction that resisted his algorithms. Little by little our points of view came closer, even if staying unique in their own ways.

Digital humanities promise a lot: to put at the disposal of researchers such efficient tools that each of the soft sciences will be transformed. From geography to sociology to literature, each of these disciplines will finally have the tools it dreamed of. Big data will allow to take into consideration all the literary production of a century and not just one or two recurring authors. We will update new connections between texts, we will embrace entire libraries instead of pecking at a few sentences. This perspective fascinates me, it will change the work of scholars.

But what I did not realize when I first met Cyril Bornet was that it was not only the analysis, but the production of texts that would benefit from these new tools. He offered me a few for my toolbox, I tested them, all interested me. Some of them seemed trivial, others were so promising that I came to think that, if Balzac, Proust or Roger Martin du Gard had them on their desk, their art would have been transformed.

## Preface

---

Science and literature have long coalesced before diverging. For a century, since the advent of movies, TV and smartphones, each of them believes that they alone hold the one way to account for the world. Writers cling to the idea they have of their relation to reality, which could only be filtered by their words correctly arranged on a white sheet. Scientists, and among them the dogmatic ones from computer science, sweep with a flick on their touchscreen the sentimentalism of the writers. It is time for these points of view to return to dialogue even if this dialogue is still terribly fragile. The writer does not want to be searched in his scriptorium and the scientist thinks every text can be understood by his machinery.

Everyone will continue to have doubts about the method of the other. It is the merit of the digital humanities to open a new field, where a research whose subject has long been taboo, can flourish.

The thesis of Cyril Bornet has many merits that his peers will be able to distinguish. For me it has one more. Now those who want to tell stories will be able to count on new tools to do so. Somebody had to open the way. Thanks to Prof. Frederic Kaplan, Cyril Bornet is the pioneer who questions the “new frontier” between computer talk and storytelling.

*Geneva, December 2016*

Daniel de Roulet



# Abstract

In any creative process, the tools one uses have an immediate influence on the shape of the final artwork. However, while the digital revolution has redefined core values in most creative domains over the last few decades, its impact on literature remains limited. This thesis explores the relevance of digital tools for several aspects of novels writing by focusing on two research questions: Is it possible for an author to edit better novels out of already published ones, given the access to adapted tools? And, will authors change their way of writing when they know how they are being read?

This thesis is a multidisciplinary participatory study, actively involving the Swiss novelist Daniel de Roulet, to construct measures, visualizations, and digital tools aimed at leveraging the process of dynamic reordering of narrative material, similar to how one edits a video footage. We developed and tested various text analysis and visualization tools, the results of which were interpreted and used by the author to recompose a family saga out of material he has been writing for twenty-four years. Based on this research, we released *Saga+*, an online editing, publishing, and reading tool. The platform was handed out to third parties to improve existing writings, making new novels available to the public as a result.

While many researchers have studied the structuration of texts either through global statistical features or micro-syntactic analyses, we demonstrate that by allowing visualization and interaction at an intermediary level of organisation, authors can manipulate their own texts in agile ways. By integrating readers' traces into this newly revealed structure, authors can start to approach the question of optimizing their writing processes in ways that are similar to what is being practiced in other media industries. The introduction of tools for optimal composition opens new avenues for authors, as well as a controversial debate regarding the future of literature.

## **Abstract**

---

Key words:

Composition, Digital Humanities, Digital Media, Distant Editing, French Literature, Named Entity Recognition, Narration, Novel Writing, Storytelling, Visualization.

# Zusammenfassung

In jedem kreativen Verfahren haben die verwendeten Werkzeuge einen unmittelbaren Einfluss auf die Form des fertigen Kunstwerks. Während jedoch die digitale Revolution in den letzten Jahrzehnten die Kernwerte in den meisten kreativen Bereichen neu definiert hat, bleibt ihre Wirkung auf die Literatur begrenzt. Diese Arbeit untersucht die Relevanz von digitalen Werkzeugen in Bezug auf mehrere schriftliche Aspekte der Romane, mit der Konzentration auf zwei Forschungsfragen: Ist es für einen Autor möglich, einen besseren Roman als den bereits veröffentlichten zu kreieren, wenn Zugang zu angepassten Werkzeugen zur Bearbeitung besteht? Und werden Autoren ändern, wie sie schreiben, wenn sie wissen, auf welche Weise sie gelesen werden?

Diese These ist eine multidisziplinäre partizipative Studie – an der aktiv der Schweizer Romanautor Daniel de Roulet beteiligt ist – um Maßnahmen, Visualisierungen und digitale Werkzeuge zu konstruieren und um den Prozess der dynamischen Neuordnung von Erzählmaterialien ähnlich wie die Bearbeitung von Videomaterial zu nutzen. Wir haben verschiedene Textanalyse- und Visualisierungswerkzeuge entwickelt und ausprobiert, deren Ergebnisse vom Autor interpretiert und verwendet wurden, um eine Familiensaga aus dem Material, das er seit vierundzwanzig Jahren geschrieben hat, zu erschaffen. Basierend auf dieser Forschung veröffentlichten wir *Saga+*, ein Online-Bearbeitungs-, -Veröffentlichungs- und -Lese-Werkzeug. Die Plattform wurde an Dritte weitergegeben, um bestehende Schriften zu verbessern und dadurch neue Romane der Öffentlichkeit zugänglich zu machen.

Während viele Forscher die Strukturierung von Texten entweder durch globale statistische Merkmale oder mikro-syntaktische Analysen studiert haben, zeigen wir, dass Autoren ihre eigenen Texte auf eine agile Weise manipulieren können, indem sie Visualisierung und Interaktion auf einer intermediären Organisationsebene ermöglichen. Durch die Integration von Leser-Spuren in diese neu aufgedeckte Struktur können Autoren die Art, wie sie schreiben, auf Wegen verbessern, die es zuvor nur in anderen Medien-Industrien gab. Die Einführung von

## **Abstract**

---

Werkzeugen für eine optimale Komposition eröffnet neue Wege für die Autoren sowie ebenso eine Streitkultur über die Zukunft der Literatur.

Stichwörter:

Komposition, Digital Humanities, Digitale Medien, Distant Editing, Französische Literatur, Named Entity Recognition, Narration, Romanschreiben, Geschichtenerzählung, Visualisierung.

# Résumé

Dans tout processus créatif, les outils utilisés exercent une influence immédiate sur le façonnement du produit final. Pourtant, si la révolution numérique a redéfini des valeurs centrales à la plupart des domaines créatifs dans les dernières décennies, son impact sur la littérature est resté pratiquement nul. Cette thèse évalue la pertinence d'outils numériques pour certains aspects de l'écriture romanesque, en se concentrant sur deux questions de recherche : Un écrivain peut-il composer de meilleurs romans que ceux qu'il a déjà publiés s'il a à sa disposition les outils adaptés ? Un écrivain change-t-il sa manière d'écrire quand il sait comment il est lu ?

Cette thèse est une étude participative multidisciplinaire, impliquant activement l'écrivain suisse Daniel de Roulet dans l'établissement de mesures, de graphiques et d'outils numériques visant à faciliter le processus de réorganisation du matériau narratif, à la façon d'un logiciel de montage vidéo. Nous avons développé et testé plusieurs outils d'analyse et de visualisation, dont les résultats ont été interprétés et utilisés par l'auteur pour recomposer une saga à partir de textes précédemment publiés sur une période de vingt-quatre ans. Sur cette base, nous avons mis en ligne *Saga+*, un logiciel d'édition, de publication et de lecture. Cette plateforme a été utilisée par des tiers pour l'amélioration de textes existants, avec pour résultat de nouveaux romans d'ores et déjà mis à disposition du public.

Alors que beaucoup de chercheurs ont étudié la structuration de textes à travers des méthodes statistiques globales ou des analyses micro-syntaxique, nous démontrons que les écrivains peuvent agilement manipuler leurs propres textes par la visualisation et l'interactivité à un niveau d'organisation intermédiaire. En intégrant des statistiques de lecture à cette nouvelle stratégie de structuration, les auteurs peuvent également approcher la question de l'optimalité de leur écriture, comme cela se fait déjà dans d'autres industries médiatiques. L'introduction d'outils de ce genre, visant à une composition optimale, ouvre de nouvelles perspectives aux écrivains ainsi qu'un débat controversé autour du futur de la littérature.

## **Abstract**

---

Mots clefs :

Composition, Distant Editing, Entités nommées, Humanités digitales, Littérature francophone, Médias numériques, Narration, Romance, Visualisation.

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Preface</b>	<b>iii</b>
<b>Abstract (English/Français/Deutsch)</b>	<b>v</b>
<b>List of Figures</b>	<b>xvii</b>
<b>List of Tables</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Tools for Digital Writing and Reading . . . . .	2
<b>Document 1 (Daniel de Roulet, 11.01.2014)</b>	<b>7</b>
<b>2 Context and Methodology</b>	<b>15</b>
2.1 Context of this Research . . . . .	15
2.2 An Interdisciplinary Methodological Framework . . . . .	17
2.3 Participatory Action Research . . . . .	19
	xi

## Contents

---

2.3.1	PAR Feedback Loop . . . . .	19
2.3.2	Researcher Influence Bias . . . . .	20
2.4	Digital Humanities . . . . .	22
2.4.1	Text Mining . . . . .	23
2.4.2	Stylistics, Stylometry and Authorship Attribution . . . . .	25
2.4.3	Visualizations and Visual Analytics . . . . .	26
2.4.4	Distant Reading . . . . .	27
2.4.5	Digital Contents and Software Design . . . . .	29
2.5	Constrained Writing and Sequencing Techniques . . . . .	29
	<b>Document 2 (Daniel de Roulet, 24.03.2014)</b>	<b>35</b>
<b>3</b>	<b>(Re-)ordering Complex Stories</b>	<b>41</b>
3.1	A Narrative Sudoku . . . . .	41
3.2	Paradigms for "Reordered" Digital Sagas . . . . .	44
3.3	Structural Understanding . . . . .	46
3.4	<i>La simulation humaine</i> . . . . .	50
3.4.1	A First Example of Reordered Saga . . . . .	50
3.4.2	A Multimodal Reading Tool . . . . .	52
3.4.3	Applying Web Analytics to Digital Books . . . . .	55
<b>4</b>	<b>Methods for Digital Writing Tools</b>	<b>57</b>



4.1 Distant Editing . . . . .	57
4.2 Shaping Effective Tools for Reorganizing Complex Stories . . . . .	60
4.3 Reliable Recognition of Characters and Places . . . . .	61
4.3.1 Test Procedure . . . . .	61
4.3.2 <i>NER</i> in Novels . . . . .	62
4.3.3 A Set of Simple Rules . . . . .	63
4.3.4 Meta-Classification . . . . .	69
4.3.5 Results . . . . .	72
4.4 Visualizations Based on Named Entities . . . . .	75
4.4.1 Mention Frequencies Histograms . . . . .	76
4.4.2 Characters Networks . . . . .	79
4.4.3 Wording Features Using <i>LIWC</i> . . . . .	83
4.4.4 Entries-Exits Graph . . . . .	86
4.4.5 Timeline Graphs . . . . .	88
<b>Document 3 (Daniel de Roulet, 22.02.2016)</b>	<b>95</b>
<b>5 Introducing a Novel Editor for Authors: <i>Saga+</i></b>	<b>97</b>
5.1 State of the Art . . . . .	97
5.1.1 Text Visualization for Literary Scholars . . . . .	97
5.1.2 Connected Editing and SaaS . . . . .	98
	xiii

## Contents

---

5.2	Overview . . . . .	98
5.3	Main Features . . . . .	100
5.3.1	Importing Existing Books . . . . .	100
5.3.2	Entities Manager . . . . .	100
5.3.3	Montage and Reorganization Interface . . . . .	101
5.3.4	Editing Features . . . . .	103
5.3.5	Versioning . . . . .	103
5.3.6	Reading Interface . . . . .	104
5.3.7	Feedback Loop and Optimization . . . . .	105
5.4	Technical specifications and software licensing . . . . .	105
<b>6</b>	<b>Results and Discussion</b>	<b>107</b>
6.1	Influence on the Author(s) . . . . .	107
6.2	Public reception of simulationhumaine.com . . . . .	112
6.2.1	Medias and Scholars . . . . .	112
6.2.2	Collected Usage Statistics . . . . .	112
6.2.3	Identifying Reading Patterns . . . . .	113
6.2.4	Qualitative Books and Chapters Classification . . . . .	115
6.3	Workshop with Authors . . . . .	116
6.3.1	Produced Novel . . . . .	119
6.3.2	Usability Feedbacks . . . . .	120

6.4 The Case of Dramaturgy . . . . .	122
<b>7 Limitations and Perspectives</b>	<b>125</b>
7.1 The Present . . . . .	125
7.2 A Thought Experiment . . . . .	126
7.3 Conclusion . . . . .	128
<b>A Appendices</b>	<b>129</b>
A.1 Appendix A . . . . .	130
A.1.1 Exploratory Modeling of Gamebooks . . . . .	130
A.2 Appendix B . . . . .	131
A.2.1 La succession des chapitres dans trois versions différentes (Daniel de Roulet, 18.11.2013) . . . . .	131
A.3 Appendix C . . . . .	132
A.3.1 Nouvelles narrations numériques Le projet 3N reconstruit la SH . . . . .	132
A.3.2 Les étapes de <i>La simulation humaine</i> (Daniel de Roulet, 28.04.2014) . . . . .	136
A.4 Appendix D . . . . .	138
A.4.1 Numeric Results of Named Entities Classifiers . . . . .	138
A.4.2 Semantic Classifier Words . . . . .	141
A.4.3 OpeNER classification results . . . . .	147
A.5 Appendix E . . . . .	149

## Contents

---

A.5.1 Attribution of chapters to originating books with LIWC features . . . . .	149
A.6 Appendix F . . . . .	150
A.6.1 Questionnaire aux écrivains ayant évalué <i>Saga+</i> . . . . .	150
A.7 Appendix G . . . . .	153
A.7.1 Frankenstein (version courte) . . . . .	153
<b>Bibliography</b>	<b>175</b>
<b>Curriculum Vitae</b>	<b>177</b>

# List of Figures

1.1 Recommended software for writing novels, according to novelists' advices on Quora.com . . . . .	4
2.1 The ten books constituting «La simulation humaine» (Copyright Canevas Editeur, Editions du Seuil, Largeur.com SA, Buchet-Chastel) . . . . .	16
2.2 Overview of the fields of study . . . . .	18
2.3 Typical iterative process core to our research, that directly echoes to <i>PAR</i> methods.	21
2.4 Technical overview of the components in the analysis/editing pipeline . . . . .	23
2.5 Visualization of <i>Le démentèlement du coeur</i> using <i>Voyant Tools</i> by Stéfán Sinclair and Geoffrey Rockwell (CC BY 4.0), December 2016. . . . .	27
2.6 The 16 possible readings of Alan Ayckbourn's <i>Intimate Exchanges</i> . Source: Wikipedia. . . . .	31
2.7 Original manuscript by Claude Simon, exposing his structural coding strategies when working on <i>La Route des Flandres</i> . Copyright © Réa Simon. . . . .	32
2.8 Qualitative comparison of constrained writing techniques . . . . .	33
3.1 Highlighted differences (in red) made between December 2013 (Left) and April 2014 (Right). Horizontal lines are shown instead of identical portions of text. Notice how almost all changes are related to simple local disambiguations. . . .	42

## List of Figures

---

3.2	Author's representations of the narrative structure components in <i>La simulation humaine</i> . . . . .	47
3.3	Number of words (top/blue) and sentences (bottom/green) in all chapters of the ten books of <i>La simulation humaine</i> . . . . .	48
3.4	Average sentences lengths in <i>La simulation humaine</i> . . . . .	48
3.5	Narrative focalizations transitions in <i>SH10</i> . . . . .	49
3.6	Joined representation comparing the narrative structures from <i>SH7</i> (left) and <i>SH2</i> (right): states diagram (up), structural foreseeability and its first order derivative (bottom). . . . .	50
3.7	Original sequencing of <i>La simulation humaine</i> (10 books in a row, each colored square represents a chapter from of the original books) . . . . .	51
3.8	New sequencing of <i>La simulation humaine</i> (The colored squares are the same as in fig. 3.7) . . . . .	51
3.9	New sequencing of <i>La simulation humaine</i> (9 new self-standing novels inside the reordered saga) . . . . .	52
3.10	Home screen of the digital saga <i>La simulation humaine</i> . . . . .	53
3.11	Reading interface of the digital saga <i>La simulation humaine</i> . . . . .	53
3.12	Textual navigation elements, located between two chapters exclusively. . . . .	54
3.13	Glossary function. . . . .	55
3.14	Example of raw usage statistics: User identifier, Chapter number, Story identifier, Date and Device Type. . . . .	56
4.1	Comparison between writing, structuration and <i>distant editing</i> . . . . .	59
4.2	Typical mean positions of uppercased words in their respective tokenized sentences vs. their number of occurrences (on a logarithmic scale). . . . .	65

4.3	Relative mean position of characters and places names for three classical French novels . . . . .	66
4.4	Comparison between Precision and Recall for each classifier, on each book . . .	72
4.5	Comparison between Precision and Recall for each meta-classification . . . . .	73
4.6	Graphical comparison between meta-classifications . . . . .	74
4.7	Repartition of uses of the word " <i>atomique</i> " in the ten books of <i>La simulation humaine</i> . . . . .	77
4.8	Frequency of mentions to "Cosette" in <i>Les Misérables</i> . Each tick represents one chapter, in reading order. Percentages are given relative to the total number of occurrences of the word. . . . .	77
4.9	Possible protagonists mentions comparisons on <i>Madame Bovary</i> , using histograms. . . . .	78
4.10	Possible protagonists mentions comparisons on <i>Madame Bovary</i> , using line charts. . . . .	78
4.11	Social network graphs comparison between two novels. . . . .	80
4.12	Social network computed from <i>La simulation humaine</i> (also available at <a href="http://simulationhumaine.com/socialnetwork.pdf">http://simulationhumaine.com/socialnetwork.pdf</a> ) . . . . .	81
4.13	Detail comparing the computed network with the author's vision (in green, matching parts vs. in red, missing edges and nodes). . . . .	82
4.14	Bipartite graph of the characters-places network in <i>Madame Bovary</i> . . . . .	83
4.15	Example of features computed by LIWC . . . . .	84
4.16	Positive and Negative Emotions Graph on <i>SH.1</i> . . . . .	85
4.17	Positive and Negative Emotions Graph on <i>SH.2</i> . . . . .	85

## List of Figures

---

4.18	Entries-Exits graph for <i>SH.2</i> . Top: characters lasting, entering and exiting between each chapters. Bottom: first and last seen curves. . . . .	87
4.19	Comparison of Entries-Exits graph of <i>SH.5</i> (left) and <i>SH.6</i> (right). Top: characters lasting, entering and exiting between each chapters. Bottom: first and last seen curves. . . . .	88
4.20	Original xkcd storyline by Randall Munroe (2009, Creative Commons 2.5). . . . .	89
4.21	<i>Storyline</i> visualization automatically generated with <i>Around the World in Eighty Days</i> . . . . .	90
4.22	<i>Storyline</i> visualization automatically generated with <i>Madame Bovary</i> . . . . .	90
4.23	<i>Storyline</i> visualization automatically generated out of characters and places extracted from <i>La simulation humaine</i> . . . . .	91
4.24	Manually reordered <i>Storyline</i> visualization of <i>La simulation humaine</i> . . . . .	91
4.25	Timeline visualization of <i>La simulation humaine</i> (Top: detail showing the first two volumes / Bottom: global view on 298 chapters) . . . . .	92
4.26	Timeline visualization of <i>Around the World in Eighty Days</i> . . . . .	93
4.27	Timeline visualization of <i>Madame Bovary</i> . . . . .	94
5.1	Text editing interface . . . . .	99
5.2	<i>Saga+</i> entities manager (here displaying the character <i>Fumika</i> in <i>La simulation humaine</i> . . . . .	101
5.3	Structure montage interface (left) and layout of a chapter tile (right) . . . . .	102
5.4	<i>Saga+</i> montage and reorganization interface (featuring <i>La simulation humaine</i> )	102
5.5	<i>Saga+</i> text editing interface . . . . .	103
5.6	<i>Saga+</i> versioning manager window . . . . .	104



6.1 Two examples of exponential drop-offs with different rates / slopes . . . . .	114
6.2 An example of a curve with a plateau in <i>La réalité, mais digitale</i> . . . . .	114
6.3 Two skimming gaps in the number of readings per chapter in <i>L'ingénieur et la fillette</i> . . . . .	115
6.4 Continuously decreasing readers in <i>Fissions</i> . . . . .	115
6.5 J48 tree predictor for critical and immersion reading modes . . . . .	116
6.6 Advertising material for the workshop . . . . .	117
6.7 Presentation of <i>Saga+</i> to participants of the workshop. Photo: Frédéric Kaplan.	118
6.8 Collective writing time at the workshop. Photo: Frédéric Kaplan. . . . .	118
6.9 Structure visualization of <i>Frankenstein (version courte)</i> in <i>Saga+</i> . . . . .	120
A.1 Possible readings of Raymond Queneau's <i>Un conte à votre façon</i> . . . . .	130
A.2 Possible readings of <i>Pierrick remonte le temps</i> . . . . .	130
A.3 Possible readings of <i>Yoël-Krahn le magnifique</i> . . . . .	130



# List of Tables

3.1	Nomenclature of the books constituting the corpus <i>La simulation humaine</i> . . .	43
4.1	Rules for grammatical structure classifier . . . . .	67
4.2	Protagonists presence in the 18 chapters of <i>SH.10</i> . . . . .	86
A.1	Overall precision, recall and $F_1$ scores. Lowest and highest scores are underlined in bold. . . . .	139
A.2	Overall precision, recall and $F_1$ scores. Lowest and highest scores are underlined in bold. . . . .	140
A.3	Benchmarked <i>OpeNER</i> results on “ <i>Les Malheurs de Sophie</i> ” . . . . .	147
A.4	Benchmarked <i>OpeNER</i> results on “ <i>Germinie Lacerteux</i> ” . . . . .	148



# 1 Introduction

Digital media are reshaping many aspects of our everyday life, from leisure to collaborative and individual professions, as well as the economics models behind them. The sector of entertainment, starting with music, suffered dramatic changes due to the democratization of broadband Internet access [Meisel and Sullivan(2002)]. Smartphone and tablet PCs amplified this phenomenon by globalizing Internet connections and rising countless new business models around software, services and social networks [Snickars and Vonderau(2012)]. Although product-wise similar to their predecessors, the various models of digital entertainment differ a lot in terms of market dynamics, patterns of consumption, power relations between market actors and, more and more, the actual way contents are produced.

These kinds of dynamics have already lead to various studies, for instance from the very pragmatic standpoint of security concerns in digital entertainment [Eskicioglu et al.(2003)] to more theoretic views on industry convergence [Frøslev Christensen and Maskell(2003)], as well as in a variety of fields ranging from the well-established motion pictures industry [Zhu(2001)] to emerging sectors like video games [Arakji and Lang(2007)] [Kline(2003)]. Digital media are also reshaped as quickly as (especially young) people adapt to their use and grab the opportunities they offer [Buckingham and Willett(2013)], which opens a complex game of mutual adaptation and feedbacks. For instance, during the last decade, most digital media have become connected, often including a social component. The so-called social medias quickly received an unprecedented importance in our society. Within just a couple of years, this revolution triggered endless debates about privacy and personal data monetization, played an important role in cascading political changes [Khondker(2011)], and as a consequence became legitimate both as an object and as a tool for scientific studies and, of course, was at the origin and a catalyst for a tremendous number of business opportunities [Jansen et al.(2009)].

Curiously, the book industry is one of the only sectors where those effects have had only a limited influence until now [Robinson(2010)]. Even if the domain at first welcomed digital publishing and major actors placed high expectations on the possibilities offered by digitization at the end of the 20th century, little has been achieved and the positive feeling changed to reluctance and skepticism [Thompson(2013)]. In the domain of digital storytelling studies, the typical literature is mainly concerned by highly-technological media like Internet, games, TV and immersive environments, whereas pure reading is generally left out of the equation [Meadows(2003), Miller(2004), Lambert(2012)]. So far, very few writers and even fewer publishers have attempted to explore the mind-blowing possibilities offered by digital approaches of their works. In the end, many digital counterparts to paper best-sellers ended up as disappointing experiences, and the very digital reading experience has yet to reach another stage to get favored to paper among the (conventionally technology enthusiastic) younger public [Merga and others(2014), Cesário et al.(2016)]. The market reflects well this situation, as even the proportion of regular books available in digital formats is marginally low and their overall sales figures hardly reach the few percent mark [Wischenbart et al.(2012)].

### 1.1 Tools for Digital Writing and Reading

The paradoxical situation of the fictional novel, being an emblematic cultural good and yet failing to transition gracefully to the digital era, can be seen through the prism of the market but also from the perspective of available writing tools. As a matter of fact, we know that in the core of any manufacturing process, tools, making techniques and final products are interdependent and mutually influencing components. This is especially the case with digital tools, which due to dematerialization play a role of unique interaction point and thus have a strong influence on creative processes, to the point of noticeably and distinctively influencing artworks [Madigan(1984)].

However, as far as the novelist's work is concerned, the tools that are currently available to them are not very different in nature from those of pre-computers era. Of course, word processors technologies evolved quickly and improved the general writing experience tremendously, by taking care of tedious tasks like spellchecking, easing access to all kinds of writing dictionaries or various improvement factors at wording level. Acknowledging the growing principles of user-centered design [Abrams et al.(2004)], modern software editors recently started to focus on the mechanics of writing processes instead of the text per se. Several trends emerged, each time with interesting tools leading to new paradigms and ways of working:

- **Minimalist approaches.** Stepping back from the excessively complex and over-featured

text editors from the 2000s, several editors proposed minimalist and distraction-free approaches as a recipe for plentiful and richer outputs [Nield(2015), Babauta(2015)], yet their target audience is unclear, and the usability in the context of writing full books is subject to debate<sup>1</sup>.

- **Scriptwriting assistants.** On the contrary, Literature and Latte's Scrivener<sup>2</sup> made a noticed technological advance by providing a set of functionalities tailored for writing large documents. Among them are an environment dedicated to nonlinear editing by means of various organizational tools such as virtual index cards, hierarchical outlier view or fragmentation of documents into constituent components, as well as scriptwriting assistance and tools supporting the actual writing activity, such as statistics, user set targets, automatic text formatting and snapshots. Those features were viewed as quite unique in that field (even more so when brought together as one single tool) and made it a very appreciated tool among novelists and humanities scholars [Meister(2014)].
- **Collaborative editing tools.** By nature, connected tools allow to explore different ways of working together, in particular at the same time and on the same work, often from different physical locations. The concept of several novelists coworking on a common text has notably been suggested and explored by Penflip<sup>3</sup>. One can also cite several examples of largely sourced writing experiences with no central guidance, like Foldingstory<sup>4</sup> or Ficly<sup>5</sup>, yet many of them closed down shortly, after encountering strategic issues or not finding their public (see WikiStory<sup>6</sup> or Fabulate<sup>7</sup>).
- **Distribution engines.** Some other initiatives, such as Describli<sup>8</sup>, promised to aggregate different aspects related to the publication chain, like the multiplicity of distribution sources and marketing automation through social media.

Even so, a recent poll on the question-and-answer site Quora<sup>9</sup> aiming to get actual novelists' insights about their habits in writing showed that today, although some tools are definite default choices, none gets a very clear preference amongst book authors (A synthetic view of the 85 produced answers is shown in fig. 1.1). In this fragmented landscape, nearly 50 tools

---

<sup>1</sup>Stack Exchange | Are distraction-free text editors effective? - <http://ux.stackexchange.com/questions/62766/are-distraction-free-text-editors-effective>

<sup>2</sup>Literature and Latte | Scrivener 2 - <http://www.literatureandlatte.com/scrivener.php>

<sup>3</sup>Penflip - <http://www.penflip.com>

<sup>4</sup>Foldingstory - <http://foldingstory.com>

<sup>5</sup>Ficly - <http://ficly.com>

<sup>6</sup>WikiStory - <http://www.wikistory.com>

<sup>7</sup>Fabulate - <http://www.fabulate.co.uk>

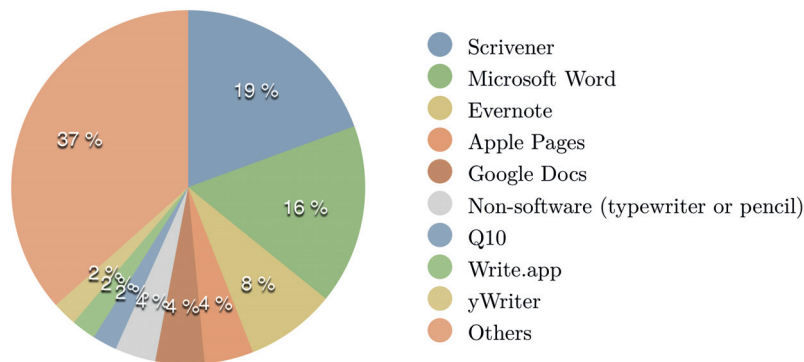
<sup>8</sup>Describli - <http://describli.strikingly.com/>

<sup>9</sup>Quora | What are the best software tools for writing books? - <http://www.quora.com/what-are-the-best-software-tools-for-writing-books>

## Chapter 1. Introduction

---

were nominally mentioned. Out of them, 33 are mentioned only once, and only 6 more than three times. Interestingly, the cumulated share of general purpose text processors (Microsoft Word, Google Docs and Apple Pages) exceeded by far the mentions of software dedicated to fiction writing, and more people agreed on the rejection of digital technologies altogether than on any specific novel writing tool.



**Figure 1.1** – Recommended software for writing novels, according to novelists' advice on Quora.com

This striking situation can probably be explained by global market issues, and by the little added value those tools provide compared to other technological imperatives, such as the wide cross-platform or long term compatibility that are offered by general purpose word processors. The British award-winning science fiction writer Charlie Stross summarized his views on this situation (and by that, possibly a feeling shared by many authors) in an enlightening blog post [Stross(2013)], where he stated notably:

*Microsoft Word is a tyrant of the imagination, a petty, unimaginative, inconsistent dictator that is ill-suited to any creative writer's use.*

and:

*I do not write novels using Microsoft Word. [...] But somehow, the major publishers have been browbeaten into believing that Word is the sine qua non of document production systems. They have warped and corrupted their production workflow into using Microsoft Word .doc files as their raw substrate, even though this is a file format ill-suited for editorial or typesetting chores. [...] by its very prominence, we become blind to the possibility that our tools for document creation could be improved. It has held us back for nearly 25 years already; I hope we will find something better to take its place soon.*



## 1.1. Tools for Digital Writing and Reading

---

Nonetheless, the potential exists for tremendous shifts in the ways digital literature is conceived and consumed. As some already theorized [Chaplin(2010)], in the near future, the work of a novelist may tend towards a form of collaboration significantly involving his readers. Digital publishing enables authors to have a direct connection with their audience, completely removing the publisher's role from the equation. Conversely, readers are given the power to influence the contents, by sharing their thoughts among themselves or with the authors, or taking an active part in the artwork, as it is the case with fan fiction [Black(2005)]. Very recent works indeed showed the feasibility of introducing crowdsourcing at the heart of writing tools, as a mean of supporting an author otherwise working alone [Bernstein et al.(2015)]. For some time now, academia also showed interest for the possible uses that could be made out of digital books, for instance with the use collaborative annotation strategies [Marshall(1997)] or wiki platforms for teaching [Kuteeva(2011)].

One can not say if these new writing ecosystems will still match our current definition of (e-)books or if they will give rise to a new generation of literary items that will have few to share with their ancestors. Nevertheless and as argued by Marie-Laure Ryan[Ryan(2004)], it seems obvious that the art of storytelling is far from disappearing, even if the way stories are built and perceived is to be drastically transformed through digital adoption. Whatever the possible outcomes, the purpose of this research is to cast a light on some of the possible changes that can be foreseen in the way that the books of tomorrow will be written.



# Document 1

Daniel de Roulet, 11.01.2014

## ***La simulation humaine, comment je l'ai écrite, construite, déconstruite, et à peine reconstruite***

J'ai commencé à écrire *La simulation humaine* (SH) en 1990. Lors d'un séjour de six mois à New York en 1992, j'ai terminé deux romans. L'un se passait à Harlem, l'autre pendant le marathon de New York. Les personnages de ces deux romans étaient vaguement parents. Le scénario se déroulait au début des années 90, donc à l'époque de son écriture.

Pour le premier de ces deux romans, que je rapportais de New York, j'ai trouvé un éditeur. Et c'est devenu *Virtuellement vôtre*, publié par Canevas en 1993. Le héros en est Vladimir vom Pokk, dit Work.

Le deuxième roman, *La ligne bleue*, a été publié l'année suivante au Seuil. Il met en scène l'architecte Max vom Pokk, né en Franche-Comté et l'ingénieur Shizuko Tsutsui, née à Nagasaki. Le roman raconte en parallèle un attentat anti-nucléaire.

L'année suivante chez le même éditeur paraît *Bleu Siècle* qui, par ce bleu dans le titre, annonce la continuité avec le précédent. Les personnages principaux en sont le grand-père de Max et la petite-fille de Shizuko. Le temps du récit est contemporain de son écriture, et 1996 est l'année où ce grand-père devient centenaire. Le chapitre 13 est intitulé *La simulation humaine*, ce titre étant

annoncé comme une « épopée » dont mes personnages seraient les héros. Bien que les deux familles, vom Pokk et Tsutsui, soient présentes désormais dans chaque roman, il n'est pas nécessaire que le lecteur ait connaissance des romans précédents pour comprendre le nouveau. Ce principe ne changera pas au fur et à mesure de la publication de la saga. En 1999 chez le même éditeur, paraît encore un titre en bleu, *Gris-bleu*, dont il est dit en 4ème de couverture qu'il fait partie de *La simulation humaine*. C'est le journal intime du fils de Shizuko. Mais cette fois l'action se situe en 1991, donc avec un certain décalage par rapport à la période d'écriture.

En 2001, aurait dû paraître *Fusion* qui se passe dans un gratte-ciel avec des gens qui se jettent par les fenêtres, mais je n'ai pas trouvé d'éditeur, comme je l'explique ailleurs (*Ecrire la mondialité*), en raison notamment de la trop grande proximité de la fiction avec la réalité du 11 septembre 2001. En 2002, j'ai publié, simultanément en anglais, en allemand et en français, un feuilleton à épisodes sur internet, *Davos terminus*. A Davos, cet hiver-là, Max vom Pokk fréquente le World Economic Forum. L'action se déroule le jour même de la parution de chaque épisode, comme s'il s'agissait d'un compte-rendu de faits réels. A ce

## Document 1

---

point de l'écriture de la SH, je considérais que les cinq tomes parus formaient un tout que je ferais publier comme tel une fois que j'aurais re-travaillé *Fusion*.

En 2005 chez Buchet-Chastel est paru *L'homme qui tombe*. C'est l'histoire écrite à la première personne des quelques secondes au cours desquelles Georges vom Pokk, le cousin de Max, tombe d'un toit. Il est ingénieur nucléaire et revient du Japon. Le temps du récit est contemporain de son écriture. L'année suivante, j'ai présenté à mon éditeur un très gros roman qui reprenait les éléments de *Fusion*, toujours inédit, précédé d'une vaste fresque retraçant les premières années du nucléaire militaire puis civil. Les personnages en étaient la jeune Shizuko et le jeune Max. Le récit commençait en 1938, avant leur naissance. Mais ce gros roman, dont l'action s'étalait de 1938 à 1988, a été jugé trop volumineux, il se terminera donc en 1968, juste avant la rencontre de Max et de Shizuko et s'appellera *Kamikaze Mozart*, publié en 2007. A cette date, la SH comporte donc sept tomes, mais l'ordre de leur publication ne correspond pas à la chronologie de l'action. Le septième publié est le premier dans la chronologie de la saga.

En 2009, Buchet Chastel publie *Le silence des abeilles* qui met en scène la fille et le fils de Shizuko comme personnages secondaires. L'action est contemporaine et parle de l'entreprise *Bleu Siècle* qui était propriété du vieux vom Pokk, mort en 1996.

La publication de *Fusions* (avec s cette fois) qui raconte le destin des deux familles vom Pokk et Tsutsui entre 1968 et 1988 tarde jusqu'en 2012. Il avait été prévu de faire paraître ce roman en même temps qu'une version numérique de l'ensemble de la saga qui aurait commencé en 1938 et ce serait terminé en 2009. Ceci n'a pas été possible et a brouillé la réception de *Fusions*, dont la publication a finalement attendu plus de treize ans et dont

les réécritures successives ont rendu l'intrigue de plus en plus complexe : de nombreux retours en arrière, une foule de personnages, jusqu'à sept points de vue différents sur les événements. La tâche du lecteur n'en était pas simplifiée.

En 2014 paraîtra *Le démantèlement du cœur* qui met en scène les deux ans qui suivent la catastrophe de Fukushima (11 mars 2011) vécue par le fils de Max et de Shizuko. Cette fois les personnages mourront, la saga est close. Elle se déroule entre 1938 et 2013, 75 ans d'histoire nucléaire : naissance, apogée, et chute.

L'écriture de la saga a donc duré 24 ans, l'ordre de publication n'est pas celui de la chronologie de l'action qui se présente ainsi :

- SH1 *Kamikaze Mozart*
- SH2 *Fusions*
- SH3 *Gris-bleu*
- SH4 *Virtuellement vôtre !*
- SH5 *La ligne bleue*
- SH6 *Bleu siècle*
- SH7 *Davos Terminus*
- SH8 *L'homme qui tombe*
- SH9 *Le silence des abeilles*
- SH10 *Le démantèlement du cœur*

Dès 1999, à la fin de la série des tomes qui ont du bleu dans le titre (*La ligne bleue*, *Bleu Siècle* et *Gris-bleu*), j'ai imaginé la recomposition d'une saga. J'ai constitué une nomenclature sous forme de fiches signalétiques de chaque personnage et de l'arbre généalogique des deux familles. J'ai parfois repris un chapitre écrit pour un tome et l'ai glissé dans un autre, mais je considérais que cela faisait partie de ma cuisine littéraire personnelle. A partir de juin 2007 et de *Kamikaze Mozart*, il était clair pour moi que j'aurais besoin d'utiliser les outils informatiques pour présenter au lecteur non pas une seule mais plusieurs lectures de la saga.

A l'époque, j'envisageais simplement de remettre

dans l'ordre chronologique les sept tomes à disposition. A la publication en 2009 de SH9, j'ai espéré de nouveau pouvoir publier le tout comme une seule saga du nucléaire. Elle aurait commencé avec Hiroshima et serait terminée par la chute d'un toit d'un ingénieur nucléaire revenu du Japon. Le projet a été repoussé encore parce qu'il ne semblait pas indispensable à la compréhension du cycle, d'autant plus que SH2 n'était toujours pas publié. Au moment de sa publication, le projet s'était enrichi et j'envisageais donc de faire publier sous forme numérique les neuf tomes déjà publiés en version papier en utilisant l'informatique dans trois directions :

1. la publication homothétique des neuf tomes dans leur ordre chronologique
2. la publication de la documentation de mes recherches sur la saga de l'atome
3. la publication des parcours biographiques de quelques personnages.

A ce moment-là non plus, je n'ai pas réussi à convaincre de la nécessité éditoriale de ce projet. Comme je ne tenais pas à y renoncer, j'ai obtenu de mon éditeur que la prochaine fois serait la bonne. La SH compterait alors dix tomes et je me résoudrais à faire mourir mes personnages principaux pour marquer la fin du cycle. La sortie prévue en 2014 de SH10 se ferait donc en même temps que la sortie d'un document numérique. Pour ce qui est du support physique de cette saga, j'avais pensé dès 2005 à un CD inclus dans le livre papier. Aujourd'hui tout ça se passera dans le Nuage virtuel et sera également lisible aussi bien sur tablettes que sur téléphone intelligents.

Mes contacts avec le DHLAB de l'EPFL ont commencé en 2012 et très vite j'ai compris que cette collaboration serait pour moi une occasion unique de renouer avec un métier que j'avais délaissé et dont je peinais à suivre le développement fréné-

tique. Après une ou deux séances, il n'est pas resté grand chose de mon projet initial et de ses trois objectifs.

Pour la publication homothétique, j'ai compris qu'il s'agissait d'un autre circuit : celui des libraires et des fournisseurs de téléchargements. Ceci pourrait se faire, mais ne permettrait pas de nouveaux parcours de lecture.

Pour mes idées de documentation multimédia et publication de mes documents de recherche, je me suis trouvé face à l'opinion bien arrêtée que le métier du romancier, c'est la production de textes et que c'est cela qui devait être le point de départ et d'aboutissement. J'ai été surpris de voir à quel point l'immersion dans le texte et rien que le texte était un credo solide du DHLAB. Je revois Frédéric Kaplan me disant : « Retour au corpus » et Cyril Bornet souriant pour marquer son scepticisme. Il disait qu'un lecteur à qui on laisserait le choix de s'évader hors du texte par des hyperliens conduisant à des documents extérieurs ne retournerait plus vers mon texte. Je me suis plié à leur avis, d'abord de mauvaise grâce, mais je considère ce point, pour moi, comme l'acquis le plus novateur de ma collaboration.

Pour ce qui est des cinq à quinze biographies constituant des parcours de lecture nouveaux à travers le corpus, j'ai moi-même fait l'exercice pour m'apercevoir ensuite, à la lecture de ces biographies que leur intérêt était faible voire inexistant. Dans mes romans, c'est la confrontation entre les points de vue portés par différents personnages qui soutient l'intérêt du lecteur. Remettre à plat les biographies des personnages comme autant de romans d'éducation, ou suite d'épisodes, détruit toute tension dramatique.

Après peu de temps avec le DHLAB, j'ai donc dû abandonner les trois aspects du projet initial. Il me restait cependant la conviction qu'il y avait là

## Document 1

---

quelque chose d'original à chercher et à construire, quelque chose qui profite du fait que je travaille un corpus homogène, malgré l'évolution du style de l'auteur et la succession des événements historiques, avec une thématique et des personnages récurrents.

Au début, j'étais assez étonné de la manière dont le DHLAB empoignait mon corpus, presque comme une objet mathématique et sans beaucoup de respect pour les affres de l'écriture qu'avait traversées le romancier que j'étais avant d'en accoucher. Pendant vingt ans, j'avais travaillé seul avec ma plume et mes petits papiers, plein de réserve face à la numérisation, et voilà que je me trouvais face à des têtes faites autrement, mais peut-être finalement moins carrées que la mienne. Heureusement que j'avais décidé que SH10 était mon dernier tome, j'avais donc la tête libre de toutes mes fictions pour me pencher avec une certaine distance sur ce que j'avais produit et sur une autre manière de le recomposer.

Et pourquoi n'ai-je pas abandonné quand les objectifs que je m'étais fixés ont été torpillés à juste titre ? C'est que j'ai toujours considéré que mes textes pouvaient être améliorés, qu'en les relisant vingt ans plus tard, je découvrais toutes sortes d'endroits pour lesquels je n'avais plus l'indulgence, mais l'agacement du lecteur égaré qui se disait : mais où l'auteur veut-il m'emmener ? A quoi fait-il allusion dont je n'ai jamais entendu parler ? En devenant le lecteur distancé de mes textes, j'en ai vu les défauts. Je comprenais seulement maintenant les critiques intelligentes de mes lecteurs d'alors, remarques aimables ou dures que j'avais balayées alors par le mépris que je croyais dû aux lecteurs inattentifs aux profondeurs de mes phrases. Je ne crois pas au progrès moral ou au progrès de l'émotion que provoquerait la réception des œuvres, mais je crois aux progrès de l'artisan dans son métier, à la maîtrise du savoir-faire. Et chez moi, si d'un roman à l'autre j'ai continué,

l'écriture du livre suivant me paraissait un progrès objectif. Je ne voulais pas être un romancier touche à tout, j'avais investi dans le dur métier d'écrire une saga sur le nucléaire et voulais aller jusqu'au bout. Je me disais : ton apprentissage autodidacte durera plus longtemps que prévu, bientôt tu pourras retoucher tes pièces manufacturées une dernière fois. Et c'est pourquoi j'ai décidé de faire du tome dix le dernier, d'appliquer à mon écriture les nouveaux principes de construction que je m'étais enseigné par déduction de mes erreurs. Et si c'est raté encore ou pas abouti, j'aurai mis au moins toutes les forces que me permet un cerveau qui commence à manquer de souplesse face à une telle complexité.

J'avais déjà étudié le parcours de quelques auteurs confrontés à l'écriture d'un long cycle, comme je l'explique ailleurs (*Écrire la mondialité* [de Roulet(2013)]). Zola, Proust, Martin du Gard et Sartre ont été mes modèles. Ceux-là avaient eu de la peine à composer leur cycle. Martin du Gard avait loué une maison avec de très nombreuses tables sur lesquelles il empilait les feuillets se rapportant à chaque tome prévu. Sartre avait déclaré forfait, Zola et Proust avaient dû tricher avec la chronologie. Mais tout cela c'était pendant les longues années d'écriture. Ceux-là, une fois le travail fini, ne s'étaient pas mis à essayer de réordonner leurs chapitres comme ces quelques auteurs qui avaient construit ce qu'on appellerait plus tard des hypertextes. J'ai croisé ainsi les exercices de Queneau dans *Un conte à votre façon* (cf. analyse de Cyril appendix A.1.1) et *Marelle* de Julio Cortazar. Ce que le DHLAB me proposait de faire ressemble surtout à ce dernier cas.

Je me suis donc attelé à la décomposition. Il a fallu choisir d'abord l'unité de décomposition du corpus. Tout naturellement, après la subdivision en dix tomes se présentait la division en chapitres. Ça paraît évident même si leur nombre et surtout leur taille varient fortement d'un tome à l'autre. SH9,

par exemple, est constitué de 49 chapitre parfois guère plus longs que deux pages. Dans SH2, les chapitres sont longs et nombreux, dans SH4 il n'y en a que six. A la relecture il m'a fallu comprendre ce qu'était pour moi un chapitre, trouver une définition et voir si elle s'appliquait ensuite à tous les 297 chapitres des dix tomes.

Dans la majorité des cas, le chapitre peut se caractériser par le point de vue d'un seul personnage à la troisième personne du singulier. Ce personnage qui est le narrateur implicite (focalisation interne, dirait Genette) raconte une histoire qui se déroule chronologiquement du début à la fin du chapitre. Le temps employé est le présent sauf pour raconter des événements déjà échus qui sont relatés au passé composé ou à l'imparfait (je n'utilise pas le passé simple). L'époque historique à laquelle se passe ce récit est désignée parfois de manière très précise (jour et heure), parfois de manière plus vague, juste une saison et l'année (printemps 1943). Les lieux dans lesquels ces chapitres se déroulent sont toujours des lieux réels, les villes sont nommées, parfois de manière encore plus précise en indiquant le nom de la rue. Même quand le lieu est inventé (une chambre à coucher, un restaurant) ses coordonnées topographiques sont définies. Les trois quarts des chapitres correspondent à cette définition : un seul point de vue par chapitre, un lieu, et une date.

Que faire alors des chapitres écrits à la première personne du singulier (tout SH3 et tout SH8, 4 chapitres de SH2) ou à la première du pluriel (le « nous » d'un double personnage, la mère et sa petite fille dans SH6) ou à la deuxième du pluriel (un vous de politesse pour s'adresser à un bâtiment dans SH2) ? J'ai longtemps hésité à récrire ces chapitres, craignant de me retrouver avec une voix narrative mal assurée. J'avais peur que ça sonne faux et me rendais compte que ce serait un travail gigantesque.

Dans certains tomes, en particulier SH5, surgissait un autre problème : un fil temporel double à l'intérieur du même chapitre. Ceci m'a amené à diviser certains chapitres en deux pour séparer ces deux temps du récit. J'aboutissais alors à une unité de décomposition plus courte que le chapitre. J'ai étudié cette variante, tout décomposer en petites unités. Après de nombreux essais, j'ai finalement trouvé une séquence d'environ cinq mille signes qui correspondait à un rythme personnel de mon écriture dont je n'avais pas conscience. J'ai donc repris tout le corpus en le divisant de cette nouvelle manière. J'aboutissais finalement à une décomposition en près de 680 séquences, mais j'entrevois déjà que la reconstruction en serait rendue encore plus difficile. Je suis retourné à mes 297 chapitres.

Au fur et à mesure que je subdivisais mes textes, je me suis éloigné de la position de l'auteur pour n'en devenir que le lecteur critique. Au point que je me suis forcé à écrire d'autres textes, de petits portraits sans lien avec le corpus, pour me persuader que j'étais encore capable de tenir ma plume. Peu à peu la SH m'est apparue un corpus comme un autre, plus ou moins réussi, mais dont l'architecture d'ensemble pouvait être disséquée et remodelée.

C'est alors que j'ai entrepris de récrire peu à peu toutes les parties rédigées autrement qu'avec le point de vue d'un personnage exprimé à la troisième personne du singulier. SH3, qui se présentait sous la forme d'un journal intime a donc pris l'allure d'une biographie. SH8, L'homme qui tombe, qui était un long monologue intérieur est devenu l'histoire d'un ingénieur qui perd ses repaires. J'avais conscience non seulement de déconstruire, mais de démolir ce qui constituait l'essence même de mes romans. Mais c'était avec en tête l'idée que, si je réussissais à mener l'entreprise jusqu'au bout, je pourrais ensuite disposer de briques unifiées pour la reconstruction

## Document 1

---

d'un mur solide.

Je crois pouvoir dire qu'entre les versions papiers et les briques que je préparais, aucun chapitre n'est sorti indemne du processus. Car en plus des changements de personnes dans la conjugaison, il fallait s'assurer que dans chaque chapitre la narration puisse être comprise du premier coup. Il a fallu rendre explicite dès le premier paragraphe du chapitre le nom des acteurs et leur situation dans le temps. Des formulations vagues en relation avec les chapitres précédents ont dû devenir très précises. On ne pouvait plus écrire en début de chapitre « *A quelque temps de là Max s'est installé non loin de là...* » Il fallait dire : « *En juin 1988, Max s'est installé à Manhattan...* »

A la fin de la déconstruction, à l'automne 2013, je me souviens d'avoir eu un moment de panique, comme un maçon devant un tas de briques récupérées et qui ne sait plus s'il va réussir ou pas à remonter un mur. Mes amis du DHLAB, eux, n'avaient pas ce scrupule. Pour eux une brique est un brique et le nombre des combinaisons possibles est infini. Possible, oui, mais signifiantes, non. Il me restait assez d'orgueil d'auteur pour ne pas proposer à mes lecteurs des reconstructions que je ne pouvais pas assumer. Je ne tenais pas à constituer une arborescence où le lecteur doit choisir. Je continue de penser que même si l'auteur ne contrôle pas tout de l'effet qu'il produit, il doit rester maître du jeu, cautionner et imposer ses choix. Laisser au hasard le soin de reconstruire une narration est un défi pour informaticien, mais ne contribue pas à entretenir le lien de confiance entre l'auteur et son lecteur. Je ne voulais donc, quant à moi, proposer pour le moment que des parcours que j'ai moi-même suivis et dont je puisse garantir l'intérêt de lecture.

Je raconte ailleurs (appendix A.3.1) les quelques principes qui peuvent guider une recombinaison en utilisant la surprise et le suspense pour recréer

une tension narrative. Pour le moment, dans cette première phase du projet qui sera disponible publiquement en avril 2014, j'ai procédé pour mon lecteur comme si j'étais un libraire à qui il vient demander conseil. Entrant dans ma librairie (mon corpus), il commence par feuilleter sur les tables les livres exposés pour en avoir un avant-goût. Ensuite, il demande au libraire de lui conseiller un texte court, une nouvelle, par exemple, cinq à huit chapitres à lire en moins d'une heure. Ou bien il demande un vrai roman, de la lecture pour deux à cinq heures au plus, mais pas les cinquante heures qu'il lui faudrait probablement pour lire le corpus entier. Je me suis donc fixé de proposer à mon lecteur six avant-goûts ou six nouvelles ou trois romans. La question qui se pose est donc de savoir s'il existe au moins un corpus recomposé qui offre une lecture plus intéressante que la simple réorganisation chronologique de SH1 à SH10. Si ce corpus existe, j'aurai atteint mon but, j'aurai appris quelque chose sur l'organisation de ma propre écriture. La contrainte supplémentaire introduite pour cette reconstruction est qu'il faut que les six nouvelles ou les trois romans puissent être soustraites du corpus et que le reste demeure signifiant, que le sens ne s'épuise pas avec le rétrécissement du corpus.

Pour obtenir une nouvelle, j'ai choisi des chapitres concernant le même personnage dont la juxtaposition formait un récit d'un début à une fin. Pour obtenir trois romans à cheval sur plusieurs tomes, j'ai commencé par opérer des groupements de personnages, en général par deux ou trois. J'ai pris dans l'ordre chronologique tous les chapitres qui se rapportaient à un personnage dans un tome et les ai combinés avec les chapitres qui se rapportaient à un autre personnage dans un autre tome jusqu'à trouver quelques combinaisons qui fassent sens.

Pour savoir lesquels retenir, j'ai considéré les « trous » que ces romans faisaient dans la narra-



tion du corpus pour que ce qui reste continue d'être un récit intéressant. Dans ce travail, j'ai agi de manière intuitive en essayant d'avoir en tête autant de paramètres que mon cerveau peut en retenir. J'espère avoir été aussi convaincant que ce qu'un algorithme automatisé aurait pu proposer.

Je n'exclus pas que, une fois que le résultat aura été modélisé par le DHLAB, que d'autres reconstructions, encore plus intéressantes pour le lecteur, puissent être proposées. Je laisse volontiers ce travail à la perspicacité du DHLAB.



## 2 Context and Methodology

### 2.1 Context of this Research

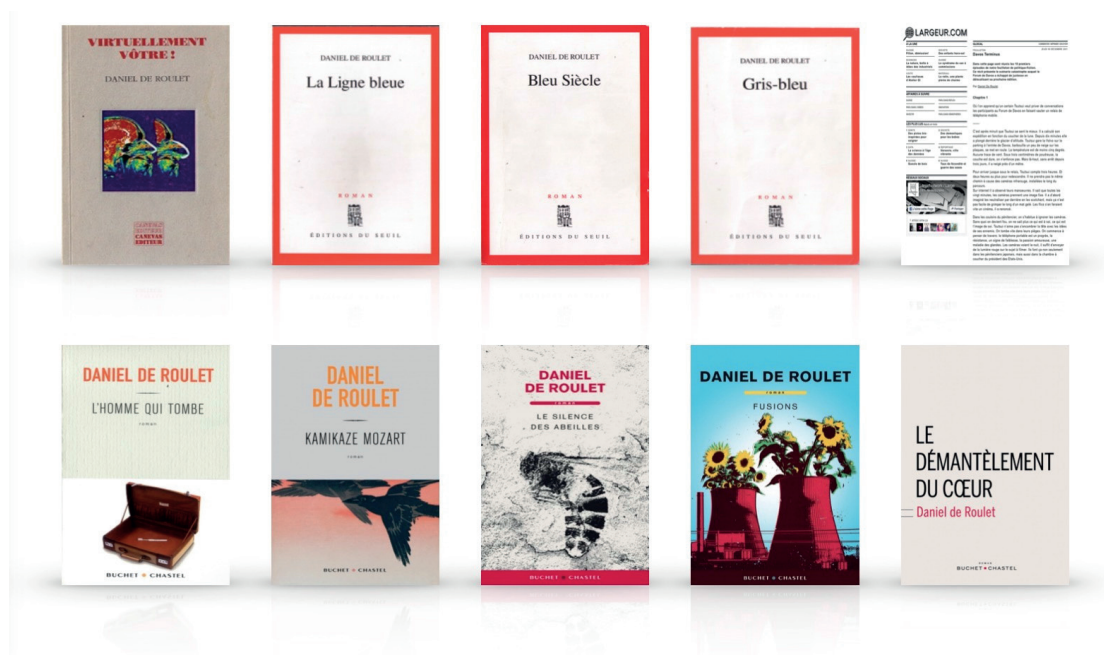
Born in Geneva in 1944, Daniel de Roulet<sup>1</sup> is a French speaking Swiss novelist. First an architect, then software engineer, he worked on some of the largest early computer systems infrastructures deployed in Switzerland. A full-time writer since 1997, his works, ranging from collections of personal letters to purely fictional compositions, are deeply meaningful and denounce political and ethical issues of our times. He notably published several critique essays, among which *Tu n'as rien vu à Fukushima* and *Ecrire la mondialité*. Globality ("*la mondialité*") is for him, in opposition to globalization, the positive side of the new contemporary order that allows thinking of what will come after modernity, in writing as well as in common goods politics. In 2015, he was awarded "Culture and Society" prize of the City of Geneva [Grosjean(2015)].

Between 1993 and 2012, he published, among other works, a succession of nine books (section 2.1) [de Roulet(1993), de Roulet(1995), de Roulet(1996), de Roulet(1999), de Roulet(2001), de Roulet(2005), de Roulet(2007), de Roulet(2009), de Roulet(2012)] forming a family saga he later named *La simulation humaine*. This fictional work relates the adventures of two families and their complex, love and fear relationships with the development of nuclear power technologies, so-called by the author «*from Hiroshima to Fukushima*». The saga features 17 protagonists related by blood and hundreds of side characters in a fictional world constituted of half a million words. It was ended with the tenth and last book, *Le démantèlement du coeur* [de Roulet(2014a)], published one year after the beginning of this research.

---

<sup>1</sup>Daniel de Roulet - <http://www.daniel-deroulet.ch>

## Chapter 2. Context and Methodology



**Figure 2.1** – The ten books constituting «La simulation humaine» (Copyright Canevas Editeur, Editions du Seuil, Largeur.com SA, Buchet-Chastel)

*La simulation humaine* was written with one peculiar stylistic feature, in that all stories can be split up into chapters obeying a strict setting that follows one single narrative point of view, reminding to the way a lot of typical TV series are scripted. More precisely, the complete saga taken together can be divided in a chain of 298 chapters, most of them telling a part of the story as seen from the point of view of one main (and possibly several second) character(s), in some precise location and within a well-defined timeframe. In each book, the sequencing of those chapters may, however, obey various rules: some are chronological, others alternate the point of views of several characters until the day they meet, and so on. As the author noted himself (document 1.1):

*« Dans la majorité des cas, le chapitre peut se caractériser par le point de vue d'un seul personnage à la troisième personne du singulier. [...] L'époque historique à laquelle se passe ce récit est désignée parfois de manière très précise (jour et heure), parfois de manière plus vague, juste une saison et l'année (printemps 1943). Les lieux dans lesquels ces chapitres se déroulent sont toujours des lieux réels, les villes sont nommées, parfois de manière encore plus précise en indiquant le nom de la rue. »*

When Daniel de Roulet was drafting the seventh book, he conceived the idea of one day repub-

## 2.2. An Interdisciplinary Methodological Framework

---

lishing the complete saga in a digital format, and to exploit in a creative way the possibilities that would be offered by this media. The imagined «digital saga» was to be an extension of this upcoming last books, giving to both new and faithful readers another way of looking at the interleaving life paths of all characters and their ancestors. With that goal in mind, the author constituted nomenclatures and genealogical documentation for every member of the fictional families. In a final discussion with his publisher, Buchet Chastel, it was agreed that the saga would come to an end with the tenth book, whose release would be accompanied by a digital republication.

The existence of this peculiar configuration, with a novelist willing to explore the implications of such an operation on his own writing, constituted an outstanding case study, tailored for the newly founded EPFL Digital Humanities Laboratory<sup>2</sup>. It was made possible with the very generous help of the *Jan Michalski Foundation for Writing and Literature*, created in 2004 in Montricher by Vera Michalski-Hoffmann in her husband's memory to perpetuate their common commitment to those who devote themselves to the written word. The Foundation's mission is to foster literary creation and encourage the practice of reading through a range of initiatives and activities. In the past years it granted financial support to a considerable number of literary projects, ranging from festivals to public writing initiatives<sup>3</sup>.

With the active participation of all the aforementioned actors, what was initially thought to be a simple republishing on some digital media turned into a three years long research project, aiming to the study of the evolution of storytelling strategies linked to the use of reading and writing applications, through the release of at least one, to our knowledge new of its kind, digital book.

## 2.2 An Interdisciplinary Methodological Framework

The design and frame of our work imply peculiar methodological considerations that are definitely not most common in a research context. As the reader may expect by now, it would be an understatement to say that this work was done with an interdisciplinary approach. Of course, studies on emerging phenomenons often lack a well-defined field, but we will see that in our case many fundamentals can actually be anchored in well-known and documented theories. To understand the extent of the different fields involved and the subfields where their crossings happen, several interdependent aspects need to be taken into consideration:

---

<sup>2</sup>EPFL | Digital Humanities Laboratory - <http://dhlab.epfl.ch>

<sup>3</sup>Fondation Jan Michalski | Some supported projects - <http://www.fondation-janmichalski.com/en/bourses/quelques-projets-soutenus/>

## Chapter 2. Context and Methodology

- Contrarily to most experimental frameworks where researchers need to make sure the studied subjects stay free of their influence, in our situation we clearly expected the produced results to be useful, and as far as possible put into practice to in turn become part of the study. Furthermore, Daniel de Roulet lead an active role in multiple stages throughout the research which, indeed was indispensable. This kind of approach closely relates to the traditionally rather politically oriented **participatory action research (PAR)**, that we will present and discuss in section 2.3.
- From a thematic point of view, both digital medias and literary considerations are of equal importance. In practice as well, one can expect such a study to need tools and theoretical frameworks either issued in computer sciences or literary studies, in similar proportions. This problematic is far from isolated within contemporary studies, and closely relates to the kind of approach precognized in **digital humanities (DH)**, whose various benefits will be presented in details in section 2.4. In addition to that, the design of innovative and highly interactive tools as well as the nascent studies around digital medias relate to theories that are inherently part of the field of *DH*.
- Finally, the picture of course wouldn't be complete without considering the author's view on the topic. As a novelist participating in this research and if all other parts where to be considered a single black box to interact with, this work would most closely relate to a **constrained writing** experience, whose implications will be discussed in section 2.5.

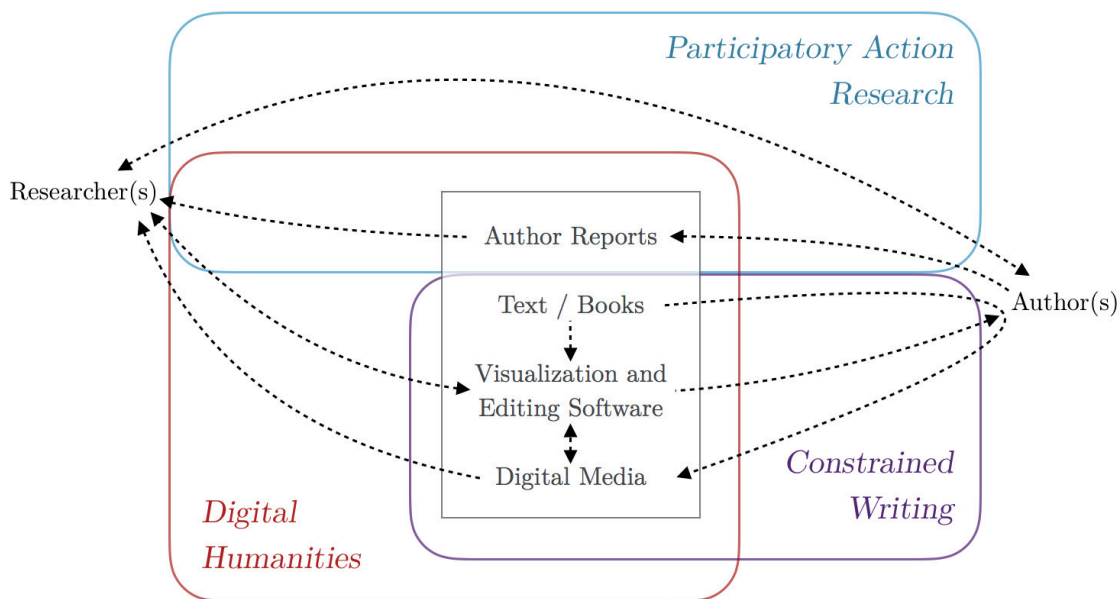


Figure 2.2 – Overview of the fields of study

All three fields are related in different ways, and connected through the various pieces of knowledge we needed, as illustrated in fig. 2.2. The continuous colored zones represent the boundaries of each field, and the dotted arrows indicate interactions and information flows. Notice that both parties interact to some extent with all fields and directly or indirectly influence all the produced and studied components.

### 2.3 Participatory Action Research

As its name suggests, Participatory Action Research *PAR* generally aims to take action or make progress happen in communities, by an iterative process that involves building common understanding between researchers and subjects through active participation. Although lacking a generally accepted definition and being considered an overly blurry field, its perimeter can be outlined by several characteristics [Baum(2006)]:

- The *Action* component implies the researcher to get actively involved and thus to claim a non neutral position towards his observations. Actually, he is literally looking for the effect that will emerge from the collaborative work with his subjects.
- Conversely, the *Participatory* component calls for the subjects to also take actively part in the research, pursuing a shared goal of building up and gaining common knowledge.
- In opposition to other research methods, *PAR* does not mainly emphasize on the reproducibility of the findings, but rather on the achievement of some desired effect.

#### 2.3.1 *PAR* Feedback Loop

The key idea of *PAR* lies in an iterative and cyclical process involving collective discussions and interaction with one's subjects. Both researchers and subjects are to learn from each other, and none have a predominant expert role. Their common goal is to recognize conflicting interests and the underlying structures in the first place, and then to elaborate an approach that would efficiently solve the identified problems [Moser(1977)]. The typical use cases for *PAR* include various sociological studies designed to understand and overcome social problems through direct and democratic interactions, including themes as variate as consumer welfare, feminist movements [Ozanne and Saatcioglu(2008)], adult education research [Mordock and Krasny(2001)] as well as miscellaneous industrial problems solving [Whyte(1989)].

At first sight, our context may seem quite remote to those, but the underlying frameworks have

## Chapter 2. Context and Methodology

---

also been successfully generalized and transposed to a variety of apolitical situations, given a context where both *Action* and *Participatory* aspects are needed [Eden and Huxham(1996)]:

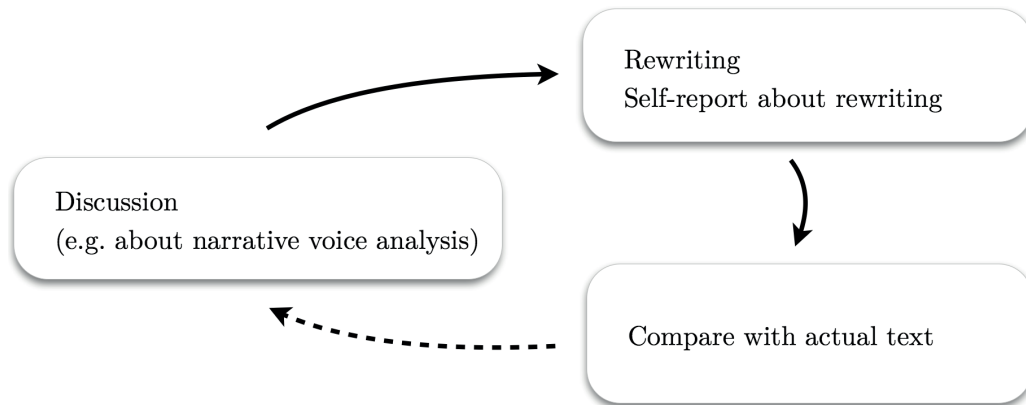
- The *Action* component is justified in situations focused on the study of upcoming phenomena, or where one has good reasons to think that noteworthy progress can be made provided a controlled amount of input. In the case of our study, this directly relates to the global context of emerging digital publication and its possible implications on the work of novelists.
- The *Participatory* component allows to take into account the commitment of the subjects (that is, if they are not neutral towards the possible outcomes or interestedly take part in the research process itself). In our case, this aspect is by design fully part of the study, and even if the single-to-single setting is quite peculiar, it has already been validly used [Horner et al.(2005), Mason(2010)].

Moreover, *PAR* fits well in a context where the research process involves the need for iterative refinements, for instance because specifications of the project are meant to evolve drastically, or the means to get to the results are fuzzy and defining them accounts for an important part of the research. *PAR* is also reputed a wiser choice when focusing on local relevance, preferring ecological validity over reliability [Kirk and Miller(1986)], or responsiveness over replicability [Dick(1993)].

### 2.3.2 Researcher Influence Bias

As we have seen, *PAR* studies often comprise an iterative process, which means that both uncovered concepts and goals are to be refined throughout the research. In our case, is typically to be expected within the rewriting loop (illustrated in fig. 2.3), where stories get modified according to initial analyses, and in turn giving way to new results. In such a loop, the general strategy pursued by *PAR* methods starts by defining the goals, means and ground rules of the research [Krishnaswamy(2004)]. This is traditionally executed by means of gatherings involving all parties, informal meetings, cooperative problem solving and information sharing, all aiming to build common understanding amongst participants. The first and essential component being to build up shared systematic knowledge through discourse, in democratic exchanges where all actors learn from each other, and where researchers must be very careful not to introduce uncontrolled influence biases. In our context, those aspects are moderated by the common grounding imposed by the creation of the digital edition, with this very piece of work standing as an evolving shared background and witness of common understanding.





**Figure 2.3** – Typical iterative process core to our research, that directly echoes to *PAR* methods.

On the one hand, results obtained using *PAR* are by construction strongly influenced by the researchers' actions and ideas. On the other hand, subjects are part of the justification for objectivity, as the results would necessarily be different (typically, out of reach) if the researchers were working alone. Both are necessary conditions for the progress of research, and conversely imply some caution to make sure the conclusions are objective and reliable. In that light, a standard criterion for objectivity is to ask oneself if the results would have been the same if driven by another researcher at the same time, given the access of the same resources. If fulfilled rigorously, it actually is a strong statement that pledges for a fully objective approach and thus reliable results. This criterion may partly be ensured by keeping track of the inputs that have been made, and by making sure those are backed up by proper information gathering beforehand. Indeed, this cannot account for creativity, and a lot of inputs may not be justifiable in *PAR* since they evolve from interactions that cannot be reproduced.

A good example of mutual influence can be seen within the frame of this thesis, as the joined effect of our cooperation lead to iterative changes in the manuscript of *Le démantèlement du coeur* (as the author explains in appendix A.3.1). Global structural modifications were applied in response to findings made through multiple visualizations computed on the early manuscripts and preceding books. In comparison to the first drafts established before the start of the project, the later ones first included a wider variety of narrative focalizations, and then later a more complex structural interleaving. An illustration of different steps of this reorganization is shown in appendix A.2.1. One immediately sees the ambiguous witness this evolution represents when rigorously compared to the previous statement, as even meticulous tracking of research input and changes made to the manuscript do not clearly account for the part of effects that could be generally described as serendipitous consequences.

Another approach enforcing validity is to use *data triangulation*, meaning to test for con-

## Chapter 2. Context and Methodology

---

sistency by confronting the results to those obtained using several different data sources or methods, whenever available [Patton(2002)]. This can be especially effective when one has access to and/or wants to mix both quantitative and qualitative methods [Jick(1979)]. In our case, triangulation could for instance be obtained by backing statements from the author with analytics on the text, some of which we will discuss in section 6.1.

Even so, achieving repeatability independently of the researcher might often not be possible in pure action research, and one may seldom be presented with several independent data sources to pick from. In *PAR*, a weaker (yet more realistic, and thus widespread) validity criterion often is to ensure the research process would be *recoverable* by third parties interested in it, by making clear «*the thought processes and models which enabled the team to make their interpretations and draw their conclusions*», in such a way that the following interpretations could be debated [Checkland and Holwell(1998)]. We enclosed in this report various original writings (documents 1.1, 2.5 and 4.4.5). providing relevant documentation with this goal in mind. Moreover, the technical results were extensively documented and made public, in the form publications ([Bornet et al.(2014), Bornet and Kaplan(2015), Bornet and Kaplan(2017)]), released open-source software code <sup>4</sup> <sup>5</sup> and a free to use web application for authors <sup>6</sup>.

Last but not least, action research experts also advised for gains in credibility by consensual pair validation [Eisner(1998)], which if done right is indeed a solid first step to recoverability. As we will see in our the late public workshop experiment results (section 6.3) and the interviews conducted thereafter (appendix A.6.1), the exposure of authors that weren't part of the study to our tools shed light on several common needs and practices in the realm of fiction writing.

### 2.4 Digital Humanities

The need for computational support in modern Humanities studies has been stressed by scholars from all walks of life since the general availability of powerful and cheap tools that potentially outdate the traditional research ways [Morón Arroyo(2002)]. Yet, many fields failed for a long time to benefit fully from those potential advances due to lack of shared knowledge, procedures, or tools tailored for their needs. Directly addressing this issue, the rapidly expanding field of Digital Humanities (*DH*) has become a growing topic amongst scholars of top universities in the recent years, within boundaries that are still subject to discussion [Svensson(2010)].

Nested at a crossing between Humanities and Computer Science, *DH* merely exist in response

---

<sup>4</sup>3N Tools source code (GitHub) - <https://github.com/dhlab-epfl/3n-tools>

<sup>5</sup>Saga+ source code (GitHub) - <https://github.com/dhlab-epfl/saga>

<sup>6</sup>Saga+ - <http://saga.plus>

to methodological gaps and respective lacks in approaches from both fields to use and interpret results with their pairs' methods. Conversely, where both scholars from Humanities and Computer Sciences typically strive to build knowledge by narrowing down their field of study, the Digital Humanist tries to widen one or several fields by including foreign data and methods, to come up with innovative approaches and possibly conclusions that differ in their very nature from preceding studies. Combining those views to produce new knowledge, the *DH* way typically merely implies an efficient and innovative reuse of otherwise well-known methods and techniques. Computational literary studies and their methods were thus implicitly made part of *DH*, driving the interests of a wide range of scholars that proceeded to document the techniques best suited for their needs, and produced various implementation recommendations.

One can immediately see why the (even blurry) bounds of *DH* perfectly encompass all the practical tools and methods needed for our research. Figure 2.4 summarizes the typical editing pipeline with a technological perspective, whose components we will discuss in details in the following sections.

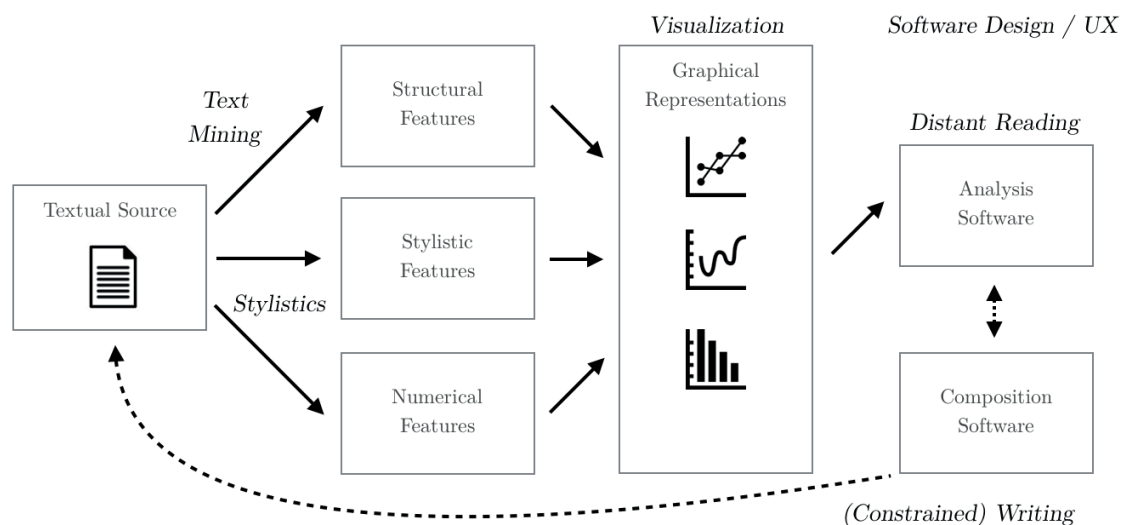


Figure 2.4 – Technical overview of the components in the analysis/editing pipeline

### 2.4.1 Text Mining

Most raw datasets used in social sciences are not fit to be processed directly by a computer software. Identification of the parts that are relevant for further computation may seem easy for an human, but are usually non-trivial for computers due to formatting issues, naming conventions, or, even worse, implicit information and metaphorical statements. This problem can be encountered with any kind of data involving natural language, and is particularly relevant

## Chapter 2. Context and Methodology

---

in the case of novels, in which the prose typically tries to avoid obviousness and frequent repetitions. In order to make a bridge between the raw data and analysis software, one possible approach can be the manual annotation of the texts, as it is done in domains requiring a higher accuracy, such as health and political sciences [Tapuria et al.(2007), Cardie and Wilkerson(2008)]. For very wide or less sensitive corpora, Text Mining technologies, who have their roots in the larger field of Data Mining, are a primer. They encompass the general process of pulling data of a certain kind, in a form that we want usable for further processing, out of bigger and non-formatted (or not standardized) datasets.

The specific task of Text Mining thus refers to the gathering of useful bits of textual resources, or more specifically, retrieving selected pieces of information out of text corpora that we may want as big and diverse as possible. It plays a core role in almost any literary computation since extraction of meaningful information is a mandatory step to any subsequent analytics, and of course gained much interest with the advent of *big data*. The general field is very broad, and ranges from pure information extraction from raw text, which plays an important role in the tools developed for our research, to supervised learning methods, automatic summarization and text classification, as well as expanding towards very recent data sources, such as multimedia and social media [Aggarwal(2012)]. Yet, it has to be emphasized that, as to be expected for an large and dynamic field spawning lots of diverging interests, its state of the art evolves quickly and varies a lot depending on the precise subfield one is considering, and the set goals to be attained.

Text Mining tasks, in a similar fashion as Information Retrieval, get usually evaluated on a known test corpus using two metrics named *Recall* and *Precision* [Van Rijsbergen(1979)]. To apply it on a said corpus  $S$ , one needs to know beforehand the amount  $R$  of relevant occurrences for a defined query  $q$  (for instance, we could consider  $R$  pages related to the query "*nuclear power*" within the  $S$  pages of *La simulation humaine*). After running the text mining task,  $S_q (\leq S)$  documents will be retrieved, out of which  $R_q (\leq R \leq S_q)$  documents will be relevant to the query (the others being false positives). We then define the *Recall* =  $\frac{|R_q|}{|R|}$  (as the overall proportion of relevant documents were returned), and the *Precision* =  $\frac{|R_q|}{|S_q|}$  (the proportion of returned documents that were relevant to the query). Since they are two distinct measures that usually act as a trade-off (as most tweaks set to improve one tend to be detrimental to the other), it is usually easier to compare evaluations by combining them beforehand. One way to do this is to compute the harmonic mean between them,  $\frac{2}{\frac{1}{P} + \frac{1}{R}}$ , which is often called the *F<sub>1</sub> Score*.

### 2.4.2 Stylistics, Stylometry and Authorship Attribution

One of the most obvious specific subfields in literary computing is modern stylistic analysis, which aims at featuring differences in styles across genres, authors or individual pieces of work. From there follows authorship, the counterpart whose purpose is to relate a text to the person that wrote it, in particular by using those very stylistic features. Both ideas are by no means new and have probably been topics of interest since early ages of writing and forensics, but the use of computer methods obviously quickly came in as a dramatical game changer [Schreibman et al.(2004)]. Nowadays, cross-analyses on wordings in large corpora allows for findings that may differ a lot in their nature, such as detecting evolution of linguistic features spanning several centuries [Buntinx et al.(2017)].

Conversely, where stylistics broadly comprise any way of underlining and separating formal structures from the meaning, approaches featuring only numerical analyses spawned the sister field of stylometry, which also follows a century old tradition of research with notable recent improvements and indeed received a lot of interest with the help of growing computational power [Holmes(1998)]. Furthermore, beyond the automation of simple word counting and statistics computing, new methods such as Principal Components Analysis (*PCA*) allow to easily look for the most differentiating stylistic features. Yet, the validity of those results is still subject to debate as there is no consensus about the interpretation they deserve, as variety of genres and languages as well as their evolution through time make methods difficult to generalize, and consistent failure to do so in turn made some scholars dubious about the field as a whole [Craig et al.(2006)].

Fortunately, establishing ground truth is much easier in the more particular case of authorship attribution, the subfield which aims at using stylistic features to determine a posteriori the author of a text. Paradoxically, this discipline, driven by the needs of forensics and possibly helped by a better defined and tangible frame, federated scholars that started putting up conclusive theoretical backgrounds. The extensively formalized methods, aiming to fight disorganization by distinguishing the good from mediocre ones, contributed to a more principled frame [Juola(2006)]. Various recent authorship attribution methods were now proven to be able to relate texts to unknown authors with a reasonable precision [Stamatatos(2009)], and thus backed the idea of consistent stylistic features that, with the help of recent text processing technologies and given enough examples, can be trusted even for more general categorizations, such as genres [Malrieu and Rastier(2001)] or datation [Smith and Kelly(2002)].

### 2.4.3 Visualizations and Visual Analytics

Data visualization plays a core role in many scientific applications, including more efficient pattern and knowledge discovery, intuitive understandings of all sorts of phenomena, central means in data mining related tasks and decision support activities [Fayyad et al.(2002)]. Here too, increase in computational power did lead to research paradigms shifts, especially in literature where studies have shown the benefits of using data visualization techniques by allowing to massively expand the available data in comparison to otherwise well known datasets [Kirschenbaum(2007)].

As theorized by Martyn Jessop [Jessop(2008)], textual data visualization can either aim to present statistical or numeric data features previously extracted, or, more recently, visual representations of the texts themselves as actual forms of text analysis. Like it is often the case in emerging fields, those techniques still tend to be fitted to specific use cases, and rose multiple debated questions regarding the most suitable ways of representing textual information, and their relevance in relation to the interpretation of texts.

A very good example of lesser conventional forms of text visualization that emerged specifically with the interdisciplinary nature of *DH* are *word clouds*. Word clouds (or *tag clouds*) are textual representations where single or groups of words are agglomerated in an unsorted way, with their relative sizes defined in relation to their usage frequencies. Even though methodologically controverted, they became recently very popular both as a mean of displaying final data and for pattern discovery [Viégas and Wattenberg(2008)]. The trend went on to get gradually accepted in the field, first as a mean to pull features out of sets of texts [Gottron(2009), Clement et al.(2009)] before to become a core feature of both specialized [Culy and Lyding(2011), Pence(2016)] and general purpose state of the art data visualizations tools, such as *Wordle* [McNaught and Lam(2010)] and *Voyant* [Xu et al.(2014)].

At a greater scale, projects aiming to centralize the most broadly used visualization resources into toolsets, such as *TAPoR*<sup>7</sup> and *MONK*<sup>8</sup>, received enough interest to federate substantial communities. Those resources are very precious for fast prototyping and establishing a baseline against which to try new approaches. However, despite the plethora of projects they encompass, they don't eliminate the need to invent and develop derivatives (or sometimes entirely new solutions) in order to bring innovation to end-users.

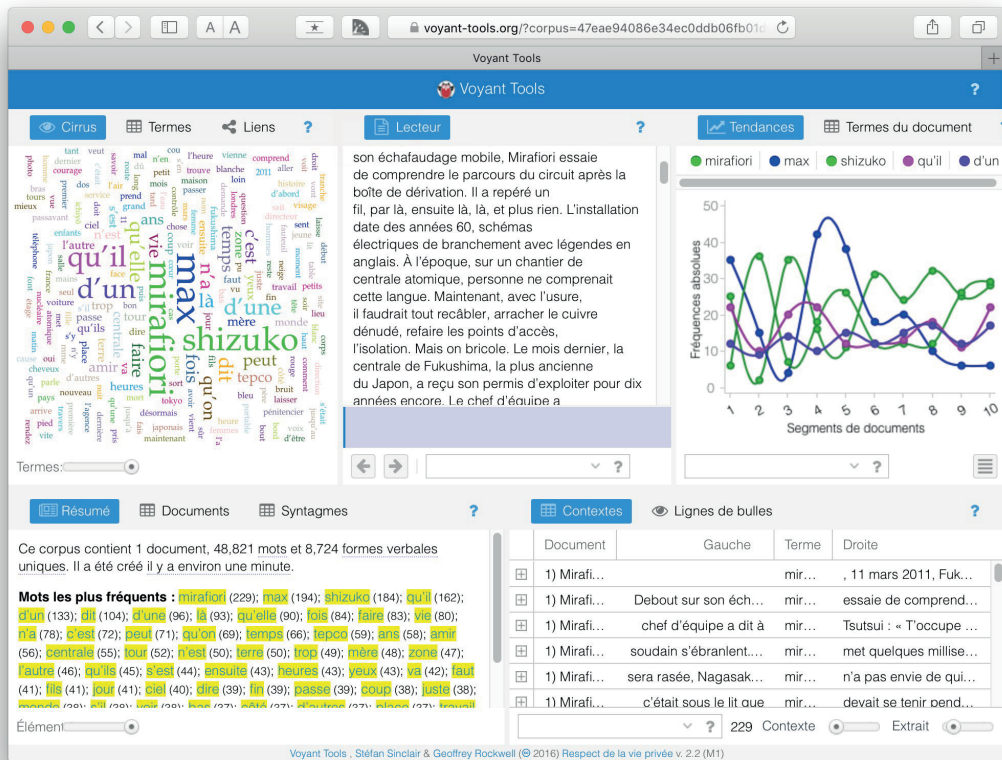
When getting to try out custom visualizations in the frame of this thesis, the open source general purpose tool *GraphViz*<sup>9</sup> represented an interesting trade-off, as it gives the possibility

---

<sup>7</sup>TAPoR | Discover research tools for studying texts - <http://tapor.ca>

<sup>8</sup>Metadata Offer New Knowledge (MONK) - <http://monk.library.illinois.edu>

<sup>9</sup>Graphviz | Graph Visualization Software - <http://www.graphviz.org>



**Figure 2.5** – Visualization of *Le démentèlement du coeur* using *Voyant Tools* by Stéfán Sinclair and Geoffrey Rockwell (CC BY 4.0), December 2016.

of generating a textual description for various predefined types of graphs that can be automatically rendered as either vector and raster images [Ellson et al.(2002)]. For more custom attempts and since we knew we would want most visualizations to eventually get interactive, the web framework *d3.js*<sup>10</sup> provided a sound programming environment [Qi Zhu(2013)]. Despite a steep learning curve, its advantages are a wide compatibility and a large adoption throughout various web developer communities including in academia that featured endless implementation possibilities. Performance-wise, it has also been argued to be suitable for big data applications [Fan Bao and Jia Chen(2014)].

### 2.4.4 Distant Reading

With literary studies today facing challenges that were only scarcely explored in the field of NLP (and vice-versa), the *DH* approach to literature is a very sensible, and actually amongst its most obvious and direct field of influence. As a matter of a fact, Computer Science has had from its

<sup>10</sup>D3 | Data-Driven Documents - <http://d3js.org>

## Chapter 2. Context and Methodology

---

early days on a long history of contributions in text processing, be it for automatic translation or attempts to understanding human language, yet scarcely tackling issues related to fiction writings. Furthermore, if stylometry indeed targets those corpora (amongst others) and uses massive and numerical interpretations, the expected conclusions are usually detached from the contents and meanings of the texts, and with the notable exception of the *NovelTM* initiative<sup>11</sup>, very few approaches explored the possibilities of applying computational methods to literary works with an humanities' perspective.

Fortunately, as top researchers in the field noted [Rommel(2004)], new tools and corpora made widely available in the late 1980s recently made literary scholars consider the importance and significance of the literary computing approach. Indeed, in a world where traditional literary studies almost exclusively relied on well known selected excerpts and painstakingly acquired knowledge about the works, computers made complete surveys approaches possible, expending drastically the studied material both in depth and scope, and turning results to factual and reproducible data. Conversely, the release of a new generation of tools, tailored for literary studies and built to meet the requirements of scholarly computing, made possible new kinds of approaches that drove the whole field to change.

As a precursor of this newly unfolding field, Franco Moretti showed that looking at entire corpora through statistical representations and projections of their features could indeed constitute a valid approach to literature [Moretti(2005), Moretti(2013)]. This so-called *distant reading* mainly relies on empirical and fact-driven methods, and is aimed at getting a broad understanding of the general underlying structures of literature rather than the traditional, meaning understanding driven, way. As Moretti explains himself:

*«At bottom, [close reading] is a theological exercise - very solemn treatment of very few texts taken very seriously - whereas what we really need is a little pact with the devil: we know how to read texts, now let's learn how not to read them. Distant reading: where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems. »* [Moretti(2000)]

Therefore, *distant reading* approaches are thought with the goal of understanding systems in their entirety by looking at the big picture and shadowing the details. By doing so, they allow to take into consideration the very vast majority of works that are contemporary, similar, or otherwise relevant to any well known classics, as opposed to the regular scholarly tradition that mainly focus on few select samples. As a direct effect, the analysis paradigm shifts from purely

---

<sup>11</sup>NovelTM | Text Mining the Novel. A Multi-University Digital Humanities Initiative - <http://novel-tm.ca>



---

## 2.5. Constrained Writing and Sequencing Techniques

qualitative to arguably more refutable, quantitative methods. As a matter of fact, despite their novelty and lack of ground truth that sparked debate about their legitimacy [Arac(2002)], such techniques have been used in recent studies in order to point out very general trends in literature, without necessarily claiming a literary approach [Jockers(2011), Michel et al.(2011)].

### 2.4.5 Digital Contents and Software Design

When confronted to questions of forms, *DH* scholars explored and theorized about the best ways to encode and use electronic documents, and their possible impacts on future research and preservation [McCarty(2011), Siemens and Schreibman(2013)]. Furthermore, the generalized use of hypertext sparked many debates ranging from its applications to update existing contents to whole new economics and ways of consuming cultural goods, where the text paradoxically became of unprecedentedly central importance [Landow(2006)].

Last but not least, several scholars in *DH* proposed noticeable contributions relevant to our research, relating to improve the software development cycle through dealing efficiently with end users. Claire Warwick [Warwick(2012)] summarized a methodological theory reflecting years of practical experience in improving digital access to data for typical Humanities users. Her recommendations include establishing a good knowledge of the end users, and to design the products with them mind and if possible, their active participation. Well known psychological biases can be reduced by replacing user surveys with hard data whenever possible, which is one of the rich features made possible by connected platforms with their access logs. Development should be conducted with the help of staff knowledgeable in *DH* techniques and access to good technical support, and the final product should be released with proper documentation and plans for active updates.

## 2.5 Constrained Writing and Sequencing Techniques

The idea of prosaic works following external or self-imposed constraints, though not most common, is not unique in the wide field of fictional literature. Sometimes used as a source for inspiration, or sometimes to bound their uncontrollable creativity, various authors used sets of predefined rules to compose their works [de Geest and Goris(2010)].

A very representative and essential part of constrained writing was explored by the members of *l'Ouvroir de Littérature Potentielle (OuLiPo)*, which includes Georges Perec and Raymond Queneau. To them, there is a need in literature for gratuitous constraints that set the choice of words, or the order of chapters and even sentences [Ouvroir de Littérature Potentielle(1990)].

## Chapter 2. Context and Methodology

---

The case of Georges Perec is very well documented. In his autobiographical reports, he himself explains how his writing was about finding constraints involving mathematical games to finally compose the text of his novel *La Vie mode d'emploi* [Perec et al.(1995)]. This can be seen as nostalgia of the classic forms of French poetry, with its formal rules that determine the forms of the poem (number of syllables, rhymes, verses arrangement): at a time when the 20th century poetry is all about getting rid of rules, *OuLiPo* is enacting new, even more binding ones, some even going as far as the prohibition or compulsion to use specific words or letters (e.g. *La Disparition* was written without ever mentioning the letter *E*). In that respect, the author's work is done under extreme constraints and promising for an aesthetic result similar to what a crossword puzzle does.

In a similar but different way, Raymond Queneau's *Exercices de Style* [Queneau(2004)] exaggeratedly and extensively explored the various ways of retelling over and over a one page short story by opposing very different and caricatural narrative styles. This work is another good example of the genre, because out of a plain story without narrative significance, the repetition and variation of styles build up another dimension, almost out of the literary world, carrying its own suspense and giving to the chapters a transversal sense of unity.

In a broader sense, it is interesting to think about constrained writing at the higher level of sequencing (in the sense of ordering meaningful pieces, and thus shifting the composition process from the sentences level to the fragments). When structuring and ordering the different parts of the story, the author is imposing a constraint on the way the events will unfold to the reader. Alan Ayckbourn stressed this idea with his play *Intimate Exchanges* [Ayckbourn(1985)], where a single choice at a moment in the story (e.g. to smoke a cigarette or not) will lead to a chain of different narratives, and in the end a different outcome for each characters (fig. 2.6).

## 2.5. Constrained Writing and Sequencing Techniques

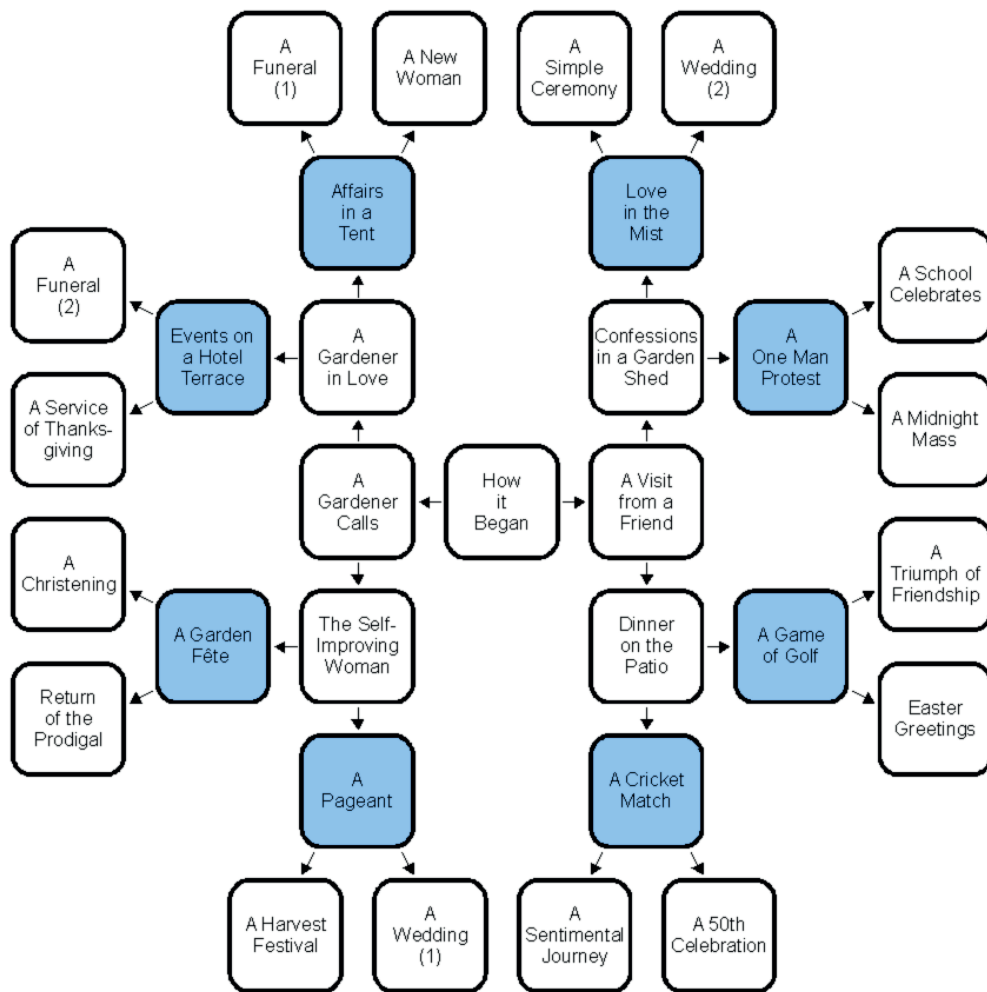


Figure 2.6 – The 16 possible readings of Alan Ayckbourn’s *Intimate Exchanges*. Source: Wikipedia.

In 1834, Honoré de Balzac wrote a letter to Ewelina Hanska about his project consisting of reuniting previously written novels in an enormous depiction of the society of that time. His saga, that he would later call *La Comédie Humaine* was to consist, even though unfinished, of 90 books. For the following 16 years, he wrote new books that exactly fitted in missing spots of his master plan, and redacted 23 of the existing ones to make them fit in different spots. Unlike other writers, Balzac considered his world a present and living artifact, whose 4 000 characters thus could not be followed chronologically, exactly like in real life [Neefs and Barel-Moisan(2015)]. The writing constraint here is methodological and acts in a balanced way both as a structure for existing pieces and a guideline for new ones, aiming at a detailed organized depiction of the society, with recurring characters whose life paths are

## Chapter 2. Context and Methodology

scattered among several books.

This reminds of the methodologies used by Claude Simon, who is indeed also known for his narratives comprising multiple trajectories meeting over and over in the same places, depicting different visions of the same reality [Calle-Gruber and Simon(2011)]. About his novel *La Route des Flandres*, he commented how even though the whole idea was present from the beginning, he proceeded to write in small unorganized fragments that he called *tableaux détachés* and needed a further assembly strategy. He then came up with a color scheme for every character and event, an example of which is shown in fig. 2.7, allowing him to determine his optimal narrative strategy.

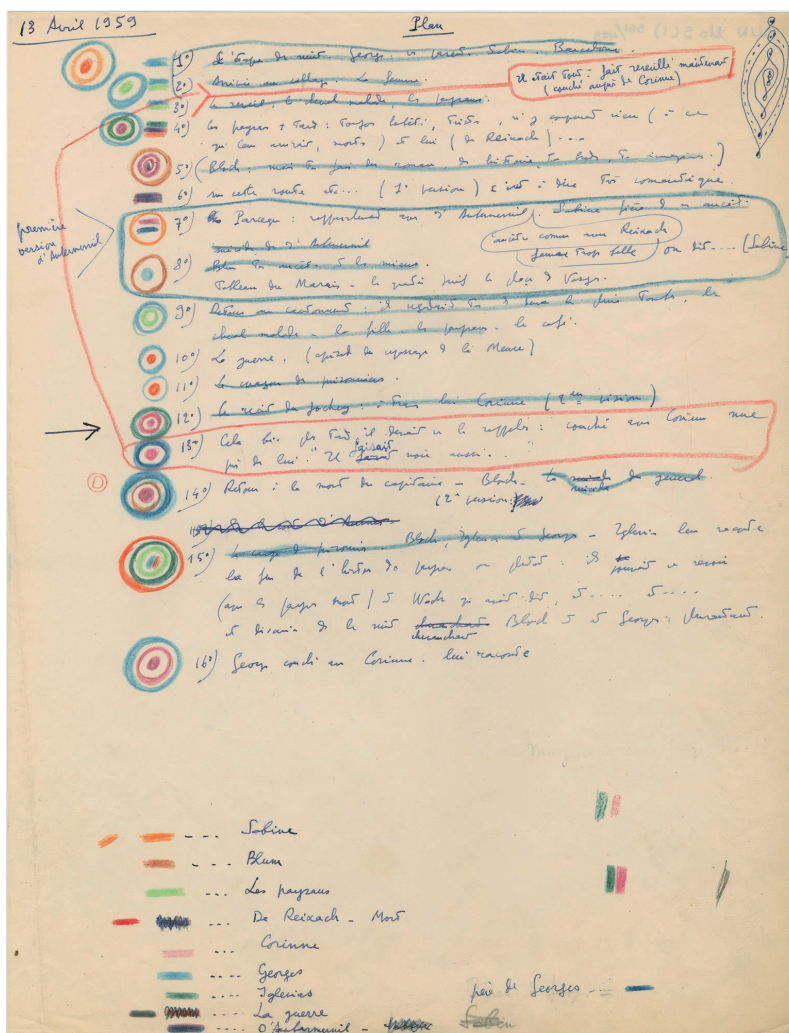


Figure 2.7 – Original manuscript by Claude Simon, exposing his structural coding strategies when working on *La Route des Flandres*. Copyright © Réa Simon.

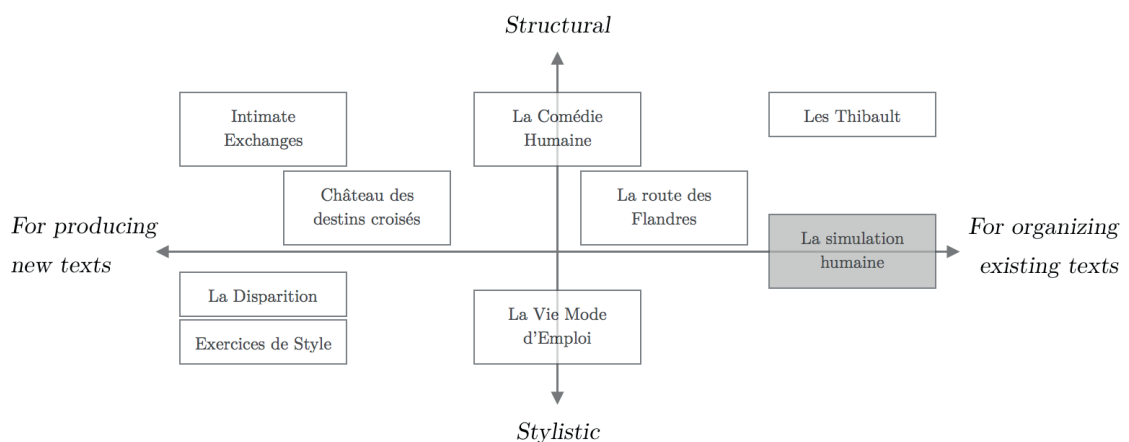
In some places where the colors seemed too monotonous for a too long period of time, he'd

## 2.5. Constrained Writing and Sequencing Techniques

fit in another part recounting of other events or seen from the eyes of other people, or wrote some new bits purposely in order to balance the various elements. According to the author, some of the best parts are amongst those that were drawn by this method [Dällenbach(1988)].

In comparison, these strategies are in nature again quite different from the one used by Roger Martin du Gard when he wrote *Les Thibault*. As he explains, the structure of his saga was built up from paper notes stacked on tables, involving a painstaking process of writing down tiny fragments and carefully choosing where they should belong to, slowly forming consistent and ordered records. The rationale of doing so was to have the entire raw material ready before he even started writing out the first actual words of the story, in a similar way an architect preconceives a piece of work in its entirety before starting building it [Martin DuGard and Camus(1993)]. In that case, one can see a structure emerging from the contents and acting as an organizational constraint that, contrarily to preceding authors, doesn't drive the writing but the final composition process.

We see that constrained writing can take plenty of forms, that can be summarized following two main axes. Where constraints can range from purely stylistic to structuring, they can at the same time be used as a canvas to produce contents, or to organize preexisting texts. We summarized these projections in fig. 2.8, which allows us to introduce the approach used within the context of this research, with its constraints touching both structural and stylistic effects, but used exclusively for organizing an existing corpus. One will immediately notice the absence of major works in the lower right quarter of this topology, which is indeed a combination that doesn't easily relate to original works, but would typically be the kind of process followed when producing a thematic compilation of existing works with an intended stylistic effect, as it is for instance the case in collections of poems.



**Figure 2.8** – Qualitative comparison of constrained writing techniques

## **Chapter 2. Context and Methodology**

---

Of course, in a way or another, any novelist conceiving a lengthy masterpiece is facing such complexity they need additional tools to overcome it. Structuring and sequencing stories both call for tools that can be drawn from writing constraints or built out of visualization, statistical or organizational strategies. In particular, sagas are by essence artworks where the number of parameters and possible outcomes make linear writing a limiting method. With that perspective in mind, it becomes relevant to think about the potential contributions of digital solutions to this process, via or leading to a form of constraint writing, and in particular what their impacts on modern authors' workflows might be.

# Document 2

Daniel de Roulet, 24.03.2014

Hello Cyril,

Je me suis bien amusé avec cette nouvelle version, bon travail de ta part. J'ai quelques améliorations à proposer.

Sur la page d'accueil, je trouve qu'il y a trop de choix, j'ai donc revu le tout surtout pour qu'on ne retrouve jamais un chapitre déjà lu. Jusqu'ici les chapitres isolés (avant-goût) étaient remis en lecture, ce qui rendait difficile le compte de ce qui restait. Désormais on peut vraiment enlever les chapitres lus du corpus. Si on a lu tous les choix proposés désormais (3 avant-goûts, 3 nouvelles, 3 romans) il reste 174 chapitres qui forment un tout cohérent (cf. les 2 fichiers annexés).

Pour la nomenclature, j'ai corrigé des tas d'ambiguïtés qui venaient du fait que des personnages ont des synonymes. Ainsi il y a Kumo et Kumo le pianiste, et Shizuko et Mlle Shizuko, il y a Elisabeth tout court et Elisabeth II, la reine d'Angleterre. Et quand il y a un surnom il est aussi mentionné Corbu pour Le Corbusier et Sid pour Sidharta, Oppie pour Oppenheimer, etc. Manquaient aussi quelques noms comme Hiroshima et Jean Tatlock. D'où une nouvelle nomenclature où il faut pointer tous les mots en gras. Et un nouveau corpus où j'ai corrigé les ambiguïtés de noms.

Pour l'application elle-même, à part cette première page que je simplifie, j'ai rencontré

quelques questions dans le désordre :

- quand on surligne un mot qui n'est pas dans la nomenclature, mon écran d'iPhone dit COPY/DEFINE et si je presse l'un ou l'autre, ne peut rien en faire. Peut-on désactiver cette fonction.
- parfois certains mots sont accolés et je n'ai trouvé aucune logique à ce phénomène et ça ne vient pas du corpus
- il est arrivé que le programme se plante en affichant « null » ??
- est-ce qu'il y a une raison pour laisser un blanc en haut d'un chapitre affiché? et un blanc après le chapitre? Est-ce qu'on pourrait écrire en gras ou en capitale pour mieux différencier du texte: CHAPITRE SUIVANT et CHOISIR UN AUTRE TEXTE (plutôt que choisir un autre livre)
- comme tu l'as proposé, il faudrait sans doute un bouton pour retourner à la page des choix.
- dans ma version, le compte des chapitres lus s'arrête à 9 (Vous avez lu 9 chapitres sur 297) et ne s'incrémente plus
- le titre La simulation humaine, je l'écris avec des minuscules. Ou bien en capitales, ce serait encore mieux

## Document 2

---

- quand on tourne l'écran (tu l'as vu), on perd l'agrandissement après un nouveau choix de chapitre
  - à l'horizontale on a deux bandes bloquées, une en haut, une en bas de l'écran. Elles gênent la lecture quand on se met à lire vraiment. E-books et Kindle, les suppriment
- quand on presse au milieu de l'écran)
- Je vais encore réfléchir à l'écran et aux informations qu'il faudrait dans « A propos » L'idéal serait de pouvoir y mettre aussi l'arbre généalogique. Je te le mets en annexe pour que tu me dises ce qu'on pourrait faire.
- 

### Réordonnement des chapitres de *La simulation humaine*

*Une fleur de cerisier* (1 chapitre)

1.28

*Sur les barricades* (1 chapitre)

2.29

*Tchernobyl en Ukraine* (1 chapitre)

2.1

*Air Force One* (4 chapitres)

2.4, 2.15, 2.40, 2.53

*Fukushima au début* (5 chapitres)

10.1, 10.4, 10.6, 10.10, 10.12

*Le pucelage d'un Helvète* (10 chapitres)

9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 9.10

*L'ingénieur et la fillette* (30 chapitres)

8.1, 8.2, 8.3, 6.2, 6.4, 6.6, 8.4, 8.5, 8.6, 8.7, 6.8, 6.10, 6.12, 6.14, 8.8, 8.9, 8.10, 8.11, 6.16, 8.12, 6.18, 8.13, 6.20, 8.14, 6.22, 8.15, 6.24, 8.16, 8.17, 6.26

*Fissions* (33 chapitres)

1.6, 1.7, 1.8, 1.9, 1.11, 1.12, 1.20, 1.23, 1.26, 1.30, 1.32, 2.6, 2.7, 2.8, 2.9, 2.12, 1.33, 2.19, 2.21, 2.22, 1.35, 1.36, 2.25, 2.27, 2.31, 2.33, 2.35, 2.38, 2.42, 2.43, 2.45, 2.47, 2.1, 2.51



*La réalité, mais digitale* (38 chapitres)

4.1, 4.2, 4.3, 4.4, 4.5, 3.1, 3.2, 3.3, 3.4, 3.5, 4.6, 4.7, 4.8, 4.10, 4.9, 3.6, 3.7, 3.8, 4.11, 4.12, 4.13, 4.14, 4.15, 3.9, 3.10, 3.11, 3.12, 3.13, 4.16, 4.17, 4.18, 3.14, 3.15, 3.16, 3.17, 4.19, 4.20, 4.21

*La simulation totale* (298 chapitres)

5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 1.6, 1.7, 1.8, 1.9, 1.11, 1.12, 1.20, 1.23, 1.26, 1.30, 1.32, 2.6, 2.7, 2.8, 2.9, 2.12, 1.33, 2.19, 2.21, 2.22, 1.35, 1.36, 2.25, 2.27, 2.31, 2.33, 2.35, 2.38, 2.42, 2.43, 2.45, 2.47, 2.51, 5.8, 5.9, 5.10, 5.11, 5.12, 5.13, 1.1, 1.3, 1.5, 1.10, 1.13, 1.15, 1.16, 1.18, 1.19, 1.21, 1.22, 1.25, 1.29, 1.31, 1.34, 5.14, 5.15, 5.16, 1.2, 1.4, 1.14, 1.17, 1.24, 1.27, 1.28, 2.17, 2.18, 2.20, 2.23, 2.24, 2.26, 2.28, 2.29, 2.30, 2.32, 2.34, 2.41, 2.44, 2.46, 2.48, 2.49, 2.1, 2.2, 2.3, 2.5, 2.14, 2.16, 2.10, 2.11, 2.13, 2.39, 2.36, 2.37, 2.50, 2.52, 2.54, 2.4, 2.15, 2.40, 2.53, 5.17, 5.18, 5.19, 5.20, 4.1, 4.2, 4.3, 4.4, 4.5, 3.1, 3.2, 3.3, 3.4, 3.5, 4.6, 4.7, 4.8, 4.10, 4.9, 3.6, 3.7, 3.8, 4.11, 4.12, 4.13, 4.14, 4.15, 3.9, 3.10, 3.11, 3.12, 3.13, 4.16, 4.17, 4.18, 3.14, 3.15, 3.16, 3.17, 4.19, 4.20, 4.21, 3.18, 3.19, 3.20, 3.21, 3.22, 3.23, 3.24, 3.25, 3.26, 3.27, 3.28, 3.29, 5.21, 5.22, 5.23, 5.24, 5.25, 5.26, 5.27, 6.1, 6.3, 6.5, 6.7, 6.9, 6.11, 6.13, 6.15, 6.17, 6.19, 6.21, 6.23, 6.25, 6.27, 7.1, 7.2, 7.4, 7.5, 7.7, 7.8, 7.10, 7.11, 7.13, 7.15, 10.1, 10.4, 10.6, 10.10, 10.12, 10.15, 10.17, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 9.10, 9.11, 7.3, 9.12, 9.13, 9.14, 9.15, 9.16, 9.17, 9.18, 9.19, 9.20, 7.6, 9.21, 9.22, 9.23, 9.24, 9.25, 7.9, 9.26, 9.27, 9.28, 9.29, 7.12, 9.30, 9.31, 9.33, 9.34, 9.35, 7.14, 9.36, 9.37, 9.38, 9.39, 9.40, 9.41, 9.42, 9.43, 9.44, 9.45, 9.46, 9.32, 9.47, 9.48, 8.1, 8.2, 8.3, 6.2, 6.4, 6.6, 8.4, 8.5, 8.6, 8.7, 6.8, 6.10, 6.12, 6.14, 8.8, 8.9, 8.10, 8.11, 6.16, 8.12, 6.18, 8.13, 6.20, 8.14, 6.22, 8.15, 6.24, 8.16, 8.17, 6.26, 8.18, 8.19, 8.20, 8.21, 8.22, 8.23, 10.2, 10.3, 10.5, 10.7, 10.8, 10.9, 10.11, 10.13, 10.14, 10.16, 10.18

*La simulation totale* (298 chapitres) reste : 175 chapitres

5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12, 5.13, 1.1, 1.3, 1.5, 1.10, 1.13, 1.15, 1.16, 1.18, 1.19, 1.21, 1.22, 1.25, 1.29, 1.31, 1.34, 5.14, 5.15, 5.16, 1.2, 1.4, 1.14, 1.17, 1.24, 1.27, 2.17, 2.18, 2.20, 2.23, 2.24, 2.26, 2.28, 2.30, 2.32, 2.34, 2.41, 2.44, 2.46, 2.48, 2.49, 2.2, 2.3, 2.5, 2.14, 2.16, 2.10, 2.11, 2.13, 2.39, 2.36, 2.37, 2.50, 2.52, 2.54, 5.17, 5.18, 5.19, 5.20, 3.18, 3.19, 3.20, 3.21, 3.22, 3.23, 3.24, 3.25, 3.26, 3.27, 3.28, 3.29, 5.21, 5.22, 5.23, 5.24, 5.25, 5.26, 5.27, 6.1, 6.3, 6.5, 6.7, 6.9, 6.11, 6.13, 6.15, 6.17, 6.19, 6.21, 6.23, 6.25, 6.27, 7.1, 7.2, 7.4, 7.5, 7.7, 7.8, 7.10, 7.11, 7.13, 7.15, 10.15, 10.17, 9.11, 7.3, 9.12, 9.13, 9.14, 9.15, 9.16, 9.17, 9.18, 9.19, 9.20, 7.6, 9.21, 9.22, 9.23, 9.24, 9.25, 7.9, 9.26, 9.27, 9.28, 9.29, 7.12, 9.30, 9.31, 9.33, 9.34, 9.35, 7.14, 9.36, 9.37, 9.38, 9.39, 9.40, 9.41, 9.42, 9.43, 9.44, 9.45, 9.46, 9.32, 9.47, 9.48, 8.18, 8.19, 8.20, 8.21, 8.22, 8.23, 10.2, 10.3, 10.5, 10.7, 10.8, 10.9, 10.11, 10.13, 10.14, 10.16, 10.18

## Document 2

3N	Entreprise spécialisée dans le traitement des déchets.
Big E	Abréviation pour Big Enterprise, spécialisée dans le traitement des déchets.
Bleu Siècle	Entreprise de traitement des déchets, née de la fusion de Big E et de 3N.
Greenwar	Organisation écologiste radicale
Ordre des Prétendants	secte et entreprise industrielle.
Saint Bellevue	Hôpital universitaire fictif à Harlem
Fermi Enrico	(1901-1954) physicien nucléaire italien, prix Nobel 1938.
Gorbatchev Mikhaïl	dit Gorby, né en 1931, dirige l'Union soviétique de 1985 à 1991.
Heisenberg Werner	(1901-1975) physicien allemand, prix Nobel 1932, fondateur de la mécanique quantique.
Le Corbusier	dit Corbu (1887-1965), architecte autodidacte né à la Chaux-de-Fonds.
McCarthy Joseph	(1908-1957) sénateur des Etats-Unis.
Meitner Lise	(1878-1968) physicienne autrichienne.
Oppenheimer Robert J.	dit Oppie, (1904-1967) physicien responsable du projet Manhattan.
Pauli Wolfgang	(1900-1958) physicien autrichien, prix Nobel, professeur à Zurich.
Reagan Ronald,	dit Ronny, (1911-2004) président du syndicat des acteurs de Hollywood, puis président des Etats-Unis.
Sakharov Alexis	(1921-1989) physicien nucléaire, père de la bombe à hydrogène soviétique.
Scherrer Paul	(1890-1969) physicien, directeur du programme de la bombe atomique suisse.
Staline Joseph	(1878-1953) secrétaire général du parti communiste, a dirigé l'URSS jusqu'à sa mort.
Sweeney Charles W.	(1919-2004) aux commandes du bombardier qui a largué la bombe atomique sur Nagasaki.
Tatlock Jean	(1914-1944), psychiatre, maîtresse de R. Oppenheimer, membre du parti communiste des Etats-Unis.
Truman Harry S.	(1884-1972) vice président des Etats-Unis puis président de 1945 à 1952.
Turing Alan	(1912-1954) mathématicien britannique, fondateur de la science informatique.
von Neumann Johnny	(1903-1957), mathématicien hongrois, émigré aux Etats-Unis.
Zuse Konrad	(1910-1995) ingénieur allemand, père du premier ordinateur électro-mécanique programmable.
Davos	station touristique d'altitude en Suisse.
Fukushima	site d'une catastrophe nucléaire au Japon, le 11 mars 2011.
Hiroshima	ville japonaise détruite par une bombe atomique le 6 août 1945.
Kaiseraugst	site prévu pour une centrale nucléaires aux portes de Bâle.
Kolyma	région à l'extrême nord-est du continent asiatique où se trouvaient les pires camps staliniens.
Los Alamos	haut plateau du Nouveau Mexique, centre du projet Manhattan dès 1942.
Lucens	site souterrain à 25 km de Lausanne, première catastrophe nucléaire mondiale par fusion du cœur.
Malville	site au fil du Rhône du surgénérateur nucléaire Superphénix.
Nagasaki	ville japonaise détruite par une bombe atomique le 9 août 1945.
Tchernobyl	site d'une catastrophe nucléaire en Ukraine, le 26 avril 1986.

Three Miles Island	site d'une catastrophe nucléaire aux Etats-Unis, le 28 mars 1979.
Felicity	1972, née à Harlem, New York.
Fumika Tsutsui	née en 1924 à Nagasaki, pianiste, mère de Shizuko.
Georges vom Pokk	né en 1962 en Californie, fils de Marthe et Jean-Paul vom Pokk
Ichiyô Tsutsui	1975, née à Nagasaki, fille de Shizuko, sœur de Mirafiori.
Jean-Paul vom Pokk	né à La Chaux-de-Fonds en 1927, dit JP, ingénieur polytechnicien.
Kumo vom Pokk	née en 1991, fille Mirafiori Tsutsui et de Vania vom Pokk.
Kumo Tsutsui	ou Kumo le violoniste, né en 1967 d'une fécondation in vitro.
Marthe vom Pokk	née Rouger, en 1938 à Téhéran.
Max vom Pokk	né à Ornans en 1945, fils de Virginie et Max-Paul vom Pokk.
Mirafiori Tsutsui	dit O-Bleu, né à Nagasaki en 1969, fils de Shizuko Tsutsui.
Paul vom Pokk	dit PVP, né en 1896, baron de l'industrie helvétique, propriétaire de Bleu Siècle.
Shizuko Tsutsui	née le 9 août à Nagasaki 1945, fille de Shizuko et Tetsuo Tsutsui.
Mlle Shizuko	ou Shizuko Watanabe, née aux Etats-Unis en 1919 pianiste, amie de Fumika.
Tetsuo Tsutsui	, né à Nagasaki en 1917, pilote de guerre.
Vania vom Pokk	, née en Californie en 1960, fille de Marthe et Jean-Paul vom Pokk.
Vladimir vom Pokk	, 1946, médecin suisse, dit Professeur Work, fils de Virginie et Jean-Paul vom Pokk.
Wolfgang Steinamhirsch	, dit Wolfie, né en 1917 à La Chaux-de-Fonds.
Frénésie	1960, juriste noire employée de Tita Zins.
Elisabeth II	reine d'Angleterre, née en 1926, couronnée reine le 2 juin 1953
Elisabeth vom Pokk	née en 1900, femme de Paul vom Pokk.
Joseph Riclette	né vers 1940, homme d'affaires, beau-fils de Paul vom Pokk.
La Prétendante	origine japonaise, née vers 1950, directrice et fondatrice de l'Ordre des Prétendants.
Lucienne Rothenbühler	née à Ornans en 1945.
Marie vom Pokk	1937, fille d'Elisabeth et de Paul vom Pokk, épouse de Joseph Riclette.
Max-Paul vom Pokk	né en 1924, fils d'Elisabeth et de Paul vom Pokk.
Paul vom Pokk jr	né en 1944, fils de Virginie et de Max-Paul vom Pokk.
Siddhârta Schweizer	dit Sid, né à Berne en 1982, apiculteur.
Tchaka	née en Tchétchénie en 1968.
Tita W. Zins	neveu de Tita Zins, héritier d'une partie de sa fortune, se déclare membre de Greenwar.
Tita Zins	1917, né aux Etats-Unis. Après une carrière politique chez les Républicains, il prend la direction de Big E.
Yoji Tsutsui	fils de Shizuko, né d'une fécondation in vitro.



## 3 (Re-)ordering Complex Stories

### 3.1 A Narrative Sudoku

Studying the effects of narrative ordering on the final perception of the story, from suspense to affective impacts, is not an entirely new idea. The *Structural-Affect Theory of Stories* postulated that the order in which events were told could be used to manipulate the readers' emotional responses [Brewer and Lichtenstein(1982)], and which in turn was recently showed to be an efficient way to tune simple computer-generated stories [Albuquerque et al.(2011)]. These results were also expanded and related to cultural specificities, qualitative reader perception [Brewer(1984), Brewer and Ohtsuka(1988)] and generalized to various literary genres [Hoeken and van Vliet(2000)]. Besides, those concepts have been used for decades by scenarists and screenwriters, as brilliantly summarized in Alfred Hitchcock's famous quote about suspense and surprise<sup>1</sup>:

*Let's suppose that there is a bomb underneath this table between us. Nothing happens, and then all of a sudden, "Boom!" There is an explosion. The public is surprised. [...] Now, let us take a suspense situation. The bomb is underneath the table and the public knows it, probably because they have seen the anarchist place it there. The public is aware the bomb is going to explode at one o'clock and there is a clock in the decor. The public can see that it is a quarter to one. In these conditions, the same innocuous conversation becomes fascinating because the public is participating in the scene. The audience is longing to warn the characters on the screen: "You shouldn't be talking about such trivial matters. There is a bomb beneath you and it is about to explode!"*

---

<sup>1</sup>American Film Institute | *Alfred Hitchcock On Mastering Cinematic Tension* - [http://youtu.be/DPFsuc\\_M\\_3E](http://youtu.be/DPFsuc_M_3E)

### Chapter 3. (Re-)ordering Complex Stories

In the case of long and complex stories, such as a full saga, one can expect the global ordering to be significantly driven by editorial, historical or technological constraints, and not to be necessarily unique. To test this hypothesis, we set out to look for other possible sequencings and their narrative impacts in *La simulation humaine*. In order to achieve at least one viable reordering, the first task was to label which parts were relevant to the story of each of the main characters, and to determine in what place and time in the saga they belonged. This preliminary step could easily be achieved using a combination of simple text analysis tools and confronting their outputs to the author's knowledge for validation.

975	- Mettez-nous un disque de <b>Wolfgang</b> .	981	- Mettez-nous un disque de <b>Mozart</b> .
1000	après l'autre. Puis dans l'ordre inverse. Délier les	1006	après l'autre. Puis <b> dans l'ordre inverse</b> . Délier les
1009	grincheux. Elle imagine ce <b>premier Wolfgang</b> terminant sa	1015	grincheux. Elle imagine ce <b>Mozart</b> terminant sa prestation
1010	prestation par une révérence tandis que de belles princesses	1016	par une révérence tandis que de belles princesses viennent
1011	viennent lui caresser la joue.	1017	lui caresser la joue.
1066	Wolfgang Steinamhirsch naît à la Chaux-de-Fonds, <b> le 1er mars</b>	1072	Wolfgang Steinamhirsch naît à la Chaux-de-Fonds <b> en Suisse,</b>
1067	1917. Chaque année ce jour-là, les habitants commémorent	1073	<b> le 1er mars 1917</b> . Chaque année ce jour-là, les habitants
1068	leur révolution contre le roi de Prusse. Les quelques	1074	commémorent <b> leur</b> révolution contre le roi de Prusse. Les
1069	aristocrates qui ont conservé leur particule ferment les	1075	quelques aristocrates qui ont conservé leur particule
1070	volets de leur maison en signe de deuil et ne sortent pas.	1076	ferment les volets de leur maison en signe de deuil et ne
1071	C'est pourquoi le médecin, un certain de Rougemont, laisse	1077	sortent pas. C'est pourquoi le médecin, un certain de
1072	ses parturientes se débrouiller seules avec la sage-femme.	1078	Rougemont, laisse ses parturientes se débrouiller seules
1073	On n'accouche pas un 1er mars, qu'on se le dise.	1079	avec la sage-femme. On n'accouche pas un 1er mars, qu'on se
1186	A peu de temps de là, il essaie de faire de Heidi sa	1193	A peu de temps de là, il essaie de faire de Heidi sa
1202	du grand professeur atomiste, <b> M. Scherrer</b> . Celui-ci demande	1209	du grand professeur atomiste, <b> Paul Scherrer</b> . Celui-ci
1203	à Wolfgang s'il croit que la physique va résoudre tous les	1210	demande à Wolfgang s'il croit que la physique va résoudre
1204	problèmes de l'humanité. Le jeune homme de seize ans répond	1211	tous les problèmes de l'humanité. Le jeune homme de seize
1205	crânement : la physique, oui, alliée à l'intelligence. <b> M.</b>	1212	ans répond crânement : la physique, oui, alliée à
1206	Scherrer lui donne une bonne tape sur l'épaule : Enoch, ce	1213	l'intelligence. <b> Le professeur</b> Scherrer lui donne une bonne
1207	garçon me plaît.	1214	tape sur l'épaule : Enoch, ce garçon me plaît.
1211	lui, il la consacrerait à la physique théorique. Puis se	1218	lui, il la consacrerait à la physique théorique. Puis <b> se</b>
1216	Admis à l'école polytechnique de Zurich, <b> il</b> quitte sa ville	1223	Admis à l'école polytechnique de Zurich, <b> Wolfgang</b> quitte sa
1217	à regret. La Chaux-de-Fonds n'est pas n'importe quelle	1224	ville à regret. La Chaux-de-Fonds n'est pas n'importe quelle
1266	Puis elle se plaint du temps, disant que La Chaux-de-Fonds	1273	Puis <b> elle</b> se plaint du temps, disant que La Chaux-de-Fonds
1276	premier plan, montagnes enneigées dans le lointain. Puis ils	1283	premier plan, montagnes enneigées dans le lointain. Puis
1277	parlent de la nouvelle passion de Wolfgang, la physique <b> de</b>	1284	ils parlent de la nouvelle passion de Wolfgang, la physique
1278	l'infiniment petit. Sa mère fait-elle semblant de comprendre	1285	<b> de</b> l'infiniment petit. Sa mère fait-elle semblant de
1279	ce qu'il lui raconte ? Ou bien saisit-elle vraiment la	1286	comprendre ce qu'il lui raconte ? Ou bien saisit-elle
1280	différence entre les électrons chargés négativement et les	1287	vraiment la différence entre les électrons chargés
1281	neutrons ?	1288	négativement et les neutrons ?
1289	alignées. Un jour Wolfgang tombe sur une équation de Planck,	1296	alignées. Un jour Wolfgang tombe sur une équation de Planck,
1290	se rend compte au dernier moment que <b> M. Scherrer</b> lui	1297	se rend compte au dernier moment que <b> le professeur</b> lui
1291	tend un piège. Cette formule a été réfutée. Tout tremblant,	1298	tend un piège. Cette formule a été réfutée. Tout tremblant,
1293	vérité et le triomphe de la nouvelle. <b> M.</b> Scherrer salue sa	1300	vérité et le triomphe de la nouvelle. <b> Le professeur</b> Scherrer
1294	clairvoyance :	1301	salue sa clairvoyance :
1302	connaissances en mathématiques. Puis en chimie et en	1309	connaissances en mathématiques. Puis <b> en</b> chimie et en
1309	compte au classement des transuraniens. Pas mécontent, <b> M.</b>	1316	compte au classement des transuraniens. Pas mécontent, <b> le</b>
1310	Scherrer l'invite chez lui avec ses assistants, un privilège	1317	<b> professeur</b> l'invite chez lui avec ses assistants, un
1311	rare. Mme Scherrer, la belle Ida, lui glisse quelques mots	1318	privilège rare. Mme Scherrer, la belle Ida, lui glisse
1312	en français :	1319	quelques mots en français :
1340	A la fin de l'année 1938, <b> M.</b> Scherrer annonce à son poulain	1347	A la fin de l'année 1938, <b> le professeur</b> Scherrer annonce à
1341	que la suite de sa carrière se passera en Californie. Il	1348	son poulain que la suite de sa carrière se passera en
1342	rejoindra le professeur Robert Oppenheimer à Berkeley. Mais	1349	Californie. Il rejoindra le professeur Robert Oppenheimer à
1343	d'abord il doit se rendre à Stockholm pour une mission	1350	Berkeley. Mais d'abord il doit se rendre à Stockholm pour
1344	spéciale.	1351	une mission spéciale.
1377		1384	
		1385	
1381	Meitner chez un ancien assistant de <b> M. Scherrer</b> . Une fête	1389	Meitner chez un ancien assistant <b> du professeur</b> Scherrer. Une
1382	pour quelques scientifiques exilés en Suède. Cette petite	1390	fête pour quelques scientifiques exilés en Suède. Cette
1383	femme aux cheveux blancs lui sourit, levant son verre à la	1391	petite femme aux cheveux blancs lui sourit, levant son verre
1384	santé de la science. Elle arrive d'Allemagne. Là-bas elle	1392	à la santé de la science. Elle arrive d'Allemagne. Là-bas
1385	est une chercheuse célèbre, admirée. Elle en garde un	1393	elle est une chercheuse célèbre, admirée. Elle en garde un
1390	travaux de <b> M. Scherrer</b> sur le bombardement des atomes par	1398	travaux de Scherrer sur le bombardement des atomes par <b> des</b>
1391	<b> des</b> neutrons. Puis elle propose de remettre à demain la	1399	neutrons. Puis <b> elle</b> propose de remettre à demain la
1392	suite de leur discussion.	1400	suite de leur discussion.
1458	complicité de quelques collègues courageux, dont <b> M.</b>	1466	complicité de quelques collègues courageux, dont Scherrer.
1459	Scherrer.	1488	
1481	Steinamhirsch répète l'offre de <b> M. Scherrer</b> . <b> A</b> Zurich, le	1489	Steinamhirsch répète l'offre <b> du professeur</b> Scherrer. <b> A</b>
1482	bombardement des atomes attend toujours une explication. Au		Zurich, le bombardement des atomes attend toujours une

**Figure 3.1** – Highlighted differences (in red) made between December 2013 (Left) and April 2014 (Right). Horizontal lines are shown instead of identical portions of text. Notice how almost all changes are related to simple local disambiguations.

Although the original books presented a definite stylistic homogeneity due to having been

written by the same author, the text was not immediately suitable for a full reorganization. Amongst other things, a disambiguation work was needed to make the beginning of each chapter less dependent on the preceding context, leading to thousands of word-level details corrections, as shown in fig. 3.1. Daniel de Roulet explains the process in his own words:

*« Je crois pouvoir dire qu'entre les versions papiers et les briques que je préparais, aucun chapitre n'est sorti indemne du processus. Car en plus des changements de personnes dans la conjugaison, il fallait s'assurer que dans chaque chapitre la narration puisse être comprise du premier coup. [...] Des formulations vagues en relation avec les chapitres précédents ont dû devenir très précises. On ne pouvait plus écrire en début de chapitre "A quelque temps de là Max s'est installé non loin de là..." Il fallait dire : "En juin 1988, Max s'est installé à Manhattan..." »*

Unsurprisingly, the resulting corpus (the one we used throughout the rest of this research), is very similar in structure and contents to the material of the original paper books. Yet, due to those stylistic differences and to prevent any confusion between the previously published texts and their reworked pendants, the reorderable digital volumes were sorted according to the narrative chronology and labeled with "SH.*n*", with *n* the sequence number (detailed in table 3.1).

Label	Original Paper Edition Title	Original Publication Year
SH.1	Kamikaze Mozart	2007
SH.2	Fusions	2012
SH.3	Gris-bleu	1999
SH.4	Virtuellement vôtre !	1993
SH.5	La ligne bleue	1995
SH.6	Bleu siècle	1996
SH.7	Davos Terminus	2001
SH.8	L'homme qui tombe	2005
SH.9	Le silence des abeilles	2009
SH.10	Le démantèlement du cœur	2014

**Table 3.1** – Nomenclature of the books constituting the corpus *La simulation humaine*

A first reordering attempt spontaneously followed from the structure of the novels. Considering the pinpointed interleaving stories of recurring characters, it was easy to extract each narrative path on their own to reconstitute in a chronological order the individual story of each of the main characters, in the way of a detailed biography. Despite its simplicity, this idea found some relevance in providing a mean for the reader who didn't read the full saga to learn more about one or several people whose life crossed (and possibly ended in) the tenth book. Such an organization would also have made an interesting base for a digital book, as it allows for

### Chapter 3. (Re-)ordering Complex Stories

---

displaying linear stories that necessarily cross each other at one or several points. The readers could at that time be given the possibility to go on reading the biography they started with, or to turn to the point of view of another character from this chronological intersection on.

Once put on the table, this first attempt produced a set of 17 stories that, although well constructed and very informative, presented a recurring issue. According to the author, they were all plagued by an overall feeling of boredom, or at least not thrilling enough to pull the reader further than a couple of chapters inside the story. He commented:

*« [...] leur intérêt était faible voire inexistant. Dans mes romans, c'est la confrontation entre les points de vue portés par différents personnages qui soutient l'intérêt du lecteur. Remettre à plat les biographies des personnages comme autant de romans d'éducation, ou suite d'épisodes, détruit toute tension dramatique. »*

We indeed very quickly realized that this way of retelling wouldn't be successful in a general case, as it had the systematic side effect of losing part of the narrative tension that was driven by the shifts of points of view described earlier and the imperatives of revealing some details in a certain order to make the stories appealing. More than a failure, this result brought the author to formulate another simple question: *«Is it possible at all for this very corpus to be presented in its integrity, in another way than it was originally published?»*. With this idea in mind we realized that the challenge of finding alternative interesting reading paths would necessarily closely relate to a constrained puzzle optimization problem: in a very similar way one fills in the numbers in a sudoku grid, the puzzle would here be to satisfyingly order the chapters, under the constraint of using all the pieces exactly once.

### 3.2 Paradigms for "Reordered" Digital Sagas

Family sagas traditionally differ from popular romance by being structured as long-time processes, focusing on the maintenance of the family state instead of driving towards narrative closure, which is reflected in a widely adopted marketing strategy often aiming to help readers identify to the writer [Bridgwood and Radford(1986)]. In the realm of French literature, the saga is nowadays often seen as very conservative if not old-fashioned, even if it used to be, not so long ago, a particularly appreciated genre [Leblond(2010)]. Yet, its lengthy realization process still poses a challenge, not only of endurance, but also about composition strategy, as shown by many well-documented classical authors' cases (section 2.5). With a text widening in size and complexity, the author is likely to get overwhelmed by the complexity of the artwork and thus welcome tools and rules helping to shape the growing structure. In traditional sagas,



### 3.2. Paradigms for "Reordered" Digital Sagas

---

from the seventeenth century to Balzac and Zola, it seems this composition difficulty was less of a concern and didn't prevent authors to add further episodes to their characters' adventures [Leblond(2010)]. The genre merely became considered outdated or impossibly ambitious after the middle of the 20th century, and stays prosperous today mainly in the niche markets of youth and fantastic literature, from *Harry Potter* to *A Song of Ice and Fire*. In many cases, the success of these works also was tied to their adaptation as series of feature films or long TV series, which raises interesting questions about the perceived degree of complexity and the relation to new medias that are needed to potentially conquer a large public.

At the other end of the digital reading and publication chain, some paradigms may need to be refined in order to make the reading experience best adapted to expectations of a contemporary public. Nowadays all mass entertainment sources follow the recent commercial trend that promoted instant reward and constant (over-)stimulation as desirable qualities and de facto standards [Rushkoff(2014)]. Namely:

- Instantaneous, unlimited and global access to any kind of contents, supported by broadband Internet and more recently mobile devices. This trend caused the disappearance of physical media in the first place, and right thereafter the transition from downloading to streaming, hence removing the need for local physical storage. The tremendous success of unlimited on-demand contents providers for music and movies is a good illustration of this phenomenon (e.g. Spotify<sup>2</sup> and Netflix<sup>3</sup>).
- New forms of medias that were built upon those principles and spawned virtual social networks of users became increasingly popular (e.g. Twitter<sup>4</sup> and YouTube<sup>5</sup>). This trend didn't stop there yet, with more recent forms of medias that pushed further on the concepts of present time and short lived information. Those recently reached an incredibly fast penetration within younger demographics (e.g. Snapchat<sup>6</sup> and Periscope<sup>7</sup>), whereas others updated their models to fit this logic (e.g. Instagram's<sup>8</sup> Stories mode[Nieva(2016)]).

As far as contents are concerned, it is easy to find good examples of these effects in recent long TV series, whose seasons come in one after the other, each similar to the previous one but still keeping its own autonomy. In many cases, we can see the logic of immediate consumption in the lack of long narrative arcs or logical dependencies, or if they exist at all, they may not

---

<sup>2</sup><http://www.spotify.com>

<sup>3</sup><http://www.netflix.com>

<sup>4</sup><http://www.twitter.com>

<sup>5</sup><http://www.youtube.com>

<sup>6</sup><http://www.snapchat.com>

<sup>7</sup><http://www.periscope.tv>

<sup>8</sup><http://www.instagram.com>

### Chapter 3. (Re-)ordering Complex Stories

---

be needed to enjoy a couple of random episodes (the more complex stories being kept for an elite spectator that will actually need to put some additional effort to get them).

In that context, one can easily see why reading long books becomes an uneasy activity. Actually, a contemporary saga written with the idea of complying to those expectations should probably take into account several factors:

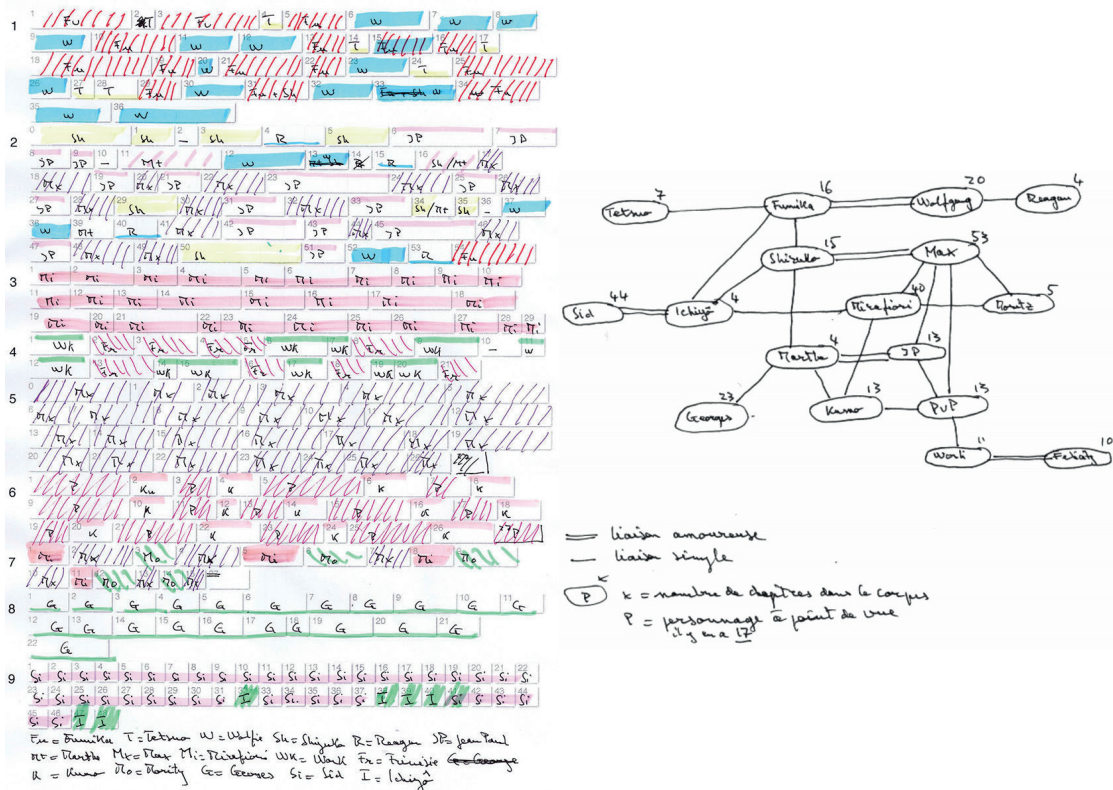
- With a narrative strategy in the form of an uninterrupted flow, the question of defining a start or an end becomes less essential. On the contrary, the capacity of getting immediate immersion and feel quickly rewarded are important success factors.
- The narration must be achieved while keeping in mind a continuous gripping effect, as opposed to the more traditional writing where some chapters are expected to tie the reader to the book until the intended resolution.
- This goes together with proposing both short term and long term denouements, whose resolutions are hardly foreseeable in order to incite the reader to cling on as far as possible.

### 3.3 Structural Understanding

In order to tackle this question, it was first necessary to get a better conceptual understanding of the narrative strategies already present in the existing saga. A family saga is indeed a story in itself, but characterized by the presence of numerous autonomous and organized sub-stories, each having their own plots and characteristics, and more or less strongly interleaved with each others. This is imposed by the will of the author, but also by the traditional needs of publishers because of the necessity for each physical book to be at the same time a reasonably autonomous item, just revealing enough of the larger plot lines for an incentive to buy the following volume. For this reason, a saga considered as a whole is likely to exhibit a collection of different organizational strategies that fulfill those purposes. We call this the *narrative structure*.

Our primary source of information to decode the narrative structure of *La simulation humaine* was gathered through a close collaboration with the author. With the help of documentation he used at the time he wrote the stories, and by completing this knowledge by hand-drawn summarizing schemas, he laid flat and provided several hints about his creative process and original intents. Examples of those documents are shown in fig. 3.2.

### 3.3. Structural Understanding



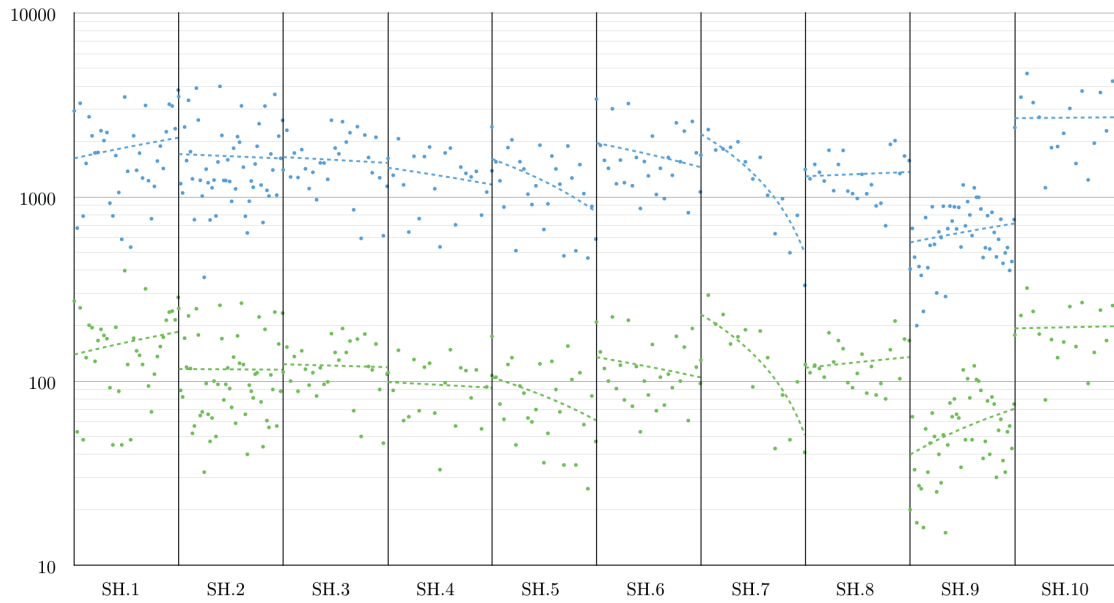
(a) Narrative focus (Each square is a chapter) (b) Social links between the 17 main characters.

Figure 3.2 – Author’s representations of the narrative structure components in *La simulation humaine*.

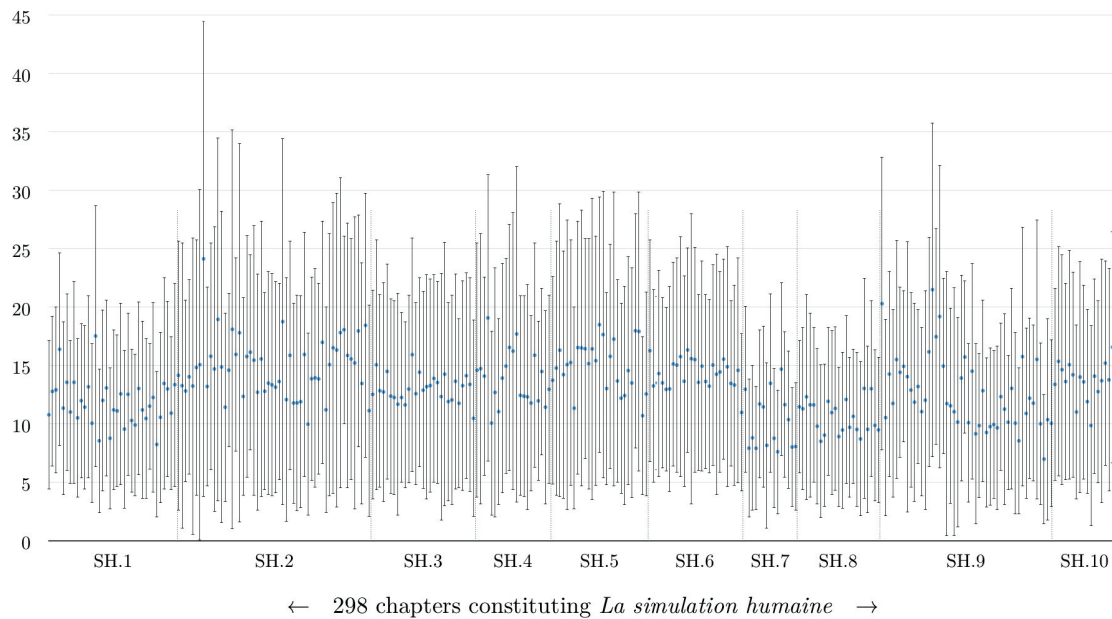
Building on top of that, we developed a set of simple multipurpose tools that, as a whole, allowed to point at unsuspected structural features and implicit organization strategies. The author’s reflection on those permitted him to see his work in a new light and to refine the necessary conditions for a felicitous full recomposition (see appendix A.3.2 for the author’s report in April 2014).

A first attempt at quantifying the corpus aimed to look at raw figures that could be drawn from simple statistics on the text of the books. Simple measures like number of words and sentences per chapter were computed and plotted to get an better idea of the corpus homogeneity. The raw comparison of chapters lengths in terms of numbers of words and sentences showed a substantial diversity (fig. 3.3), the shortest ones starting at 200 words (17 sentences), with the longest one capped at 4688 words (320 sentences). Leaving out the most extreme cases, 80% of them are linearly distributed in the 500 to 2500 words range, with an average of 1475 words (115 sentences).

### Chapter 3. (Re-)ordering Complex Stories



**Figure 3.3** – Number of words (top/blue) and sentences (bottom/green) in all chapters of the ten books of *La simulation humaine*.



**Figure 3.4** – Average sentences lengths in *La simulation humaine*.

A closer look at each book shows that, except for *SH.7* where the chapters tend to shorten with time, and *SH.9*, whose chapters are consistently shorter than in any other books', those features exhibit no obvious patterns that would potentially hinder the reorganization process, as it would be the case for instance if all books concealed a similarly shaped curve that would need to be respected for stylistic reasons.

Furthermore, this simple visualization hints at a stylistic feature that some scholars argue to be a stable marker of the author’s style, namely the average and variance of sentences lengths. A closer look at those (fig. 3.4) show that the number of words per sentence is a rather stable feature throughout most chapters and books, which is compliant with the averages pointed out by other ongoing researches [Schöch(2016)].

Another set of tools aimed at representing the alternation of narrative focalizations inside each of the original novels. We tagged all chapters with the main character they are focused on, and considered the chain of changes of point of view as a simple marker of the macrostructural sequencing. More formally, we modeled each chapter as stable states in the story during which some parameters are fixed, and changes of chapters as a potential state transition. This representation, as shown in fig. 3.5 (a), has the advantage of providing a precise and uncluttered visual feedback in the form of a state transitions diagram [Booth(1967)], even for complex stories.

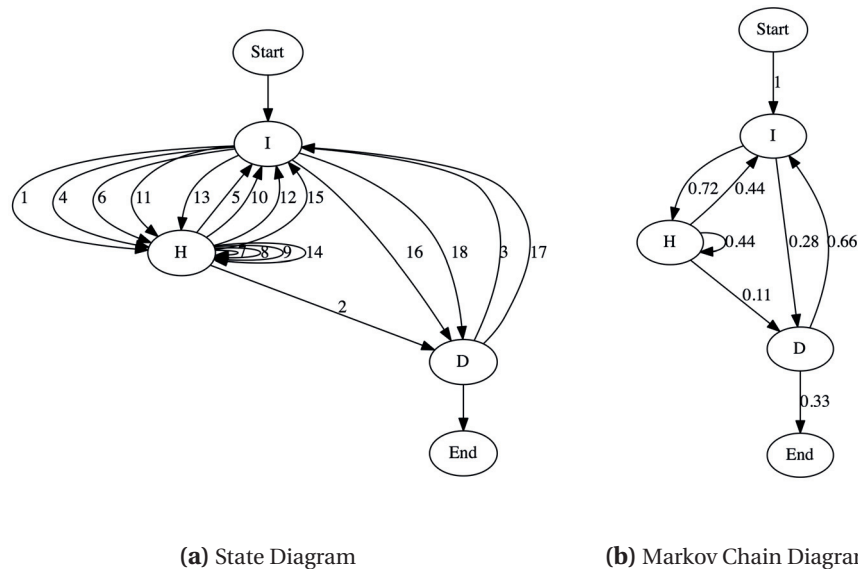


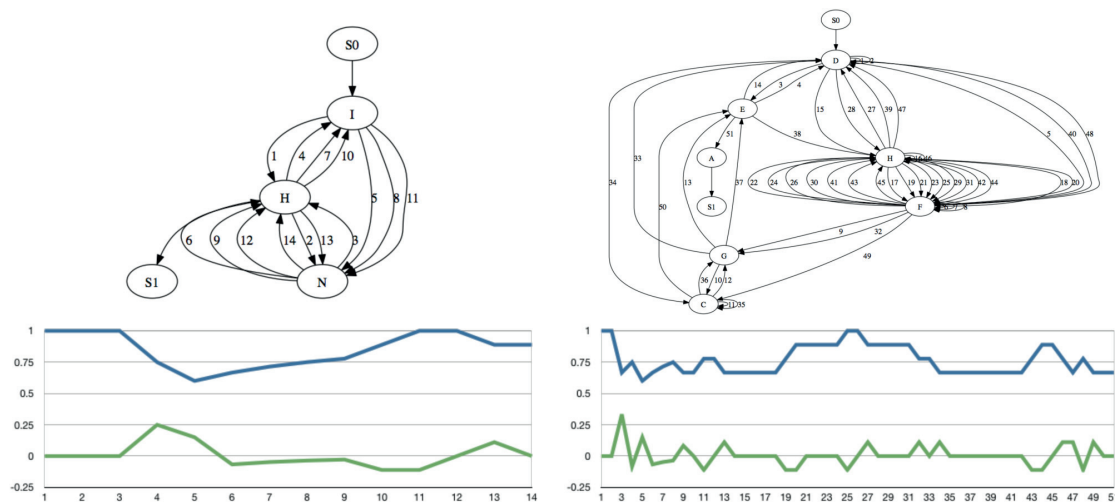
Figure 3.5 – Narrative focalizations transitions in *SH10*.

Aside from the visual feedback, it is also possible to consider a statistical understanding for each graph, by computing the probability of a transition towards another state given the prevalence of the observed alternances. To illustrate this interpretation, let us imagine a distracted reader that consistently forgets the focalization in the previous chapter (say,  $n - 1$ ) while he is reading the following one ( $n$ ). In that case, the focalization of the next chapter ( $n + 1$ ) can only be predicted with the knowledge of the current one’s ( $n$ ), and by hypothesizing the number of characters is known and finite, the sequence of predictions for the whole story

## Chapter 3. (Re-)ordering Complex Stories

can be represented as a first order Markov chain, whose states are the focalizations for each chapter [Kemeny(1976)]. This new representation is illustrated in fig. 3.5 (b).

Of course, the Markov chain representation can be extended to match more realistic models. For instance, considering readers that remember the focalizations of the  $k$  preceding chapters and will use this knowledge to predict the next one, the same interpretation would follow with a Markov chain of order  $k$ . From there on, it is possible to model to which extend the choice of one transition over another follows a likely or expected structural construct, or on the contrary if a structural surprise effect is to be expected. Figure 3.6 shows one possible visualization for the structural foreseeability (blue line). Notice how the second story, more complex and about twice as long, exhibits a relaxing point in the middle.



**Figure 3.6** – Joined representation comparing the narrative structures from *SH7* (left) and *SH2* (right): states diagram (up), structural foreseeability and its first order derivative (bottom).

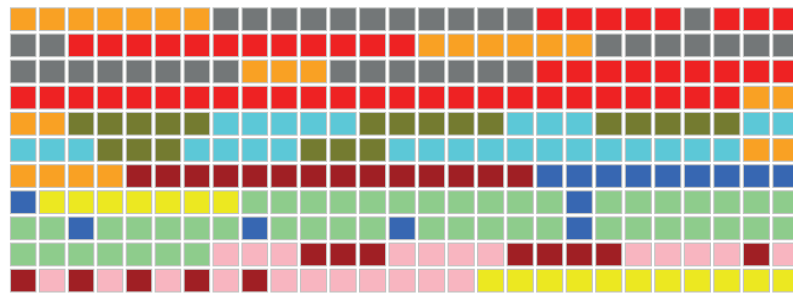
### 3.4 *La simulation humaine*

#### 3.4.1 A First Example of Reordered Saga

With the help of the aforementioned visualization tools and his irreplaceable knowledge of the corpus, Daniel de Roulet managed to produce a full reorganization of the 298 original chapters constituting *La simulation humaine*, validating the existence of at least one possible rearranging, using exactly once every chapter of the original saga. Figures 3.7 and 3.8 illustrate the principle of this reorganization, which reminds in some way Julio Cortázar's *Hopscotch* [Cortázar(2006)], whose 155 chapters can be read following an alternative sequence to the numeric order.



**Figure 3.7** – Original sequencing of *La simulation humaine* (10 books in a row, each colored square represents a chapter from of the original books)



**Figure 3.8** – New sequencing of *La simulation humaine* (The colored squares are the same as in fig. 3.7)

Additionally, according to the initial constraints we set, the author defined nine new self-standing novels inside the reorganized saga (as described in Document 2), each of them up to be read on their own or within the reorganized saga as a whole (fig. 3.9). As we will see in section 3.4.2, this feature allows to provide a greater variety of entry points to the saga, varying the minimal number of chapters constituting one coherent story and thus aiming to be more appealing to new readers by permitting a progressive introduction to the corpus without asking for upfront commitment to read it whole. For this reason, the three shortest bits (actually only one chapter in length) were subtitled as "*un avant-goût en un chapitre*" ("*A foretaste in one chapter*"), leaving no doubt about their short nature and thus expected to act as ten minutes teasers for the saga. They are followed by three short novels (4 - 10 chapters in length, 1 - 2 hours reading time), and three the length of a typical paper book (30 - 38 chapters, or 10 hours reading time).

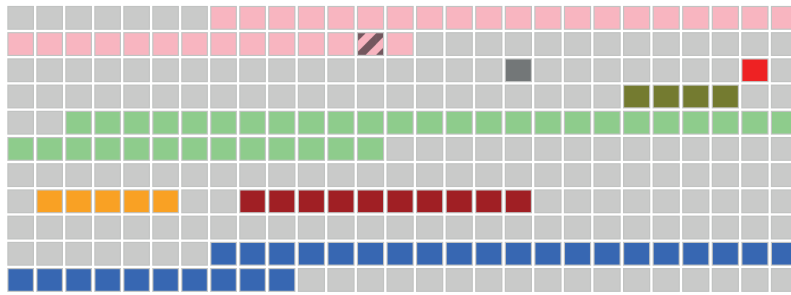


Figure 3.9 – New sequencing of *La simulation humaine* (9 new self-standing novels inside the reordered saga)

### 3.4.2 A Multimodal Reading Tool

With the exception of the very specific field of education [González et al.(2013)] where multimedia contents indeed show better learning results, most scholars agree that reducing sources of distractions is a must to leverage concentration and improve cognitive abilities [Lustig et al.(2006)]. For this reason, we renounced to introduce external items to the raw text, like additional documents or hyperlink notes, in order to guarantee the best possible immersion experience. The final interface should thus be uncluttered, display only the navigation items that are relevant at the time being, no additional documents or links to other parts of the app, nor advertisement.

The tool that was released to the public followed these ideas. It was made accessible both as a mobile-friendly website<sup>9</sup> and native mobile applications for both iOS<sup>10</sup> and Android systems<sup>11</sup>. Figure 3.10 shows the home screen of the web interface, that displays the ten possible entry points (an immediate choice between the nine new novels, or reading the full saga in the sequence order).

In order to make this page more descriptive and informative, each story is presented with a title, a short summary and one related image acting as a cover. This screen also features a short note describing the project, and a meter displaying the total count of chapters read by the user, aiming to reintroduce some sense of dimensionality that has been lost in the conversion to digital and the removal of the paper pagination.

---

<sup>9</sup><http://www.simulationhumaine.com>

<sup>10</sup><https://itunes.apple.com/fr/app/la-simulation-humaine/id833678593>

<sup>11</sup><https://play.google.com/store/apps/details?id=com.simulationhumaine.app>



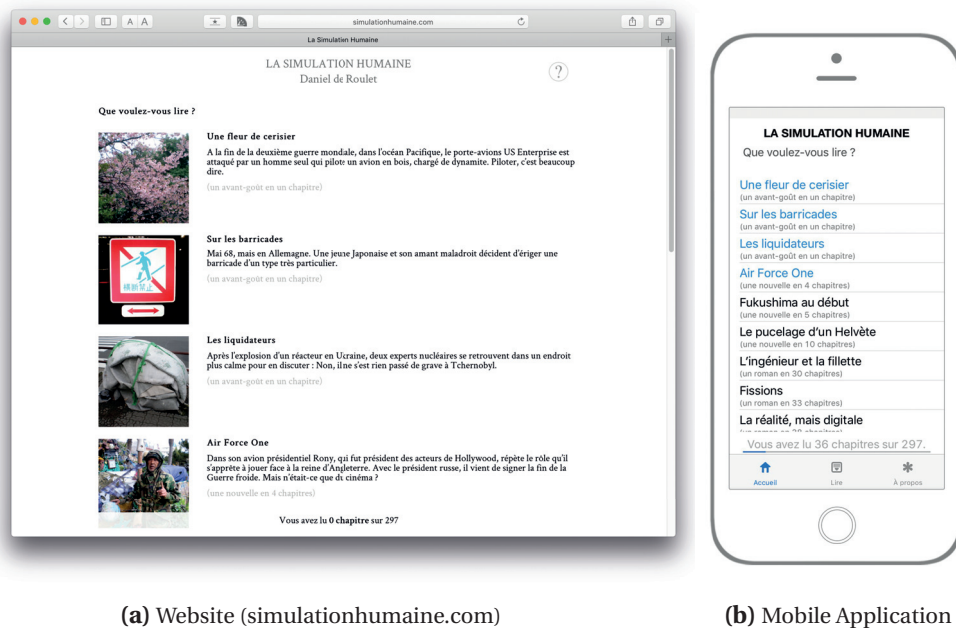


Figure 3.10 – Home screen of the digital saga *La simulation humaine*.

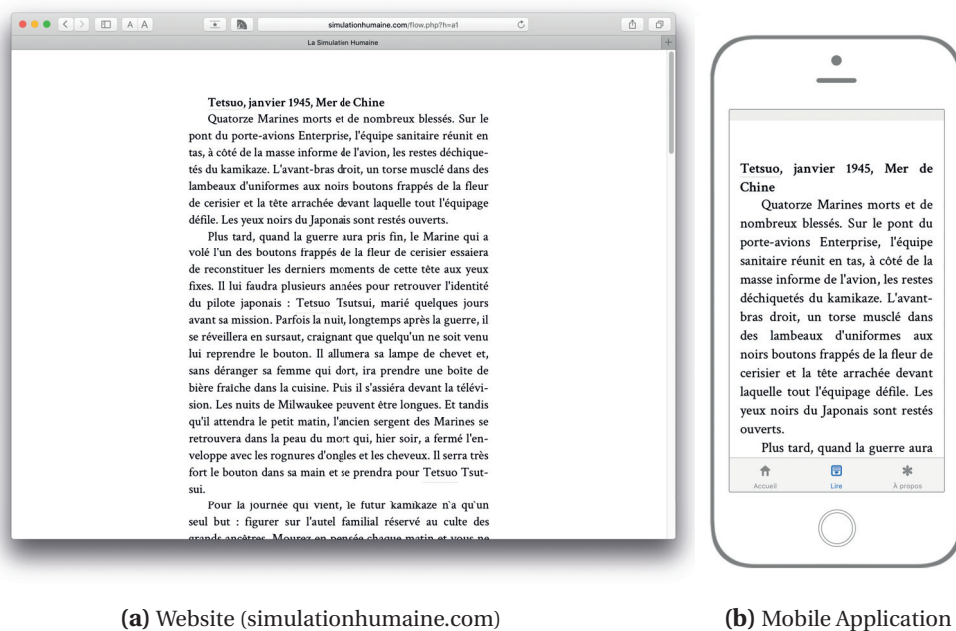
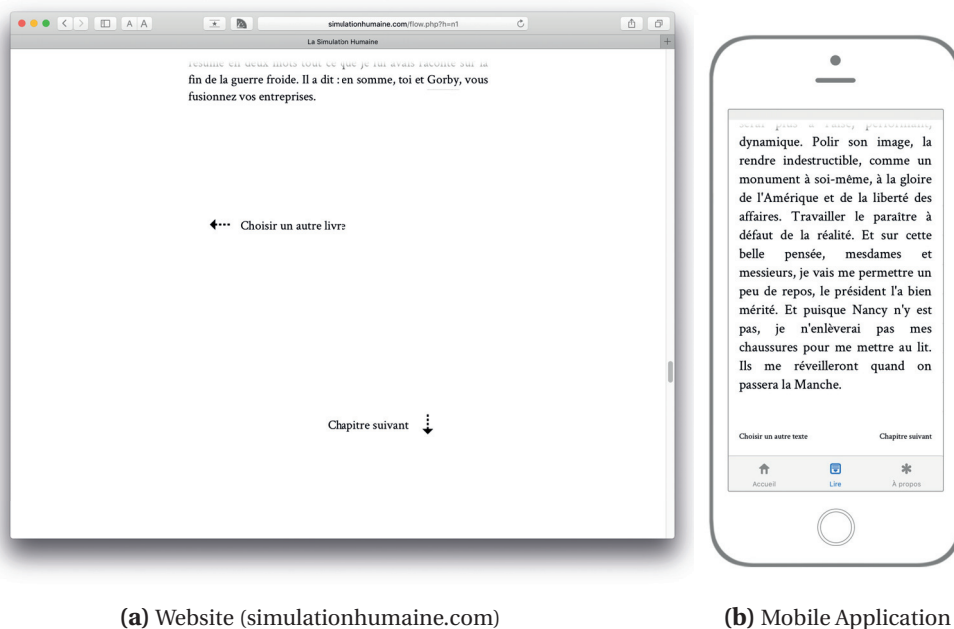


Figure 3.11 – Reading interface of the digital saga *La simulation humaine*.

Figure 3.11 shows more precisely the reading interface, directly accessed by one click or touch from the home screen. Chapters are unnumbered but start with a title line showing the name of its protagonist and the date and location frames where the action takes place. The text is

### Chapter 3. (Re-)ordering Complex Stories

presented as a single page that unrolls as the user scrolls further down, following the design principle known as *infinite scrolling*. This technique has been shown to be beneficial in the context of documents consultation, as it allows to walk through contents without having to transition between two interfaces [Whitelaw(2012)]. To reinforce the immersion, navigation interface has been kept to the strict minimum and mandatory elements are displayed only when necessary, for instance to offer the possibility of switching to another story at the end of a chapter (as depicted in fig. 3.12).



**Figure 3.12** – Textual navigation elements, located between two chapters exclusively.

Due to the non-linear nature of this reading, we introduced a tiny glossary that provides clues about elements that may not have been introduced to the reader before (shown in fig. 3.13). The names of 65 most important and recurring characters, places and companies were identified and are shown with a very slight underline markup. When clicked on, a short notice written by the author is shown, aiming to introduce the entity without giving out any important information about the story.



Figure 3.13 – Glossary function.

### 3.4.3 Applying Web Analytics to Digital Books

Usage information play a core role in understanding consumer reactions to any product, and are thus an important component both in the development and marketing stages of most commonly used products. Where in most traditional industries those are typically difficult and costly figures to get, the venue of Internet turned the tables by making precise users tracking a reality. For instance, the consultation paths of electronic commerce sites in the form of a directed graph of pages pulled from the server's log files has been successfully used as a powerful and readily available datasource for marketing purposes [Büchner and Mulvenna(1998)].

Even with reading being an activity mostly focused on the participating user (as the device plays a very passive role), many perceptive analytics can be pulled from this simple process [Alter(2012)]. To get insights about how the new books are read on all platforms, our system was designed to track precise reading usage statistics. Formally, the collected data are very similar to state-of-the-art website analytics. In response of the HTTP request needed to load the page, we logged on the server's side the precise date, time and type of device used, as well as the chapter and story identifiers (since some chapters can be accessed either through one of the sub-stories, or within the flow of the full saga). Additionally, each user is uniquely identified using a browser session and a cookie on the website, or a device identifier on the mobile app. An excerpt of those records is shown in fig. 3.14.

### Chapter 3. (Re-)ordering Complex Stories

---

a4c0aac2198f651c	2-29	a2	2014-08-25 20:10:56	and:nuclear-a86
a3631jd11956rvheodqqgnc5l7	2-1	a3	2014-08-25 20:53:29	web
aff50ba983971357	2-29	a2	2014-08-25 21:34:19	and:SmartPadGo
aff50ba983971357	2-1	a3	2014-08-25 21:36:29	and:SmartPadGo
aff50ba983971357	1-6	r2	2014-08-25 21:37:50	and:SmartPadGo
aff50ba983971357	2-4	n1	2014-08-25 21:40:00	and:SmartPadGo
aff50ba983971357	5-1	i1	2014-08-25 21:40:16	and:SmartPadGo
aff50ba983971357	1-28	a1	2014-08-25 21:40:58	and:SmartPadGo
36ac46183dac362f	2-29	a2	2014-08-25 23:02:45	and:crane-evb
8ao6mmlg0pv3ugqtf29avel3v6	5-1	i1	2014-08-26 08:29:14	web
7c2lpochkhun2nd7srj663d417	1-28	a1	2014-08-26 11:28:03	web
3613c17bef8d89cc	2-4	n1	2014-08-26 13:37:06	and:nuclear-pub
caa21e49ceae6533	5-1	i1	2014-08-26 17:09:19	and:crane-a1002jhzx
46d1cbcdbb0b7365	4-1	r3	2014-08-26 17:20:33	and:crane-evb
1a3ac13b509b939c	5-1	i1	2014-08-26 19:29:51	and:nuclear-n01
a57e8b686a415af7	1-28	a1	2014-08-26 19:59:20	and:crane-a1002jhzx
F44E5FCE-BE44-46EE-8EF2-7B9C01C1	1-6	r2	2014-08-26 21:11:06	iPad Air (WiFi)
36ac46183dac362f	2-29	a2	2014-08-26 23:27:53	and:crane-evb
o2hdvjifdsu86pbia34ll6ie6	8-1	r1	2014-08-27 13:48:00	web
o2hdvjifdsu86pbia34ll6ie6	8-2	r1	2014-08-27 13:48:06	web
056B0A14-BE0B-448E-A8E5-9BEE5501	1-6	r2	2014-08-27 18:22:30	iPad 3 (WiFi)

**Figure 3.14** – Example of raw usage statistics: User identifier, Chapter number, Story identifier, Date and Device Type.

## 4 Methods for Digital Writing Tools

### 4.1 Distant Editing

Famous authors are usually quite evasive about what makes for a good novel and how to produce one. Mark Twain, Oscar Wilde and Somerset Maugham apparently in their own ways all came up with a single advice on writing fiction, in the form of three rules being "write, write, and write"<sup>1</sup>. One can find plenty of books on the art of writing novels though<sup>2</sup>, especially very recent ones relating to the fashionable digital and self-publishing. Yet, most authors that synthesized those pieces of advice focused on the very pragmatic aspects of inspiration gathering and the technicality of plot building [Hall(1989), Noble(2011), Bransford(2013)]. Few tackled the theory of editing per se, besides the recurring observation that the process with few exceptions involves a mix of structuration by global planning and content expansion by gathering details, with some authors being prolific in detailed insights about possible strategies for efficient and appealing results [Weiland(2011), Weiland(2013)].

In the realm of novel writing, one sometimes finds a theoretical categorization between authors that obey a thorough planification (*plotters*), compared to ones following their inspiration without knowing precisely beforehand where the story will go (*pantsers*). Most novelists use a mix of both, and no strategy seems to yield better results than the other, with the means that eventually lead to completion being highly individual [Ousby(2009), Parv(2014)]. Of course, there aren't a lot of choices available to the novelist besides structuring and writing, in some arbitrary (and mostly sequential) order. Let us just stop on those two aspects to better define

---

<sup>1</sup>Quote Investigator : There Are Three Rules for the Writing of a Novel - <http://quoteinvestigator.com/2013/05/06/three-rules/>

<sup>2</sup>As of December 2016, Amazon.com lists more than 22 000 books in the *Fiction Writing Reference* category, including best-sellers like *Writing Fiction For Dummies* [Ingermanson and Economy(2010)] - [https://www.amazon.com/s/ref=lp\\_12015\\_nr\\_n\\_3?fst=as%3Aoff&rh=n%3A283155%2Cn%3A%211000%2Cn%3A21%2Cn%3A5267707011%2Cn%3A12015%2Cn%3A12022](https://www.amazon.com/s/ref=lp_12015_nr_n_3?fst=as%3Aoff&rh=n%3A283155%2Cn%3A%211000%2Cn%3A21%2Cn%3A5267707011%2Cn%3A12015%2Cn%3A12022)

them:

- **Writing** is the phase dedicated to shaping words, sentences and paragraphs into chunks of text that will potentially eventually be part of the story. It can take place in countless forms, and we have for instance illustrated with our classical examples related to constrained writing. In its simplest form, writing can be a very straightforward action, with no specific focus on syntax, style or meaning at a given time. On the contrary, it might as well be decomposed in an information gathering and drafting phase that precedes the definitive writing, in which case a simultaneous approach to structuration can be envisioned in a first time.
- **Structuration** shapes the core of the story. It's the inspirational process where the big ideas come into place, along with the key interactions that make the story progress, and the order in which they will be presented to the reader. When one follows a *plotter* writing strategy, it is indeed more visible, as a distinct and explicit process usually preceding the writing phase. Considering a *pantser* way of working, it is more likely to be an implicit feature, built on the fly and on top of the last written pieces of plot.

From there on, the available options for composition are derived from analogue tools and manual processes that do not give the freedom to tackle the task with a different strategy.

As a matter of a fact, once a string of consequent pieces of text come together to form one work, be it by following a pre-established structure or while building one, one needs to follow a constraining method in order not to get lost in the details and introduce factual or narrative contradictions. For this reason, the vast majority of further possible edits past that point are likely to stay at a syntactic level. The reverse statement is also true, as structural editing on large texts is an arduous task, which is not made easier by conventional text processing software that impose a pagination flow logic.

However, in the context of innovative edition possibilities offered by digital tools that would differ in their nature from the current ones, we would like to introduce a different editing paradigm. As a reference to *distant reading* and considering the involved necessity to operate at a decent level of abstraction from the text, we named this process *distant editing*. The key concepts allowing to define it more precisely are summarized in fig. 4.1. As shown, *distant editing* focuses on an intermediary level of edition located between the structuration, that connects scattered plot elements in a coherent story through structural features, and pure writing, which operates at the syntactic and semantic levels by turning words and sentences into autonomous fragments (or chapters). In order to make the picture complete and more

didactic, we included typesetting at the lower end of the list, which is the counterpart process that operates between words and glyphs levels.

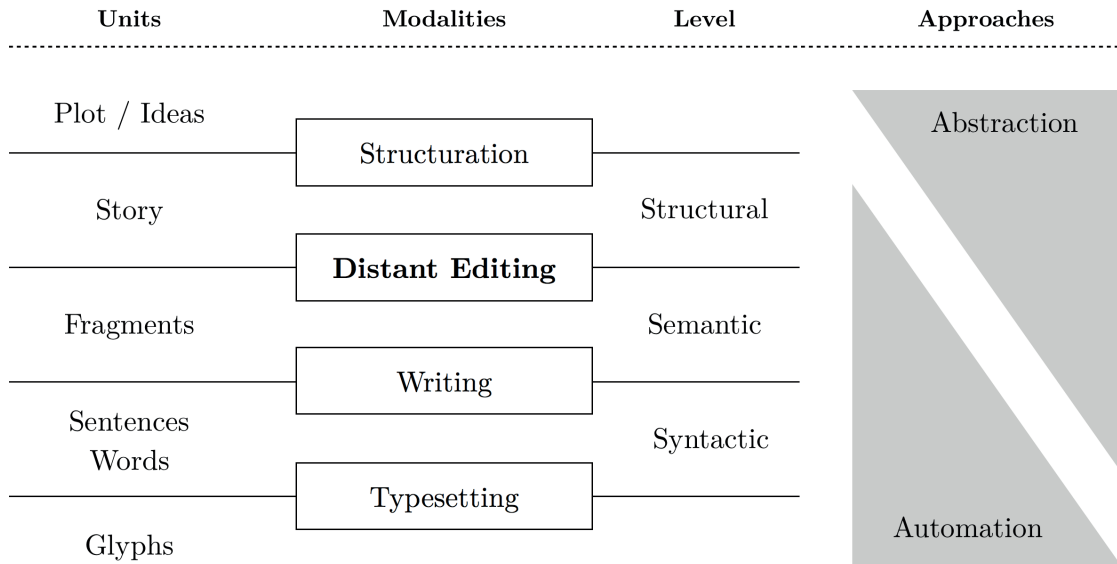


Figure 4.1 – Comparison between writing, structuration and *distant editing*

This organization also features two complementary supporting axes:

- **Abstraction** shows the distance from the text at which the author operates. It is minimal from the point of view of typesetting and at the time of writing, since those operations cannot be envisioned without an immediate and close relation to the text (e.g. to its shape for the former, and to the style and meaning for the latter). Conversely, it is maximal when structuring the story or drafting a plot, since one typically doesn't consider the actual text during those processes.
- **Automation** shows the part of the work that can be delegated to computer software with the current technological means. It is maximal in typesetting operations, in such a proportion that most users don't even realize this is being done. Writing nowadays includes a fair and growing share of automation, of course with spelling and grammar tools, but also statistics, auto-completion and stylistic suggestions. Structuration on the other hand is reputed a creative task, for which automation possibilities are almost inexistent.

On both axes, *distant editing* occupies a central position. The need for tools that provide a decent level of automation is a corollary to the balance one desires when working at the desired level of abstraction in order to consider the task such as the ordering of fragments.

### 4.2 Shaping Effective Tools for Reorganizing Complex Stories

As we have just seen, the venue of tools fostering *distant editing* could potentially change in many ways how any kind of novel is composed. In particular, one aspect that especially calls for this level of agility and that we know for having explored with *La simulation humaine* is the recomposition of different stories out of an existing corpus. The reorganization proof of concept that was achieved in the context of this research was indeed helped by several visualization tools (section 3.3), yet this process was still mostly done by hand and relied heavily on the author's implicit knowledge of his saga. Hence, the next step to get closer to real *distant editing* was naturally to devise more effective means that allow for an automated and thorough exploration of the structures of complex stories at the (re-)composition level. The aim of this section is to present and discuss methods that are efficient enough computationally wise to be applied on entire books, whilst producing feedbacks on global effects of the composition, as a whole.

Since this process is typically taking place at a level where a meaningful part of the story exists in the form of chunked text, we will assume this configuration can be exploited as a well-defined set of units whose final sequencing needs to be worked on. For existing pieces of work that thus need to be artificially segmented beforehand, chapters are the most obvious choice since they relate to a meaningful structure and will be easier guiding marks. Hence, for coherence and simplicity purposes, we will refer to the chunked units as "chapters" from now on, even if they may actually relate to a hidden granularity different from the actual "chapters" the author may settle on at publication time.

Starting with the idea of the visual representation of a set of chapters, any content-savvy method will necessarily need to identify the absence or presence of constitutive elements of the story. Unsurprisingly, among the possible features one may want to look at, characters, places, time markers and recurring thematics play a core role. Reasons for focusing on characters are:

- As shown by multiple classical examples and illustrated in our own recomposition process, the identification of subthreads following each character constitutes an inescapable task in the composition process, whose track keeping needs substantial amounts of efforts in the context of a dynamic and evolving project.
- Aside from that, the ability to reveal narrative strategies by plotting the interleaved presence of recurring characters and places has been shown to be key to interesting results in the context of *distant reading* comparison of novels [Moretti(2013)].



- Similarly, looking at possible more general applications, connecting distinct novels by the places and characters they share may offer interesting and innovative navigation and exploration means, as already suggested by some real world applications [Albanese(2011)].

In the next section, we will present a method for efficiently and reliably extracting featured characters and places names out of French texts.

### 4.3 Reliable Recognition of Characters and Places

Named Entity Recognition (*NER*) has become a reasonably well developed field, driven notably by the needs for machine translation and automatic extraction out of flows of news articles [Hirschberg and Manning(2015)]. However, only a small number of systems perform well on other languages than English [Azpeitia et al.(2014)], in particular in our case for French content. In addition, fiction is challenging since a large part of novel contents are not listed in reference databases commonly used by many *NER* systems, like DBPedia<sup>3</sup>, Wikidata<sup>4</sup> or YAGO<sup>5</sup>. One additional difficulty with non-English languages is the limited amount of readily available sets to train machine learning recognisers. Therefore, there is a strong relevance for solutions that can achieve good results without the need of training beforehand.

The solution presented in this section combines a set of different standalone recognition systems within a meta-recognition system. This approach, which originated from improved accuracy needs in biometrics, has proven to be effective in various other fields [Scheirer et al.(2011)], and among them, ones related to *NER* [Si et al.(2005)]. The rationale behind meta-recognition is to reach out for a set of independent classifiers, each looking at different features and possibly using different strategies. One would then expect that although none of the standalone method clearly outperforms the others, their combined judgement offers a robust solution to the tackled recognition problem, as their individual errors will tend to cancel out.

#### 4.3.1 Test Procedure

We built a test corpus made out of 35 classic works (5M tokens) in French, stemming from digitized books part of Project Gutenberg<sup>6</sup>. The works were handpicked in a way to get a mix of different genres and narration types. They include 3 sagas, 6 classical books that were

---

<sup>3</sup><http://wiki.dbpedia.org>

<sup>4</sup><http://www.wikidata.org>

<sup>5</sup><http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago>

<sup>6</sup>Project Gutenberg | Browse By Language: French - <http://www.gutenberg.org/browse/languages/fr>

translated from other languages and 25 French classical novels. For each piece of work, we produced a corresponding reference file by extracting the proper nouns occurring more than five times and manually labelling each of them with the category they fall in (character, place or other). The resulting database references about 3,700 categorized named entities, along with the work they come from and their number of occurrences.

Given our further needs, we considered names referring to pets, companies and groups of persons as regular characters whenever they serve the same narrative purpose (people mostly interacting with corporations or groups). Conversely, they were considered places in the case of names used to qualify where actions take place (as with corporations considered as locations), or ethnical groups used to name their originating place (as in "*le pays des Oreillons*"). This differs a lot from most knowledge databases, as the very bounds between characters and places might be fuzzy and context dependent. For instance, if the word "*France*" is part of the expression "*Roi de France*" we don't want it to be labelled as a place, but the whole entity as a *character* instead, which may vary from work to work.

For the actual test procedure, we ran each classifier on our corpus and computed the averaged yield precision, recall and  $F_1$  score. Section 4.3.2 shortly discusses state of the art of existing tools under the same constraints, and sections 4.3.3.2 to 4.3.3.6 give the details of the implementations of each of our classifiers, whose obtained performances will be discussed in section 4.3.5.

### 4.3.2 *NER* in Novels

The field of *NER* is evolving quickly, as big data approaches and access to increasingly bigger corpora brought up several projects promising interesting results. However, these tools remain mainly purposed to summarize, categorize and extract meaning out of short texts that relate to the outside world (e.g. news articles, encyclopedias or otherwise informative texts). Analyses on big portions of self-contained texts like entire chapters or full fiction books, whose named entities are difficult to look up and interpret out of their context, remain a challenging endeavor. Novels in particular are different from traditional applications for *NER* techniques for at least two reasons:

- Novels create a world of their own, with own recurring characters and places that can be hard to understand out of their context. Many traditional *NER* methods do not exploit this characteristic as they on the contrary rely on external databases to identify and classify named entities [Jovanovic et al.(2014)].
- Each novel tends to be characterized by specific stylistic features whose purpose pre-

cisely is to give the narration a unique taste. On the one hand, a novelist's writing diverges in nature from journalistic, academic or other forms of non-fiction writings, and on the other hand, it is usually purposely meant to introduce noticeable differences compared to the style of other authors. This phenomenon is obvious enough that it was shown to allow to confidently attribute texts to their authors in many contexts [Stamatatos(2009)]. This intrinsic diversity makes a one-size-fits-all approach to *NER* difficult and is problematic for pattern-inducing machine learning algorithms.

In order to establish a baseline for our study, we conducted an initial evaluation with one of the current widely used *NER* systems. *OpeNER*<sup>7</sup> is a multilingual and powerful *NER* pipeline, promising state of the art results on French corpora [Azpeitia et al.(2014)]. Even if it wasn't designed to handle big amounts of text<sup>8</sup> its processing routine is efficient and produces XML result files that can easily be processed and compared to our reference data. Since *OpeNER* is extracting a wider range of entities than what we considered (such as date and time information), we only considered the relevantly tagged words and manually resolved naming differences in order to keep the comparison fair and accurate.

#### 4.3.3 A Set of Simple Rules

##### 4.3.3.1 Preprocessing

In general, spotting proper nouns in French texts is a rather easy task due to capitalization rules, that are quite similar to English if not simpler due to less false positives (e.g. words related to languages, or days and months are written in lower case) [Geno(1992)]. Still, the task is not entirely trivial, because we need to filter out capitalization due to sentences starts and miscellaneous stylistic effects (such as subtitles, quotes, verses, etc...). Additionally, we want full names extracted as single entities and remove false positives, like honorific titles or named time periods [Grevisse and Lenoble-Pinson(2009)]. In order to do that, we designed a two-pass method:

- First, our scripts extract a list of all proper nouns that are not leading sentences. To do this, we compute all 1-grams, 2-grams and 3-grams of each sentence whilst letting out the first word. We then keep all isolated proper nouns (i.e. capitalized words surrounded by lower cased ones), pairs of juxtaposed proper nouns, and triplets of proper nouns

---

<sup>7</sup>The OpeNER project - <http://www.opener-project.eu>

<sup>8</sup>The JVM heap size limit set by JRuby needed to be increased to over 8 GB in order to process our bigger samples (yet only about 6MB in size)

joined by an hyphen or a nobiliary particle<sup>9</sup>, each of them as if they were single entities.

- Then, we run a second pass on the full text, where we look for all sentences containing each noun we identified in the first run to include legitimate nouns leading sentences we had left out. For each of them, we store an index of the sentences they appear in for further processing.

This approach allows to recover all occurrences of each capitalized word, as long as they are not systematically at the start of sentences. In practice, the resulting list needs to be refined thereafter, as some capitalized common nouns still happen to end up in it. The reasons for this may be multiple, ranging from sentence tokenization errors, typos in the source text or other stylistic effects (e.g. poetry excerpts, chapter titles,...). This refining can be done accurately and efficiently by combining three strategies:

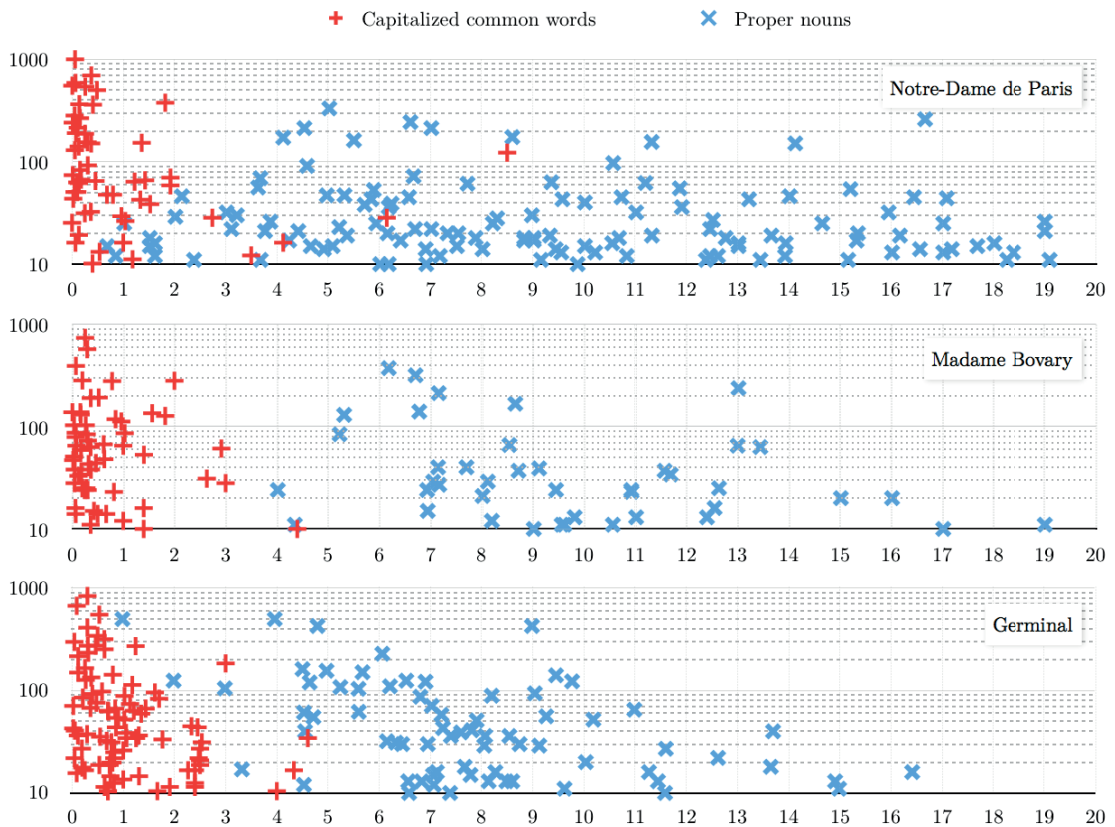
- By computing and comparing the mean positions of each word in the sentences they were found in. As we may expect, the distribution of words that are usually capitalized only when leading the sentences will concentrate towards 0, whereas the real proper nouns' will tend to even out. Figure 4.2 illustrates that the typical distributions often allow for easy separation, with very few outliers.
- The said outliers being usually connection words (e.g. pronouns and prepositions) that are often capitalized in different contexts, they can be filtered out using a simple French stopwords list.
- Thirdly, words related to nationalities or ethnic groups can be easily filtered out as they are usually present in both their singular and plural forms, whereas genuine proper nouns will usually not follow this characteristic. Hence, we remove the words ending in "s" that can also be found without the "s" throughout the same text.

Once identified, proper nouns usually fall in three main categories that serve different purposes to narration: they can namely be *characters*, *places* or *others* (brands, abstract concepts, acronyms...). We designed and evaluated six independent classifiers. Each classifier gets one word at a time as an input as well as the context that is necessary and relevant for its way of processing data, and returns the predicted category (namely *character*, *place* or *other*). We first present the implementation characteristics of each component before looking in details at the resulting scores.

---

<sup>9</sup>Since nouns can be in any language despite the text being in French, we considered the list of nobiliary particles found on the related English Wikipedia page: [http://en.wikipedia.org/wiki/Nobiliary\\_particle](http://en.wikipedia.org/wiki/Nobiliary_particle)

### 4.3. Reliable Recognition of Characters and Places



**Figure 4.2** – Typical mean positions of uppercased words in their respective tokenized sentences vs. their number of occurrences (on a logarithmic scale).

#### 4.3.3.2 Classifier #1: Obvious Context, Titles and Predicates

When one encounters a proper noun in a sentence, a good guess on its nature can sometimes easily be taken due to the immediate context. The simplest case, which we'll refer here as *obvious context*, would be if the noun is immediately preceded by a title or a predicate that hints at what it refers to. For instance, the name *Vilquin* could probably refer to anything without clear preference when out of any context, but it would be very easy to classify respectively as a person or as a place, if at least one sentence is mentioning "[...] *monsieur Vilquin* [...]" or, conversely, "[...] *rue Vilquin* [...]".

For this classifier, we compiled a simple list of obvious context classifiers that allow to make good guesses about the nature of the immediately or next to following proper noun:

- In the case of characters, the list is basically the usual name titles of the French language, such as *madame* or *docteur*<sup>10</sup>.

<sup>10</sup>Wikipédia : Titres et prédicats - [http://fr.wikipedia.org/wiki/Titres\\_et\\_prédicats](http://fr.wikipedia.org/wiki/Titres_et_prédicats)

## Chapter 4. Methods for Digital Writing Tools

- For places, we considered the most common predicates used to qualify toponyms, as for instance *ville*, *avenue* or *rue*.
- For other words we wanted specifically to filter out, we added a short list of terms referring to deities (*dieu*, *jésus*, *marie*, *vierge*, *saint*).

### 4.3.3.3 Classifier #2: Naive Position

In French, like in many other languages, the grammatical structure makes it more likely for sentences to follow a pattern that puts the subject of the action at the beginning, and the location towards the end. This characteristic can be used when one looks at enough examples to make a simple, yet quite powerful guess about the global roles of the proper nouns.

The accuracy of this classifier is indeed strongly dependent on the writing style of the author, as the frequent use of specific figures of speech may break its work hypothesis, and longer sentences may narrow the gap between the categories or blurry their boundaries. This can be seen clearly in fig. 4.3, where we show the relative positions of identified classes of names for three different stories. In those examples, we can see the effectiveness of a separation guess at around 45% of sentences' length, which is expected to yield quite good results for the first two books, yet a bit more disappointing ones for the third one.

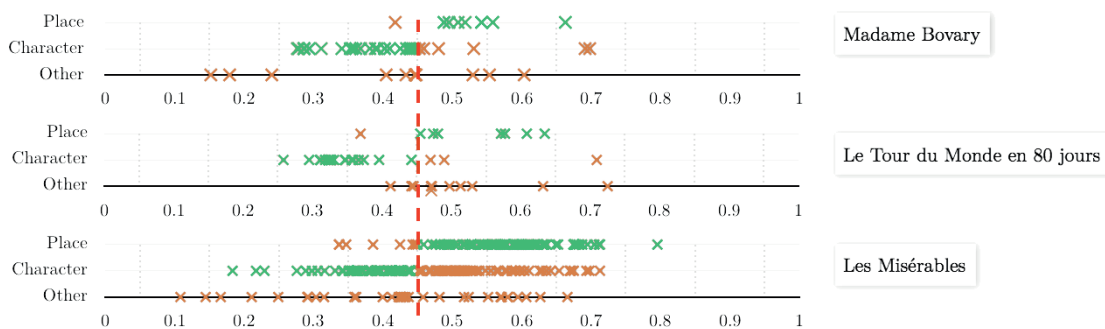


Figure 4.3 – Relative mean position of characters and places names for three classical French novels

### 4.3.3.4 Classifier #3: Semantics out of Neighboring Words

Inspired by Latent Semantic Analysis [Dumais(2005)], in which it is hypothesized the meaning of words are the result of their neighbors', a slightly more sophisticated approach consists of a broader look at neighboring words semantics. This approach is very different from section 4.3.3.2 in that we aren't expecting the surrounding words to qualify the noun per se, but to hint at its nature due to the actual meaning of the near context. For this implementation, we compiled lists of words that are more likely (but not exclusively) to appear respectively

### 4.3. Reliable Recognition of Characters and Places

nearby characters, places or abstract concepts. For instance, we expect names of characters to be more often surrounded by words related to emotions, body functions, speech or professions, whereas names of places would be more closely related to motion verbs, place features and prepositions.

Starting from common nouns that are unambiguously related to one of the categories we are interested in, we used a French synonyms dictionary service<sup>11</sup> to put up a list aiming to be as extensive as possible. The final files resulted in 4'500 words for characters, 670 for places and 50 for concepts (see appendix A.4.2 for the complete list). The script then looks for these words in the neighborhood of the nouns to be disambiguated, and returns the most probable category.

#### 4.3.3.5 Classifier #4: Grammatical Structure

As character and places serve different narrative purposes, one may expect the grammatical constructs surrounding them to differ in a significant way. For instance, place names are often preceded with prepositions or determiners, whereas it is common for characters names to be directly followed by verbs. We thus introduced a script classifying names based on its knowledge of the full text, grammatically tagged using TreeTagger<sup>12</sup> and tokenized in sentences. To guess the nature of names, it then matches all sentences containing them against a set of rules that are typical constructions one uses when writing about a person or a place.

$t_{i-2}$	$t_{i-1}$	$t_{i+1}$	$t_{i+2}$	<b>Result</b>
(at least one is a Verb)				<i>character</i>
VER				<i>character</i>
NAM				<i>character</i>
PRP	PRP		-	<i>place</i>
	VER			<i>place</i>
	DET			<i>place</i>
	PRP	(empty)	(empty)	<i>concept</i>
	PUN			<i>unknown*</i>
	(empty)	(empty)		<i>unknown*</i>
*(probable tokenization error)				
VER = Verb, NAM = Proper Noun, PRP = Preposition				
DET = Determiner, PUN = Punctuation				

**Table 4.1** – Rules for grammatical structure classifier

<sup>11</sup><http://www.synonymo.fr>

<sup>12</sup><http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger>

We tried out a set of seven manually established rules covering the most straightforward grammatical constructs (described in details in table 4.1), plus two that help filter out tokenization errors at a sentence level by flagging words that are preceded by a punctuation mark or that are alone in their sentence. Formally, rules are patterns possibly matching the set  $\{t_{i-2}, t_{i-1}, t_{i+1}, t_{i+2}\}$  of grammatical tags attributed to up to four neighboring tokens  $\{w_{i-2}, w_{i-1}, w_{i+1}, w_{i+2}\}$  for each proper noun  $w_i$ . When matched, they increase or decrease the probability score for one or several classifications, and the category yielding the highest score gets picked and returned.

### 4.3.3.6 Classifier #5: Online-Querying

A lot of proper nouns can be non-ambiguously or with a high probability related to one or several categories based on general knowledge. For instance, a human could make a guess that "Elisabeth" will most likely refer to a person, whereas "Manhattan" is likely to be a place and "Amour", a concept. But the same knowledge may equivocally tell that those same words could also potentially be related to the ship (*RMS Queen Elisabeth*), an abstract concept (*project Manhattan*), or a place (*Amur river*), probably with a lower likelihood if no other context is available.

For many nouns, the knowledge we are looking for is well captured in the categorization of their related Wikipedia pages. Using categories instead of the text of the articles also presents the advantages of being very straightforward and reduces a lot noisy signals related to text processing techniques. To test this idea, we implemented a simple algorithm that gathers the categories of the page whose name is closest to the noun we are looking for and looks for ones tagging people, places, or abstract concepts. In the case no category gives a hint (which tends to happen both with very complex or very precise pages), it tries to recursively walk up the hierarchy until the necessary clues are found.

### 4.3.3.7 Classifier #6: Quotes-Based

Several works already showed the relevance of locating direct and indirect speech parts to identify characters in novels [Glass and Bangay(2007), Goh et al.(2012), Karsdorp et al.(2012)]. Most of these approaches rely heavily on the lexical database WordNet<sup>13</sup> to find out speech-related verbs and refine their accuracy, but for performance reasons and since we wanted the classifiers to remain efficient even on very long texts we implemented a simpler version that simply checks the proximity of detected proper nouns to quotation marks. For each proper

---

<sup>13</sup>WordNet | A lexical database for English - <https://wordnet.princeton.edu/>



noun  $w$  appearing  $m_w$  times, the system would essentially count the number  $q_w$  of mentions that appear near quotations. It then computes:

$$r_w = \frac{m_w}{q_w} \quad (4.1)$$

the quotation ratios for each noun, and:

$$\tilde{t} = \text{median}\{r_w\} = \begin{cases} r_{\frac{n}{2}} \text{ for an odd number } n \text{ of words} \\ \frac{1}{2}(r_{\frac{n}{2}} + r_{\frac{n+1}{2}}) \text{ for an even number } n \text{ of words} \end{cases} \quad (4.2)$$

the median value of all ratios, to be considered the differentiation threshold for this book. Each noun  $w$  is then assigned the *character* class if  $r_w$  is higher than  $\tilde{t}$ , or *place* otherwise.

#### 4.3.4 Meta-Classification

Once all classifiers returned their answer for a given word, the last step is to compare these results and to decide on a final answer. This meta-classification step can be done by a simple voting system, where the final result will be chosen according to the majority of predictions with regards to the confidence indices, or by using a meta-recognition system, aiming to discard classifiers that encountered a problem on the considered text file. We implemented and discussed the performance of four distinct meta-classification methods.

##### 4.3.4.1 Simple Vote

The easiest and most obvious solution to average the different classifications is to use a simple voting system, i.e. the classification who gets the majority of results wins. However, since there is an even number of classifiers, ties are to be expected, in which case the choice will be nondeterministic by lack of model to support one option over the others. For this reason, we introduced a second meta-classification, which involves for each classifier to compute a confidence self-assessment score.

##### 4.3.4.2 Self-assessed confidence

For most classifiers, their internal mechanics allow themselves to evaluate to which extent the strategy they are using seems likely to return reliable results, given the current work context. Hence, a simple strategy to help the voting process in the case of ties is for each classifier to return a confidence index, between 0 and 1.

## Chapter 4. Methods for Digital Writing Tools

---

The used self-assessing strategies are as follows:

- For *Obvious Context*, *Semantic* and *Wikipedia* classifiers (sections 4.3.3.2, 4.3.3.4 and 4.3.3.6), we use  $\frac{C_{max}-C_{min}}{\sum C_i}$  where  $C_{max}$  is the count of most and  $C_{min}$  of least represented categories for which we found clues. For instance, if the *Semantic* classifier finds 6 neighboring words related to places, 3 to people and 1 to abstract concepts, the confidence index will be  $\frac{6-1}{(6+3+1)} = 0.5$ . This index is thus expected to equal 1 if the decision was made with no ambiguity and 0 if the clues were equally distributed.
- For *Naive Position*, *Grammatical* and *Quotes* (sections 4.3.3.3, 4.3.3.5 and 4.3.3.7) classifiers, we use the difference between the splitting threshold and observed decision value, normalized between 0 and 1. For instance, considering the *Quotes* classifier computes a ratio of 0.4 for some noun with a threshold of 0.2, the result will be classified as a person with a confidence index of  $\frac{0.4-0.2}{1-0.2} = 0.25$ . Again, this index is expected to see its value tend towards 0 for ambiguous cases, and towards 1 for the more definite ones.

On top of that, some classifiers are given the possibility to return 0 to mark their results as known to be invalid, and thus irrelevant at voting time. This can happen for instance when we don't find any known title preceding a word throughout the text, if no grammatical rule could be matched, or if Wikipedia doesn't have any result for the searched word.

The improved voting algorithm then first discards all classifications that have a confidence mark of 0, and proceeds to a simple vote between the remaining ones for each noun. In case of a tie, the results rated with the highest confidence will be privileged.

### 4.3.4.3 Fixed weighting

Not all classifiers exhibit the same behavior regarding precision and recall. It thus can be justified to put more confidence on some of them in cases when we know they are more likely to succeed. For this test, we used manual weights putting more importance to the *obvious context* classifier (section 4.3.3.2), due to its high precision rate, all others being treated equally. With the help of confidence rating (section 4.3.4.2), we know the low recall rate will not impact negatively on the other classifiers because it will return a confidence score of 0 if it couldn't find any classification clue. Hence, those cases will be discarded regardless of the coefficient. A good compromise can be reached by giving 3 times more weight to the *obvious context* classifier, allowing the others to still easily overpower it in the unlikely case a majority of them reach a contradictory agreement.

#### 4.3.4.4 Meta-recognition approach to optimized weighting

A meta-recognition algorithm follows the idea of improving its accuracy by entirely removing one classifier if it detects it is consistently failing, typically due to stylistic biases or other broken assumptions on the considered book. Given the global classification results, one can easily compute an agreement score between the different classifiers, for instance using Fleiss' Kappa method [Fleiss and Cohen(1973)]. In case this indicator hints at discrepancy, we can simply recompute all averages by systematically letting out one of the classifiers, until the new Fleiss' Kappa value increased. Our hypothesis here is that since the remaining classifiers reached a higher agreement, the discarded one must have globally failed in some way and needs to be put aside.

4.3.5 Results

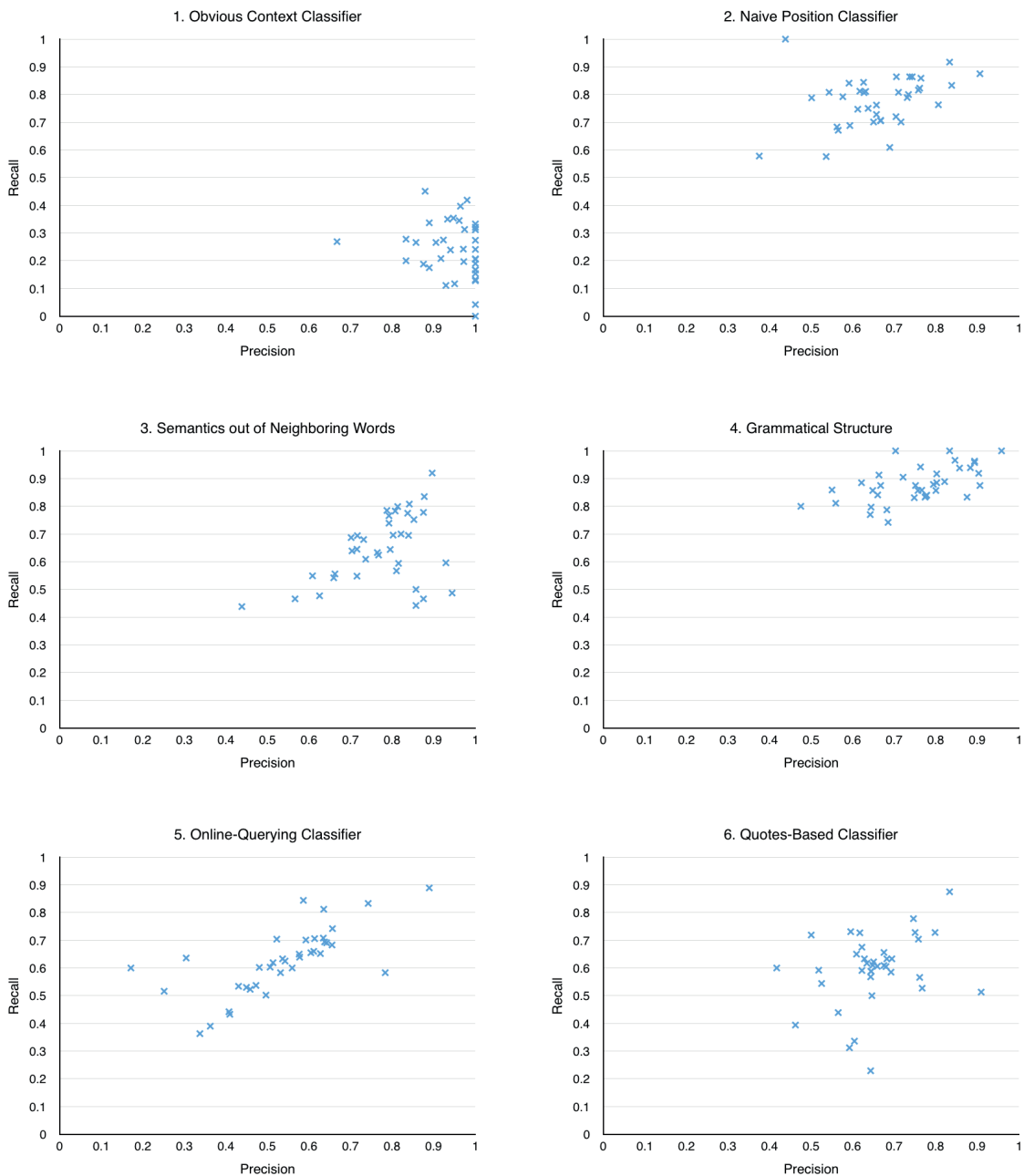


Figure 4.4 – Comparison between Precision and Recall for each classifier, on each book

Let us consider in fig. 4.4 the precisions vs. recall results each of the six classifiers got on our test corpus. One can immediately see a typical pattern in any information retrieval system: one parameter is detrimental to the other, and no two classifiers behave in a similar way. We can also see that for each of them, some books get incredibly good results, and few others

### 4.3. Reliable Recognition of Characters and Places

turn out very bad. Interestingly and as backed up by the full numerical values shown in table A.1 (appendix A.4.1), those are almost never the same, confirming our hypothesis that some methods may work way better (or worse) on some texts, giving a strong backing for the multi-classifier approach. The averaged results seem to confirm this intuition. In figs. 4.5 and 4.6, we can see that all meta-classification schemes overall pushed the results upwards, and at the same time made the clustering denser, hence reducing the differences between the books and output more constant results by removing the worse outliers.

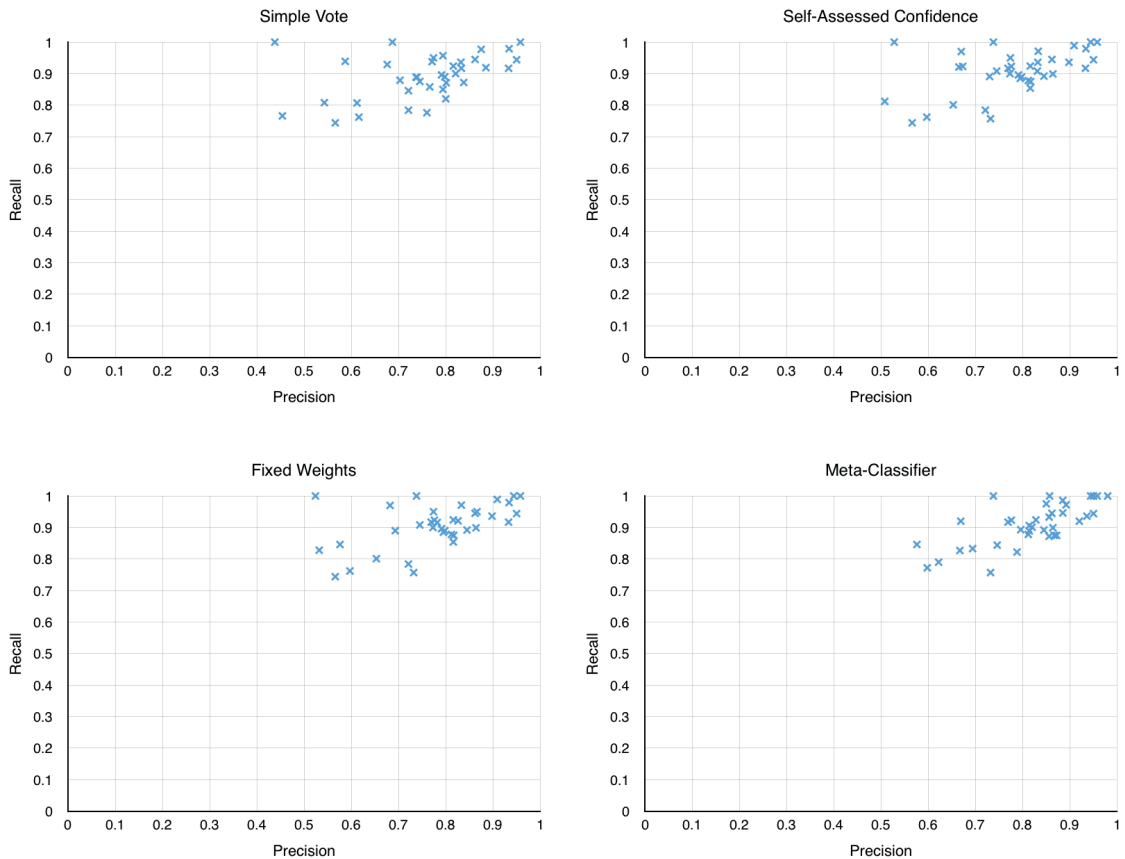


Figure 4.5 – Comparison between Precision and Recall for each meta-classification



Figure 4.6 – Graphical comparison between meta-classifications

By looking at the numerical results (appendix A.4.1, table A.1), several interesting statements can be made about each classifier:

- Classifier #1: As one may expect with this kind of simple implementation, the resulting predictions usually translate to a very high precision (100% in nearly 2/3 of the cases, and on average 0.949) but a poor recall rate (only 0.239), as the presence of enough context words leaves little doubt on the categorization, though unlikely to happen for most of the nouns we would want to disambiguate. It is also interesting to see that its highest score was achieved on *La Comédie Humaine*, one of our longest samples, with an  $F_1$  score of 0.596. As a matter of fact, all three sagas ranked quite high compared to the shorter books (*Les Misérables* and *Les Mystères de Paris* both got an  $F_1$  score higher than 0.5).
- Classifier #2: Unsurprisingly, this classifier is one of those that have the overall worst performance when compared to the others. However, one has to note that with a precision and recall of respectively 0.654 and 0.773, this very simple method still outperforms possible more complex ones.
- Classifier #3: The semantics approach surprisingly didn't perform as well as one may expect. Its precision of 0.769 is the second highest, but it comes with a cost of a recall of 0.651, which brings its  $F_1$  score even below the one of #2.
- Classifier #4: The grammar classifier resulted in the overall highest  $F_1$  score (0.807), with a precision of 0.750 and a recall of 0.879. One can notice it performs best, with a very satisfying 0.909  $F_1$  score, on *Les Malheurs de Sophie*, which could be explained by the fact that as a book for children, its syntax might be more regular than other novels.

- Classifier #5: The quotes classifier did not perform very well. Its  $F_1$  score (0.568) is the one to lowest and it has the lowest precision rate of all (0.528). Yet it still achieved a very good score on the novel *Thérèse Raquin*, where it has actually got the highest precision, recall and  $F_1$  scores amongst all classifiers.
- Classifier #6: This method resulted in a rather disappointing overall precision and recall of respectively 0.644 and 0.597, which can be very well explained by that only few fictional entities are expected to be represented on Wikipedia, and among them, many names would actually relate to real world counterparts that may be very different from the fictional use of the same word.

Regarding the meta-classification schemes (appendix A.4.1, table A.2), we first wanted to compare them to our the *OpeNER* baseline. On our test corpus and using similar evaluation *OpeNER* averaged a precision and recall of respectively 0.609 and 0.754. This may seem surprisingly low compared to the standards usually set by this tool, but actually is a good illustration of how difficult it may get to find the correct tagging on fictional texts. The two worst cases (*Les Malheurs de Sophie* and *Germinie Larcerieux*) are shown in details in tables A.3 and A.4 (appendix A.4), and show that those problems are about as much related to bad classifications as to missing entities. Our meta-classification methods perform on average better than that, and we can see an encouraging trend that the various strategies we tried out tended to get better results and to close the standard deviation gap. That being said, as far as standard deviation is concerned and taken into account, this improvement is not statistically significant. The meta-recognition method managed to push up the  $F_1$  score to no worse than 0.674, and as high as an incredible 0.99 on one of the books, which in other terms meant that all entities were found and only one of them ended up misclassified.

#### 4.4 Visualizations Based on Named Entities

In an interview from 2005<sup>14</sup>, the famous contemporary American writer Kurt Vonnegut exposed the idea of understanding stories by interpreting at their *shape* [Vonnegut and Simon(2007)], simply by looking at good and ill fortune events from the main character's perspective. Ironically, this concept allegedly didn't convince at the time he presented this idea for his Master's thesis and never was intended as more than merely underlining interesting patterns without claim of general truth. Yet, it recently ended up sparking actual research on those exact topics, with notably an investigation on the possible and most represented emotional arcs [Reagan et al.(2016)]. Similarly, a famous attempt at representing parallel storylines half-

---

<sup>14</sup>Kurt Vonnegut on the Shapes of Stories - <https://youtu.be/oP3c1h8v2ZQ>

jokingly hand drawn by Randall Munroe<sup>15</sup> sparked a lot of interest in the field, as several projects successfully proposed tools to automate the generation of those graphs, especially using modern web technologies [John et al.(2016)].

The reliable pinpointing of characters and places mentioned in each chapter, combined with the narrative focalization modeling, allows for interesting and accurate projections of the narrative structures at the whole story level. In sections 4.4.1, 4.4.2, 4.4.4 and 4.4.5, we will present different visualizations potentially useful to any novelist, that can be computed to explore various meaningful ways of projecting those dimensions automatically after extraction of the relevant named entities. For the sake of demonstration, and to include a wider range of examples the reader may easier relate to, some graphs shown in those sections were computed on sample books from our *NER* test corpus instead of *La simulation humaine*.

### 4.4.1 Mention Frequencies Histograms

Words frequencies are generally speaking interesting features when looking for patterns inside long organized texts. They have for instance proven efficient to quantify trends in cultural evolutions in a mass study covering millions of digitized books [Michel et al.(2011)]. While not computationally intensive, they are particularly relevant when looking at named entities inside of a closed corpus, where a mere presence (or, conversely, total absence) inside a section is a relevant structural marker, whereas many repetitions in one single section is expected to be a good marker of a strong narrative presence (related to descriptive acts, lots of interleaving actions or dialogues).

The representation allowing to compare the frequencies of several words among themselves is usually done using a *words cloud*, which removes the notion of temporality or repartition for each word. Conversely, one may be interested to look at the use of one specific word through an organized corpus, for which we suggest the use of a graph that follows the logic of an histogram. In comparison to some other approaches to visualize words frequencies, such as the Google N-Gram Viewer<sup>16</sup>, the histogram variant has the advantage of not displaying any data point in the places where there is no occurrence, which thus renders the view less cluttered and better suited to cases where this situation is to be often expected, as it is the case for named entities in novels. Figure 4.7 shows a frequency visualization in the typical form of an histogram representation, displaying the repartition of occurrences of the word "*atomique*" in *La simulation humaine*:

---

<sup>15</sup>xkcd | Movie Narrative Charts - <https://xkcd.com/657/>

<sup>16</sup>Google Books Ngram Viewer - <https://books.google.com/ngrams>



#### 4.4. Visualizations Based on Named Entities

- Each tick represents one chapter, in reading order,
- Percentages are given relative to the total number of occurrences of the word).

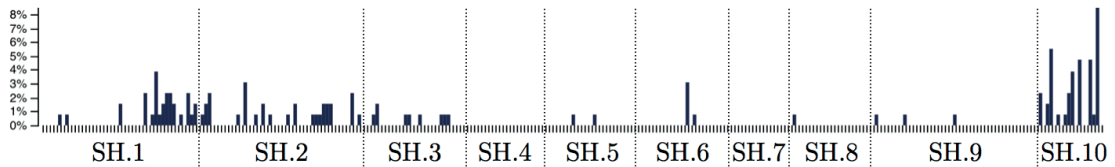


Figure 4.7 – Repartition of uses of the word "atomique" in the ten books of *La simulation humaine*.

Of course, one cannot claim this kind of representation will be informative for every possible term. Moreover, interpretations need to be used with caution and probably together with a certain knowledge of the text, which doesn't per se best reflect distant reading. Nevertheless, simple well-chosen examples can shed some light on the analytical possibilities of such a measure. For instance, one can be interested in the evolution of the representation of one specific character within a story spanning over several years. Figure 4.8 illustrates the results of such a query about the word "Cosette" in *Les Misérables*. Notice how this case allows to immediately identify the different parts of the story as well as narrative transitions such as Cosette evolving from secondary character status to main protagonist after the first quarter of the story with a dramatic increase in presence.

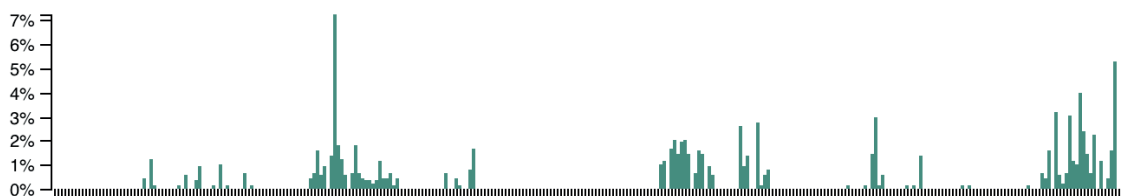


Figure 4.8 – Frequency of mentions to "Cosette" in *Les Misérables*. Each tick represents one chapter, in reading order. Percentages are given relative to the total number of occurrences of the word.

If histograms are an efficient way to hint at where one needs to look for relevant information, they make poor comparison tools when more than two datasets are concerned. As shown in fig. 4.9 (left), the joined display of data on the same line is one possible option, yet it heavily relies on colors and will understandably get difficult to read when the numbers are very uneven. Another option is to stack the bars vertically (fig. 4.9, right), which results in a slightly less cluttered view, yet makes visual appreciation less intuitive as the vertical reference point will change for each series.

## Chapter 4. Methods for Digital Writing Tools

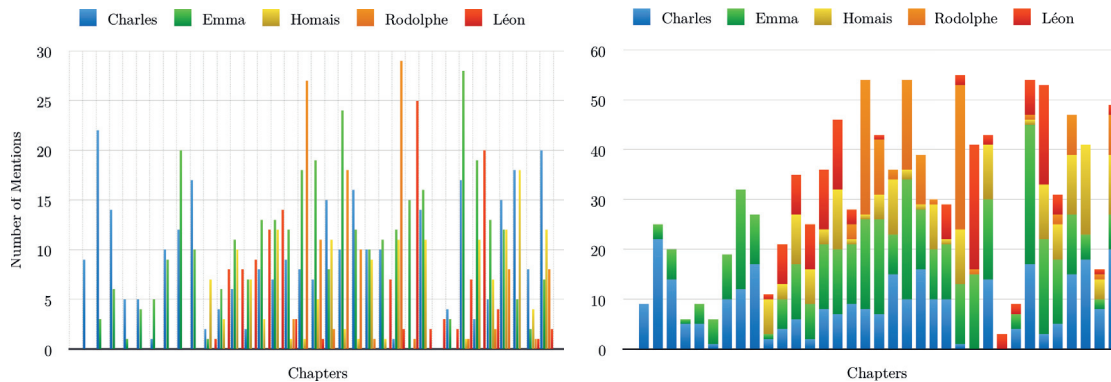


Figure 4.9 – Possible protagonists mentions comparisons on *Madame Bovary*, using histograms.

The question of visual patterns identification in data series rose early in domains such as biology, where the usual solution consisted of a projection on fewer dimensions with the help of standard deviation depiction [Hubbs and Hubbs(1953)]. Such a solution would however not be an improvement in our case, since we want all dimensions to be kept separate in order to be able to refer to the concerned parts of the text.

Technically speaking though, solutions do exist, driven by the need for such tools in fields featuring frequent pattern identification by comparison of large datasets. Alternative visualization means, such as circular layouts or parallel spirals [Weber et al.(2001)] indeed exhibit interesting properties, but at a cost of less intuitive representations for non-technical users. A simpler although less satisfying solution can also be to revert to line graphs, such as shown in fig. 4.10 (left) with the same dataset. Here again, reading might get improved by stacking the lines, resulting in a neater visual effect but at the cost of a fuzzier interpretation (fig. 4.10, right).

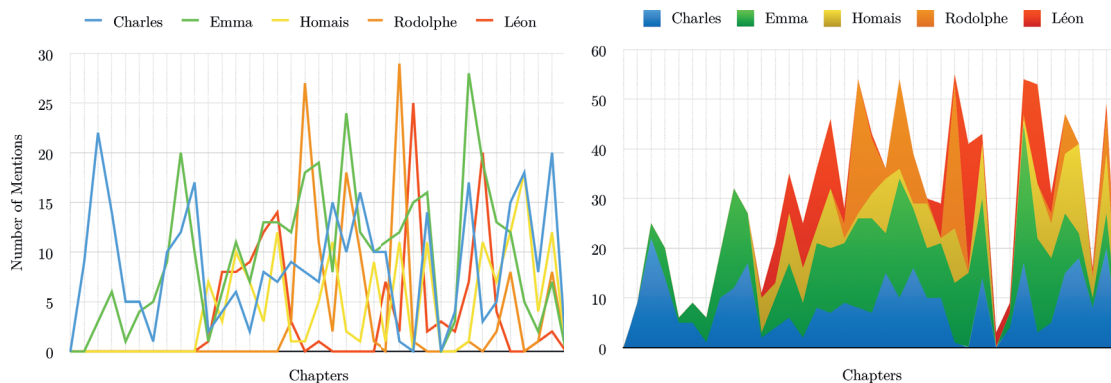


Figure 4.10 – Possible protagonists mentions comparisons on *Madame Bovary*, using line charts.

### 4.4.2 Characters Networks

Characters, be it in the form of persons, animals or any kind of acting entities, are probably the most basic and essential component of storytelling. For one thing indeed because a story cannot exist without at least one acting subject, but also because the emotional connection between the reader and the fictional protagonists is a key factor for predicting affective and memory responses [Kurkjian and Livingston(2007), Roser et al.(2007)]. This is reflected in the writing process of many authors, who keep records with detailed information about their characters, such as personal and biographical features, but also more importantly organized and possibly evolving relations to other characters.

In any typical narration, as in the real world, the development of intertwining events comes with a thickening of the interactions and relations between the various characters, which builds a mesh carrying its own memory of the story. Already back in 1934, Jacob Levy Moreno showed the predictability of roles in real life communities through connections between individuals, founding sociometry and opening the way for modern social network analyses [Moreno(1934)]. With the intermediary representation of so-called *sociograms* (e.g. the graphical representation of such networks), he managed to explain group behaviors and made key contributions to psychiatry and group psychotherapy. More recent works have shown the validity of these methods within a literary approach, precisely with sociograms structure providing insights about the narrative roles and relative importance of each character [Rochat et al.(2013)]. Different interpretation methods were suggested, with notably the various mathematical definitions of centralities opening interesting formal analysis approaches [Rochat(2014)].

Therefore, it is not surprising that automatic representations of fictional social networks have been a central topic of research spanning from rather traditional approaches [Sack(2011), Kydros and Anastasiadis(2015)] to distant readings challenging characterizations established by literary scholars [Elson et al.(2010)], and have been particularly relevant in the context of analyzing Scandinavian sagas [Tangherlini(2015)]. Of course, the range of possible representations is very wide, given networks can be built directed or non-directed, simple or multipartite, potentially clustered or weighted in very different fashions, each leading to different possible uses and interpretations. Closer to our preoccupations, recent works in artificial intelligence proposed tools and methods using sociograms as a foundation for generating simple narratives [Sack(2014)], giving a strong sign for relevance of social network analysis tools for novelists.

Where some qualitative works used manually annotated data sources and discussed ways to make the extracted representations as accurate as possible according to the contextual

## Chapter 4. Methods for Digital Writing Tools

meaning [Moretti(2011), Agarwal et al.(2012), Kydros and Anastasiadis(2015)], a simple quantitative way to determine links between characters can be to look for names mentioned one or more times together in the same sentence, vicinity, or meaningful text excerpt. Following this idea, some works use mentions present in same or neighboring pages, who can thus be computed only out of names indexes [Rochat(2014)], whereas others argued for automatic information extraction with empiric proximity rules. For instance, Elson et al. considered relations between characters appearing within the same dialog, separated by less than 300 words and mentioned at least three times in the whole novel [Elson et al.(2010)]. Ultimately, the definition of relations is as equally shaped by work hypothesis than technical constraints, since different quantities of text or annotation qualities have a big impact on what can in the end be computed.

Looking at the results produced by the tools we described in section 4.3, a first obvious choice was simply to consider a connection between characters mentioned in the same sentence, and to weight the edges using the number of those cooccurrences. This approach is expected to result in a clear lower bound on the relations and their intensities, which makes sense since we aim at designing tools working with texts long enough for these features to be representative. Section 4.4.2 shows a comparison of such an automated extraction for two novels, *Frankenstein* and *Madame Bovary*. Notice the clustered organization of the first one, where one character (i.e. the creature) interacts independently with characters that do not know each other.

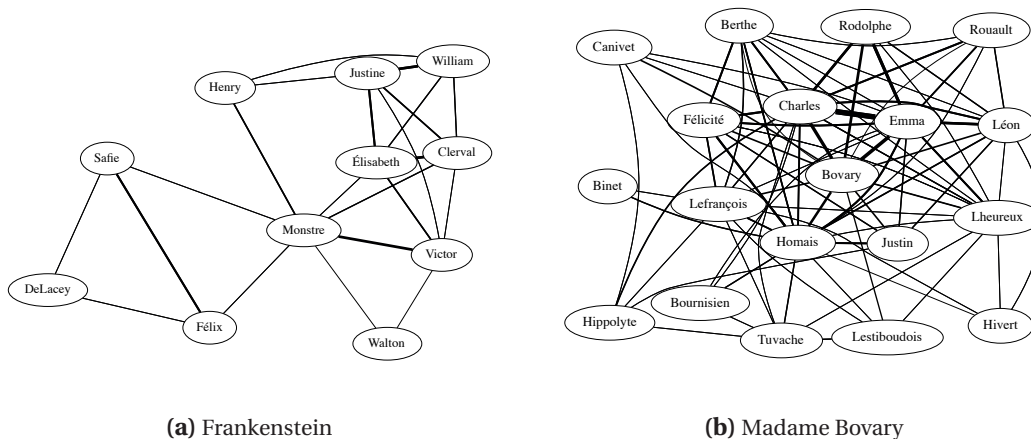
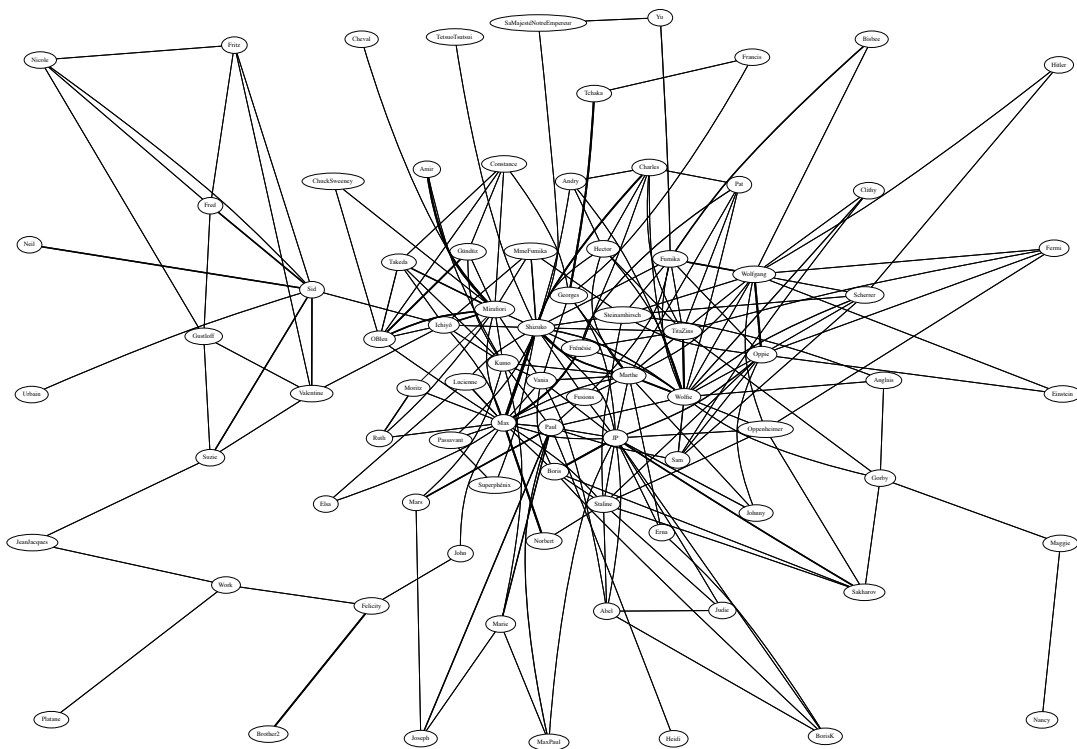


Figure 4.11 – Social network graphs comparison between two novels.

Especially when computed in an automated fashion, social network graphs can be criticized as a rather unsophisticated view that doesn't provide any clue on the natures of the relations they display. Nevertheless, they have several advantages that make them an interesting tool. First, such visualizations can be computed very efficiently and without a priori knowledge

#### 4.4. Visualizations Based on Named Entities

about the corpus other than the list of named entities to consider (it will thus independently work as well on a different language, almost any literary genre, or across a very wide range of stylistic variations). Additionally, as shown in our example, by following this principle one can use the cooccurrences counts ratios to compute an estimate of the strengths of the relations, or potentially derivate other kinds of more intricate graphs by blending in more information at our disposal, as for instance to visually differentiate links relating protagonists and second characters.



**Figure 4.12** – Social network computed from *La simulation humaine* (also available at <http://simulationhumaine.com/socialnetwork.pdf>)

As stories grow bigger, so do the number of characters and the size of the networks. Figure 4.12 shows the level of complexity one may expect from the full network extracted from *La simulation humaine*. Of course, not all aspects of such big meshes considered as a whole are noteworthy, yet the overall structure features an interesting organization showing main characters in the middle and respecting almost all the relations intended by the author. We reproduced in fig. 4.13 the central part of this network, highlighting the characters part of the two central families, as documented by the author (fig. 3.2b, section 3.3). Notice how most of the protagonists from the two central families (actually 13 out of 17) are given a high centrality

and almost all connecting edges between them were correctly established. The higher level of cluttering compared to the sole genealogical tree can be explained by the detection of more relations than hand-drawn ones (*Frénésie*, *Lucienne* and *Vania* appear to be more important to the story than what the author first intended), but also in big part because of the presence of company names (*O-Bleu* and *Fusions*) and multiple reference of the same characters under different names (*Oppie* and *Oppenheimer*, *Wolfgang* and *Wolfie*). Of course, the first two features are rather desirable in the general case, whereas the last one is a default that should be corrected for most applications, for instance by using coreference resolution techniques. We would however like to point out that coreferences seem to be positioned in a near vicinity to each other in the graph, which could thus constitute a valid approach for detecting those cases when names are similar enough.

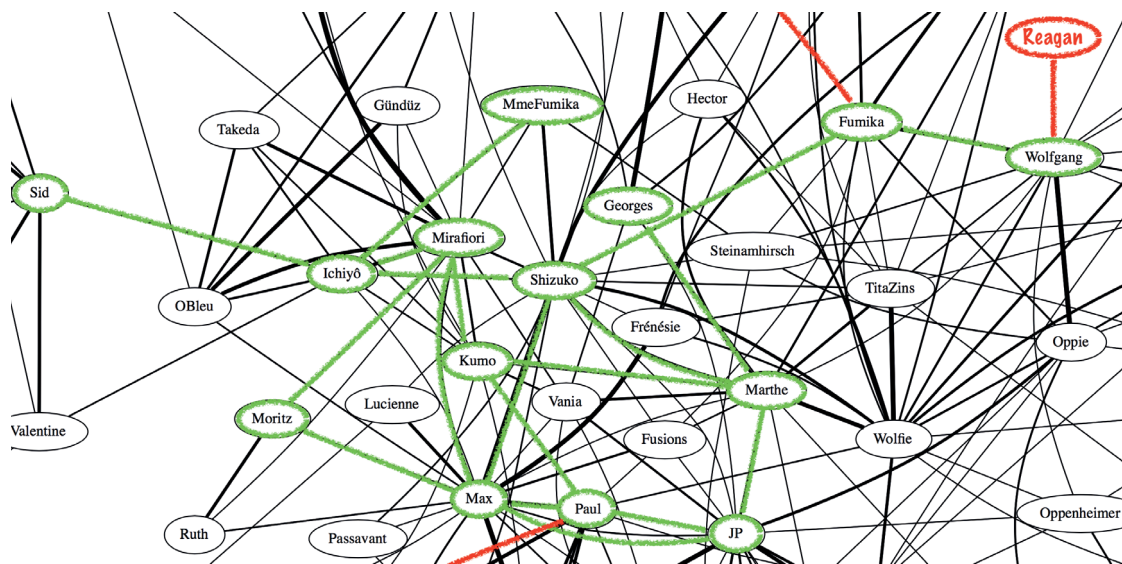


Figure 4.13 – Detail comparing the computed network with the author’s vision (in green, matching parts vs. in red, missing edges and nodes).

Of course, not all aspects of such big meshes considered as a whole are interesting, and if more than anything, this kind of representation calls for reduction techniques. For instance, an interesting compromise could be reached by considering a time range that doesn’t span the whole story, as some scholars pointed out that "static" global social network had a tendency of emphasizing wrongly some links [Agarwal et al.(2012)].

On the contrary and in cases when the size of the visualization is not an issue, the model can be extended to take into consideration more parameters. For instance, instead of representing direct relations among characters, one can look at places that mediate between them, resulting in a bipartite network. An example of such a graph is shown in fig. 4.14 on *Madame Bovary*.

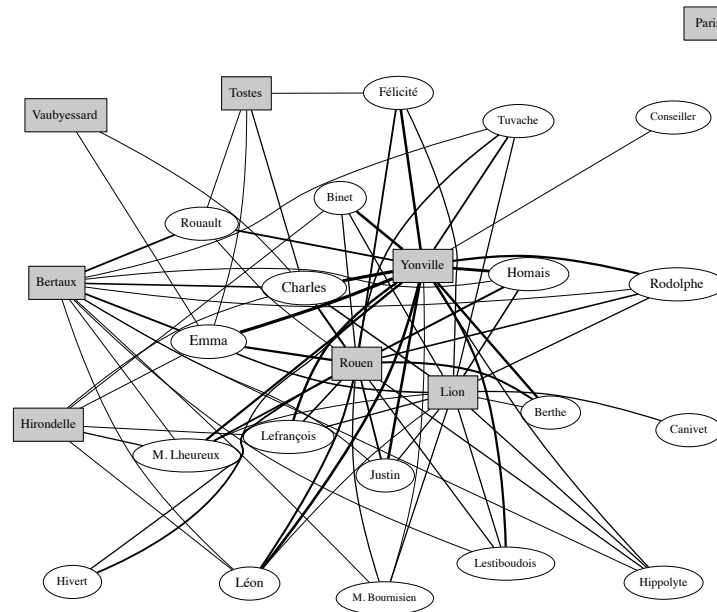


Figure 4.14 – Bipartite graph of the characters-places network in *Madame Bovary*.

#### 4.4.3 Wording Features Using LIWC

Traditional research in the analytics of storytelling focused on formalization of the narrative organization, possibly leading to efficient summarization [Rumelhart(1975)]. While these efforts hold promising possibilities with the venue of information technologies and artificial intelligence, the necessity to code by hand the knowledge representation of the story hampered heavily any possible real world application. Recent achievements in automatic stories understanding, while holding interesting propositions, stayed limited to very simple and short stories [Mueller(2003), Mueller(2006)].

Linguistic Inquiry and Word Count (*LIWC*) [Pennebaker et al.(2001), mcc(2011)] is a software designed for outputting semantical statistics based on words occurrences and categorization. It has often been used for basic natural language processing in social sciences in multiple languages, including in French [Boinon et al.(2011), Kergosien et al.(2015)]. The computed classifications results have been claimed reliable and psychologically meaningful [Chung and Pennebaker(2012)], even if the validity of some studies using them has been subject to debate, in particular when processing unfiltered datasets [Pury(2011)] or texts for which the embedded dictionary wasn't suited [Duval and Pétry(2016)]. On a corpus such as *La simulation humaine*, LIWC can indeed easily generate a wide range of features, accordingly

## Chapter 4. Methods for Digital Writing Tools

to the word categories present in the considered dictionary and relatively to an arbitrary unit of division of the text. Given the existing structure we decided to segment the corpus in the existing 298 chapters, and run LIWC on each of them, using its embedded French dictionary. The produced file is a matrix with 24 436 entries, giving the relative representations of 82 words and punctuation categories for each chapter (see fig. 4.15 for a illustrative example depicting a small excerpt).

Book	Chapter	WC	WPS	Sixltr	Dic	Num	foncti	pronom	pronom	je	nous	vous	il	ils	prono	article	verbe	verbe	verbe	verbe	verbe	verbe	adver	prepos	conjon	negati	qua
1	1	2957	12.91	26.21	75.89	0.1	47.48	12.21	11.6	1.66	0.24	0.57	7.03	2.1	0.61	18.19	14.34	2.54	4.63	6.8	1.08	4.73	12.75	4.5	3.25	6.5	
1	2	677	13.54	26	81.54	0.59	52.14	19.2	19.05	6.5	5.02	4.43	3.25	0.15	0.15	15.66	14.03	3.25	3.25	6.79	1.77	3.55	14.18	5.61	2.36	3.6	
1	3	3253	14.52	26.59	77.19	0.22	47.86	12.14	11.8	1.05	0.12	0.18	9.38	1.14	0.34	19.64	13.83	2.52	5.35	5.81	0.43	3.78	13.71	4.37	2.61	6.0	
1	4	789	17.53	27.38	82.38	0.38	53.99	19.52	19.26	7.35	2.03	5.83	4.06	0	0.25	15.34	15.34	3.04	4.06	7.22	1.39	4.31	12.42	7.22	2.41	4.6	
1	5	1524	13.03	25.2	75.39	0.33	46.85	11.42	10.89	1.77	0.33	0.33	7.28	1.18	0.52	20.41	14.04	2.36	4.13	7.41	0.66	4.79	13.45	5.12	1.97	5.7	
1	6	2763	16.16	26.89	76.8	0.14	47.45	13.93	13.54	1.81	0.07	0.36	10.39	1.12	0.4	18.42	12.27	2.17	2.97	7.02	0.43	4.34	13.03	4.52	2.71	6.2	
1	7	2157	12.4	26.38	76.63	0.37	48.63	14.14	13.91	1.76	0	0.28	10.38	1.76	0.23	19.1	12.98	1.76	3.15	7.28	0.56	3.94	13.95	3.89	2.18	6.0	
1	8	1748	16.34	27.12	74.2	0.46	46.57	12.36	11.73	1.43	0.17	0.34	9.32	0.69	0.63	19.68	11.27	2.46	2.35	6.46	0.51	3.03	14.93	4.41	2.63	5.2	
1	9	1755	12.54	28.38	74.59	0.46	46.78	9.63	9.29	1.37	0	0.06	7.29	0.57	0.34	22.74	11.74	2.17	2.91	6.61	0.51	3.36	14.87	4.1	1.88	6.2	
1	10	2303	13.01	25.92	75.34	0.13	47.89	11.16	10.73	1.43	0.09	0.48	7.25	1.48	0.43	20.97	13.42	1.39	3.99	6.08	0.96	3.86	12.64	4.34	2.3	7.0	
1	11	2032	13.46	27.46	71.31	0.3	44.73	8.42	8.17	1.43	0.15	0.3	5.86	0.54	0.25	21.31	11.76	2.21	3.4	6.15	0.44	4.23	14.17	3.4	2.12	7.0	
1	12	2253	15.54	27.03	74.17	0.4	46.52	11.45	10.96	2.62	0.09	0.84	6.39	1.02	0.49	18.24	13.49	2.26	2.44	7.99	0.44	4.13	13.4	5.15	2.44	5.6	
1	13	923	11.26	27.52	75.73	0.22	46.7	13.11	12.24	2.28	0.11	0.65	6.93	2.28	0.87	18.31	15.71	2.93	5.31	6.72	1.3	4.01	12.24	3.9	2.71	7.5	
1	14	791	18.83	24.4	82.93	0.76	53.86	17.57	17.32	4.93	1.52	5.06	5.56	0.76	0.25	14.92	15.42	4.42	6.07	6.83	0.76	5.56	12.64	6.19	2.78	6	
1	15	1681	10.99	23.22	77.1	0.06	46.22	14.52	14.04	2.08	0.2	1.06	8.92	1.01	0.48	17.07	14.52	2.44	3.99	8.39	0.36	4.46	11.48	4.28	2.22	6.5	

Figure 4.15 – Example of features computed by LIWC

Generating this features' matrix was obviously the most easy part. Much more difficult was, to determine which if any if those were actually relevant, and for what purpose they could be used. To answer the first question, we first checked if an automatic classifier could possibly predict from which book the chapters were taken, knowing the learnt features. Using a linear logistic regression model, we achieved a rate of 66% of correctly classified instances with 10-fold cross-validation (appendix A.5.1), which, despite this figure being quite low, tends to validate the rather natural first hypothesis of the wordings to be consistent amongst each of the books.



#### 4.4. Visualizations Based on Named Entities

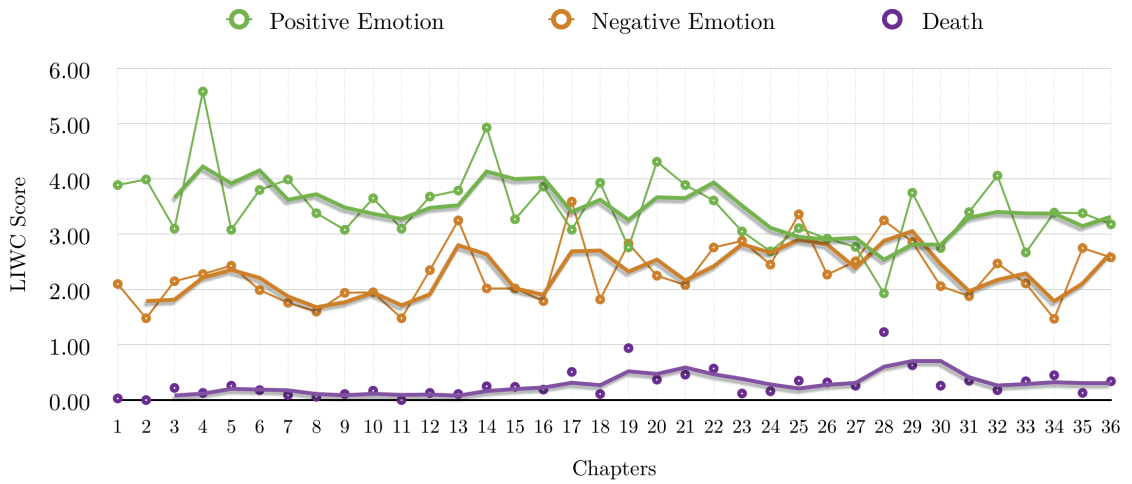


Figure 4.16 – Positive and Negative Emotions Graph on *SH.1*

One direct possible application for graphs using this information could be to quickly identify stylistic breaking points showing possible plot twists, or to isolate the big parts a story is made of. In some cases, it is indeed possible to show a faint correlation between potential plot twists and changes in wording statistics gathered from LIWC. We presented in fig. 4.16 one example where the variations of the difference between "positive emotions" and "negative emotions" words seemed to follow a pattern where the latter become consistently more important than the former during the third quarter of the story (chapters 23 to 30). In this first example, the convergence phenomenon could also be related to an increase of words semantically related to death in the same parts, which may further hint at value in this kind of graphical analysis.

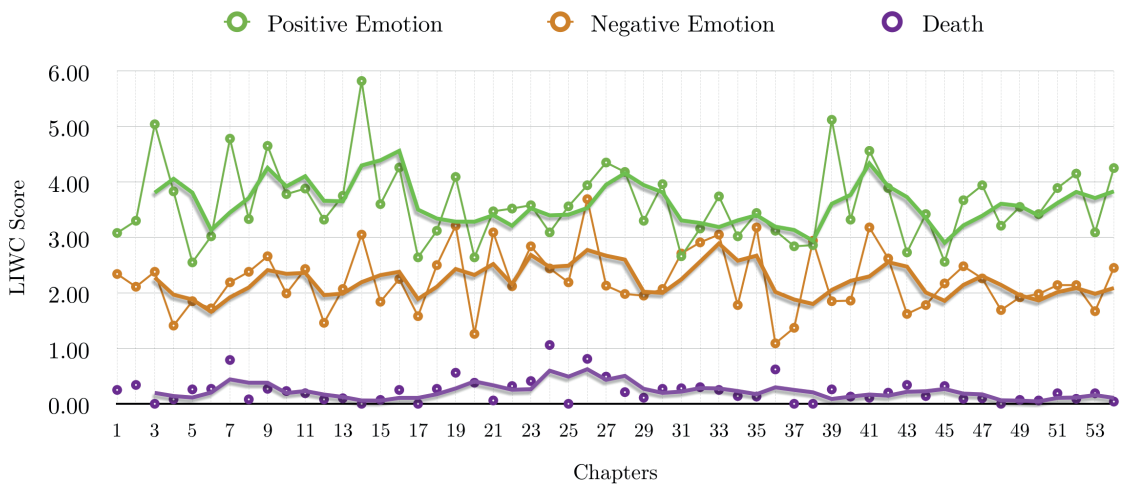


Figure 4.17 – Positive and Negative Emotions Graph on *SH.2*

## Chapter 4. Methods for Digital Writing Tools

However, as shown with another example in fig. 4.17, the same cannot be said of any story. In fact, most of the graphs produced with this method turned out to be impossible to interpret and generally inconclusive, as they did not exhibit patterns that could be unambiguously related to narrative features. For this reason, we chose to disregard emotion graphs from our writing tools.

### 4.4.4 Entries-Exits Graph

The time-oriented pendant of the characters network is the representation of characters entering, staying and exiting between the scenes, thus hinting at narrative continuity and discontinuities along the story. This representation only needs as an input the presence information of the characters in each scene. It echoes to a traditional, theatrical-oriented way of composing stories, where the author constitutes his fragments as scenes that makes a (possibly evolving) set of characters interact inside more or less well-defined place and time unities.

To introduce a first example, we considered the presence of the sole main protagonists, in the eighteen chapters of *SH.10*. This raw information is reproduced in table 4.2, and in itself shows bits of structuration that can be useful to understand the underlying composition rules, like the alternating presence of sets of characters, or the variations in total numbers of them.

Characters \ Chapters	Chapters																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Fumika								×								×		×
Ichiyô												×						×
Kumo						×									×			
Mirafiori	×			×		×	×	×	×			×			×		×	×
Marthe														×				
Max		×	×		×		×	×	×		×		×	×	×	×		×
Shizuko		×	×		×		×	×	×		×	×	×	×		×		×
Total	1	2	2	1	2	2	2	4	3	1	2	3	2	3	3	3	1	5

Table 4.2 – Protagonists presence in the 18 chapters of *SH.10*

We suggest a graphical projection to visualize this information in a way that makes it immediately interpretable, illustrated in fig. 4.18. The main view (top) represents characters with stacked colored rectangles, for each scene. A gray square represents a character that was present in the preceding scene and still is in the considered one, whereas green and red squares denote respectively characters having entered and exited the story at this point. In addition to the stacked bars, a secondary view (bottom) shows the cumulative entries and exits (i.e. the times when a character enters the story for the very first time, or disappeared for

#### 4.4. Visualizations Based on Named Entities

good).



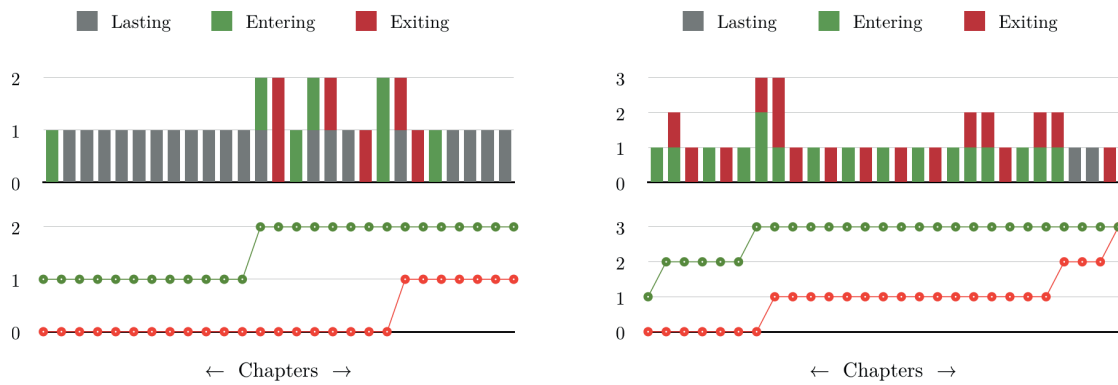
**Figure 4.18** – Entries-Exits graph for *SH.2*.

Top: characters lasting, entering and exiting between each chapters. Bottom: first and last seen curves.

While this view may of course not apply gracefully to any kind of composition, it is not hard to see how it can allow to compare the overall structuration strategies employed by an author following a compatible schema. As we can see, our example allows to easily spot structural parts of the story, such as the progressive introduction of characters in the first six chapters, a narrative break in the middle with one new character introduced and one disappearing almost at the same time, and a progressive resolution towards the end, with alternating point of views focusing one last time separately on different characters. Furthermore, the area between the two lines of the bottom part can be understood as the concurrent number of characters the reader has to keep track of in order to understand the story till the end, which is a potentially interesting variable one may want to play with in a narrative (re-)organization process.

Finally, more than a merely interpretative projection, the Entries-Exits graphs allow to devise typologies following the nature of the narration. For instance, several different narrative strategies can be identified throughout the ten volumes of *La simulation humaine*, and we chose to show two simple examples in fig. 4.19, which were computed from *SH.5* (left) and *SH.6* (right).

## Chapter 4. Methods for Digital Writing Tools



**Figure 4.19** – Comparison of Entries-Exits graph of *SH.5* (left) and *SH.6* (right). Top: characters lasting, entering and exiting between each chapters. Bottom: first and last seen curves.

Just by looking at those graphs, one can immediately see that *SH.5* is about two characters, starting with a focus on the first one, continuing with interleaving episodes and ending focusing on the second one. Conversely, *SH.6* features a very regular pattern from the second third on, strictly alternating between two characters, except when they get reunited towards the end. The latter book could be labelled "*The parallel lives of A and B*", whereas the former one would be "*The life of A, A and B, and B*".

### 4.4.5 Timeline Graphs

We already mentioned Randall Munroe's *Storyline* graph (fig. 4.20), for the coverage it received from various communities. Early works already recognized the broader value of such visualizations transposed to different contexts involving the representation of diverse forms of social interactions, such as genealogy [Kim et al.(2010)], emergence, evolution and dissolution of community structures [Reda et al.(2011)], or even cooperations between software developers [Ogawa and Ma(2010)].

Later works pushed the model further, designing composition rules to improve its legibility and making automated generation possible [Tanahashi and Kwan-Liu Ma(2012)], showed the relevant contribution of real time user interaction to the model [Shixia Liu et al.(2013), Shejuti et al.(2015)], and discussed its possible use as embedded within literacy analysis tools [Akyigit et al.(2014)]. Lately, more contributions used *storyline* representations to visualize automatically extracted interactions between the characters of movie screenplays [Agarwal et al.(2014)] and TV series [Tapaswi et al.(2014)].

#### 4.4. Visualizations Based on Named Entities

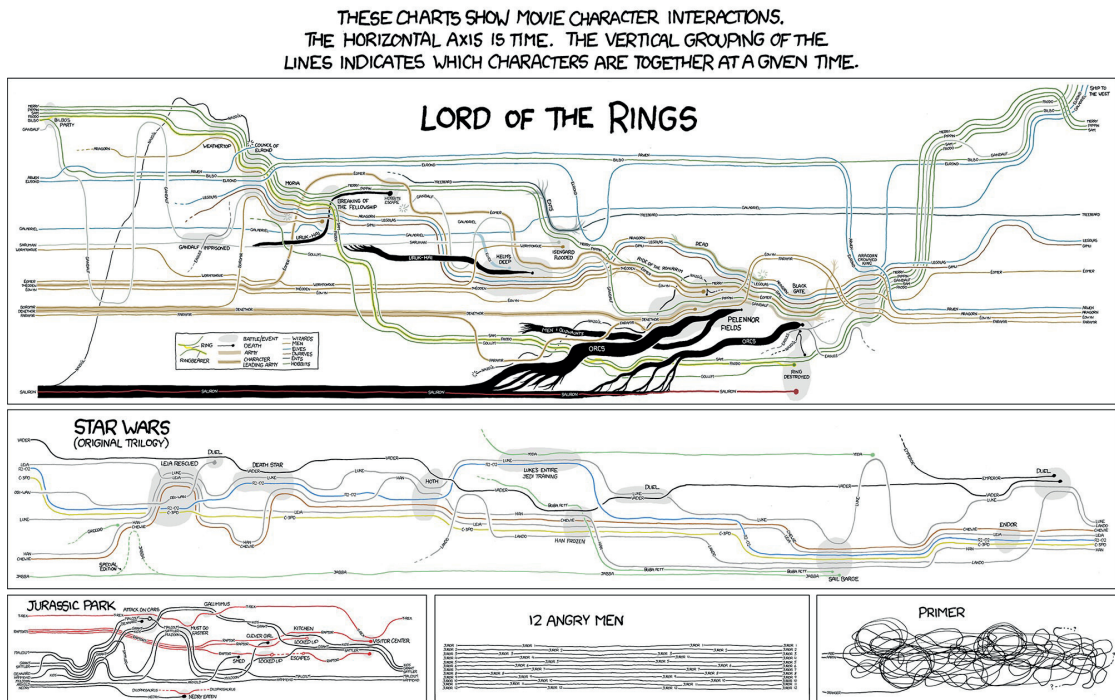


Figure 4.20 – Original xkcd storyline by Randall Munroe (2009, Creative Commons 2.5).

Of course, following their definition we can expect *storylines* to be easily derivable from full books given a reasonably precise named entities extraction. We implemented a simple tool interfacing the extracted results presented in section 4.3 with a d3.js implementation of *storyline* graphs<sup>17</sup> to evaluate this hypothesis on our test datasets, with rather convincing results (see figs. 4.21 and 4.22 for two resulting graphs).

<sup>17</sup>Nancy Iskander et al. (Computer Science Club, University of Waterloo, Canada) - [http://csclub.uwaterloo.ca/~n2iskand/?page\\_id=13](http://csclub.uwaterloo.ca/~n2iskand/?page_id=13)

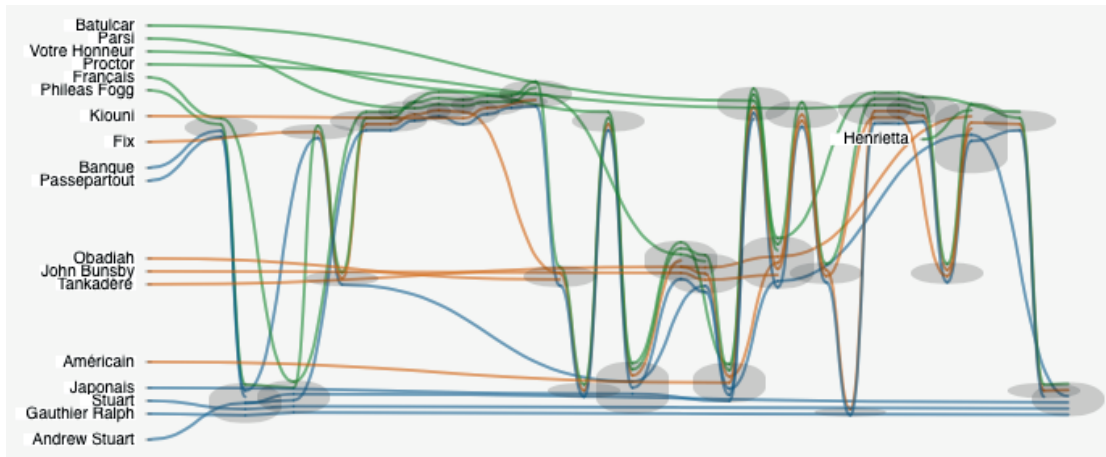


Figure 4.21 – Storyline visualization automatically generated with *Around the World in Eighty Days*

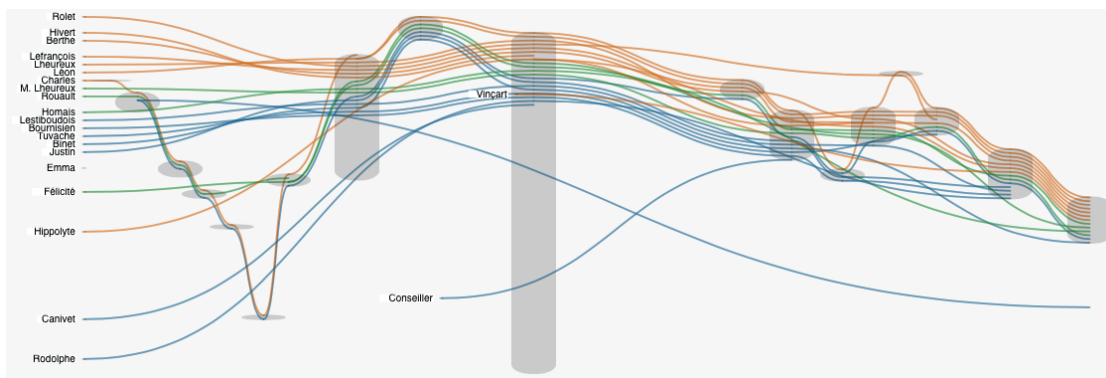
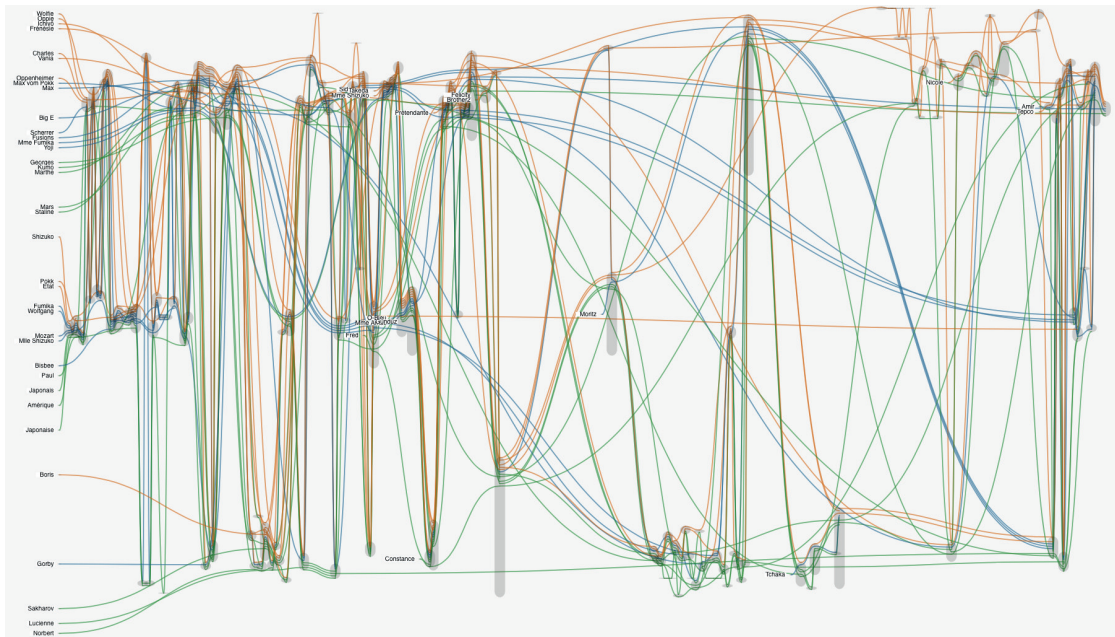


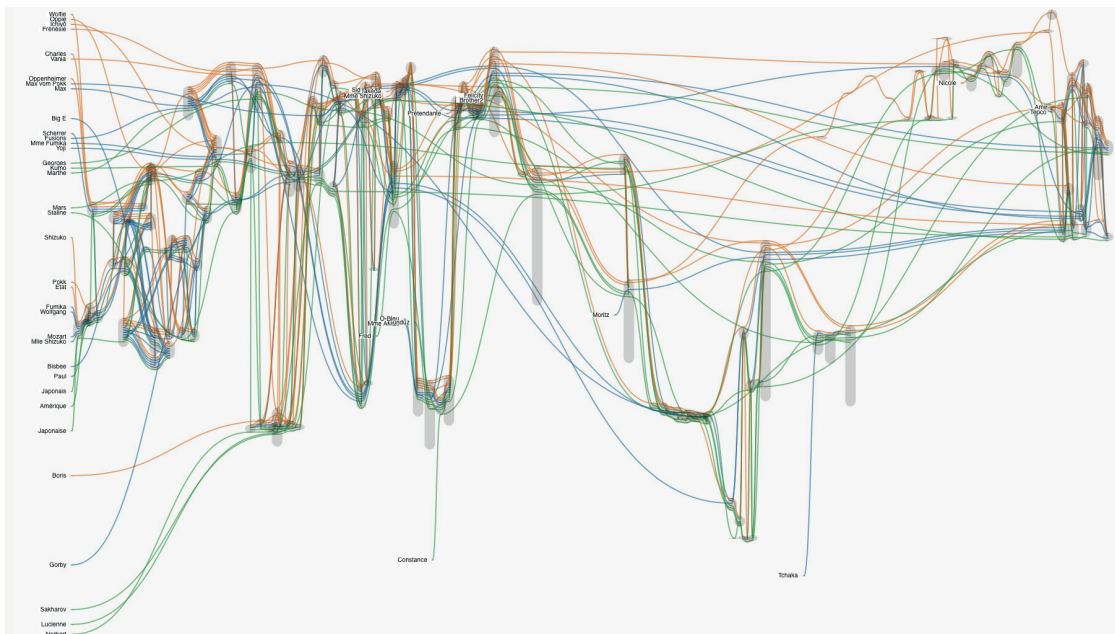
Figure 4.22 – Storyline visualization automatically generated with *Madame Bovary*

However, as some pointed out [John et al.(2016)], this kind of representation does not scale well to long and intricate stories, and comes with design constraints that make it difficult to display additional narrative information potentially useful to a novelist. Figure 4.23 shows an example of such a graph on *La simulation humaine*, illustrating that the multiplicity of scarcely recurring characters and scenes tend to make trajectories untidy and arduous to read, as the constraints imposed by both vertical and horizontal spacing end up creating very wide spaces between the scenes. This can partly be improved through careful manual reorganization of the scenes (fig. 4.24), which however may break reading assumptions, such as events chronology.

#### 4.4. Visualizations Based on Named Entities



**Figure 4.23** – Storyline visualization automatically generated out of characters and places extracted from *La simulation humaine*



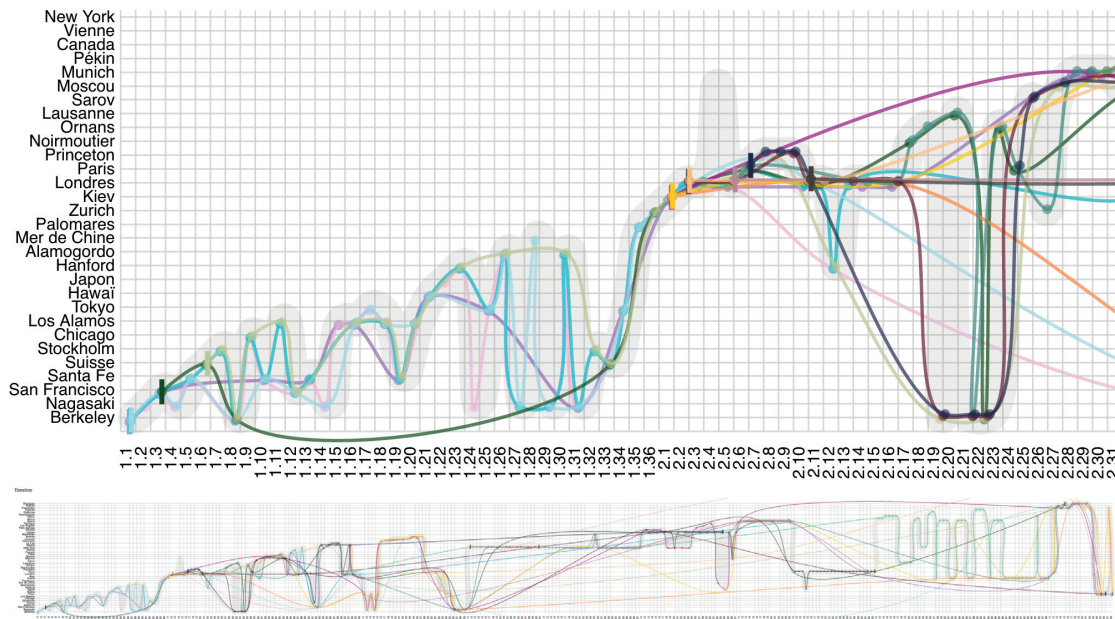
**Figure 4.24** – Manually reordered Storyline visualization of *La simulation humaine*

Furthermore, these visualizations don't deal well with spacio-temporality in representations, as events can in theory freely be positioned on the  $x$  axis, and give no information about the

## Chapter 4. Methods for Digital Writing Tools

locations where they take place. We addressed those concerns by proposing a slightly different interpretation of this visualization, that considers text fragments as event units, and possibly ties them to locations when the information is available. Of course, we want to be able to dynamically generate this graph from our processed text source, and quickly enough for the user to see effects of compositional changes in real time.

The  $y$  axis will list all the possible locations, in the order in which they were for the first time nominally mentioned in the story (from bottom to top). The  $x$  axis represents the fragments, in the order of the story, from left to right. An example of such a graph is shown in fig. 4.25. Like with *storylines*, each character is assigned one single curve that spans from the moment of its first to its last mention in the story, and can be given a color for better identification. To improve the legibility, a small vertical line is added at the start and at the end of each curve, and mentions in chapters are denoted by bulges on the line. For aesthetic reasons, the lines are smoothed using a cubic Catmull–Rom spline [Catmull and Rom(1974)] with an easing parameter set to 0.7.



**Figure 4.25** – Timeline visualization of *La simulation humaine* (Top: detail showing the first two volumes / Bottom: global view on 298 chapters)

Compared to the *storyline* visualization we tested before, this timeline allows for a more compact view in the vertical dimension, that avoids the need of two-dimensional scrolling to display all details, and eases the following of one particular character while factoring in spatial information. An overall view of the full representation of *La simulation humaine* is shown in



fig. 4.25 for comparison.

Again, this representation allows to visually explore storytelling typologies. The typical adventure story will be expected to show a group of characters exploring a variety of places together, resulting in a diagonal organization. An example of such a graph is shown in fig. 4.26, with Jules Verne's *Around the World in 80 Days*. Notice the two distinct trajectories between characters that follow Phileas Fogg's and Passepartout's journey around the world before coming back to England, whereas Andrew Stuart and John Sullivan, members of the Reform Club staying in London, follow a flat line.

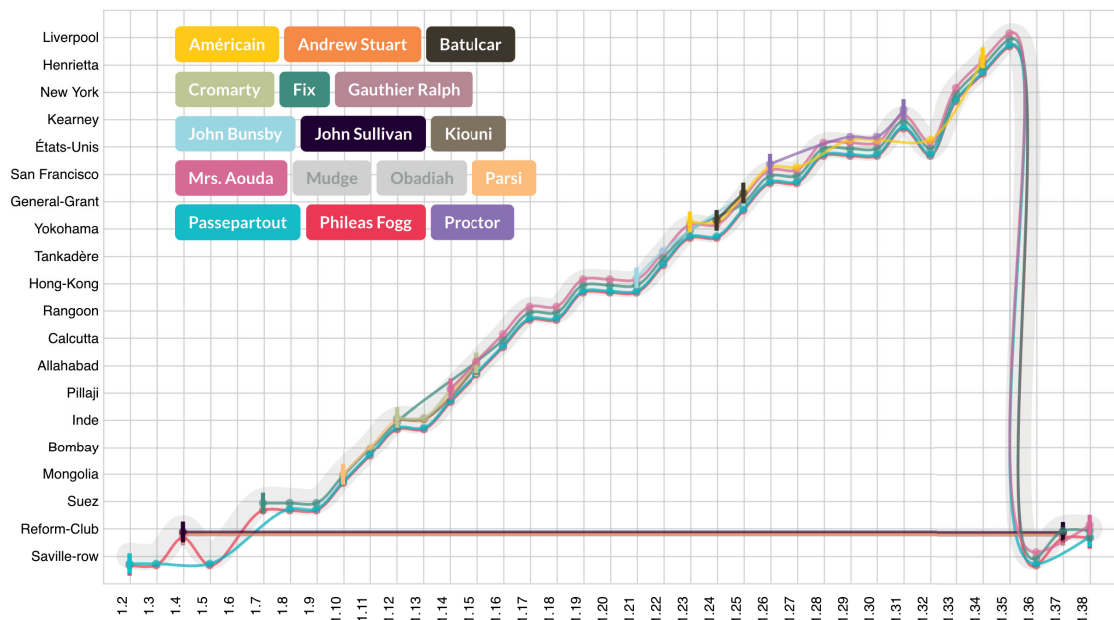


Figure 4.26 – Timeline visualization of *Around the World in Eighty Days*

On the contrary, social stories happening in recurring places will exhibit flatter shapes with lines crossing as the characters interact. An example of such a graph is shown in fig. 4.27, with Gustave Flaubert's *Madame Bovary*. Here we clearly see how the story suddenly gains in complexity of interactions starting from the second tome, and the interleaving lines between characters that stay in Yonville and those who frequently travel to Rouen. This example also illustrates one weakness of this representation: as the vertical axis is ordered chronologically and not according to geographical proximity, long lines can appear between places that are actually nearby, or lead to crossings when the locations were supposed to be aligned.

Chapter 4. Methods for Digital Writing Tools

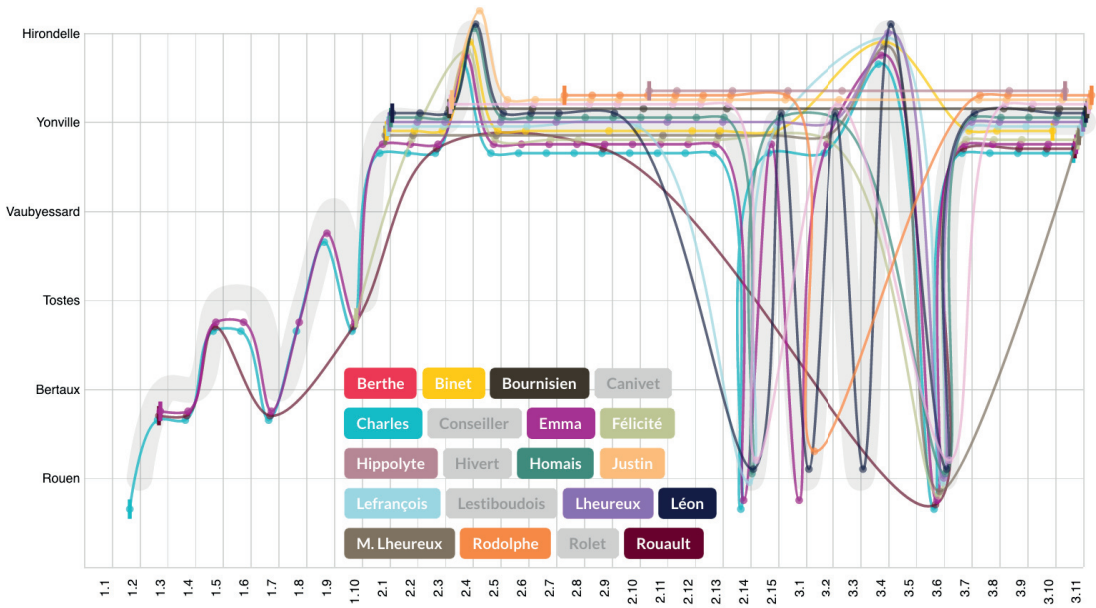


Figure 4.27 – Timeline visualization of *Madame Bovary*

# Document 3

## Daniel de Roulet, 22.02.2016

### Les outils dont je pourrais avoir besoin

En travaillant sur le corpus de la SH j'ai utilisé des outils « manuels ». Ce sont principalement des fiches et des schémas.

du corpus. Ils doivent donc se retrouver dans la nomenclature

#### Exemple de fiches : la nomenclature

Elle regroupait :

- Des noms d'entreprises fictionnelles (Greenwar, Bleu Siècle...)
- Des noms de lieux où avait lieu l'action (Davos, Fukushima...)
- Des noms de personnages illustres (Reagan, Oppenheimer...)
- Des noms de personnages de fiction (Max, Shizuko...)

Cette nomenclature (sous sa forme raccourcie) a ensuite été utilisée pour les mots qui devaient être surlignés pour permettre au lecteur qui entre dans le corpus numérisé de manière aléatoire de s'y retrouver.

Cette nomenclature a aussi été utilisée pour les personnages à point de vue. Il s'agit des personnages romanesques selon lesquels un chapitre est à chaque fois raconté. Il y en a 17 sur l'ensemble

**Exemple de schéma** : le schéma des relations entre les personnages à point de vue

Ces relations (cf. annexe) sont de deux sortes : simples (lien de parenté, lien de proximité) ou amoureuses.

#### Remarques sur le document : graphe des liaisons

La visibilité de l'ensemble est excellente et donne une idée de la complexité du tissu romanesque. Un tel outil ne pourrait pas être produit manuellement par le romancier.

Pour en faire un outil encore plus utile au romancier, je suggère :

- Il faut pouvoir unifier les noms qui désignent le même personnage (Exemple : Joseph = Joseph Riclette ; Wolfie = Wolfgang = Steinamhirsch ; Fumika = Madame Fumika)
- Il faut pouvoir enlever des personnages qui n'en sont pas (Exemple : Urbain) ou des

### Document 3

---

personnages qui sont des objets ou des animaux, même si le romancier s'est adressé à eux comme à de vrais personnages (Exemple : Cheval, Superphenix)

- Il faut pouvoir réduire les nœuds du graphe et éventuellement le réduire jusqu'à ne laisser que les 17 personnages à point de vue. Alors on pourra comparer le schéma dont le romancier s'est servi (liaisons simples et liaisons amoureuses) pour avoir de nouvelles indications. On s'apercevra sans doute alors que les persos Frénésie et Work sont extérieurs, sans liens avec les personnages plus centraux, ce qui serait une indication utile pour la construction de nouvelles intrigues

- 9 sont des personnages qui n'apparaissent pas dans la nomenclature, il faudrait les y rajouter)
- 2 sont des doublons (Wolfie = Wolfgang ; O-bleu = Mirafiori)
- 8 sont des erreurs

Dans le graphe à 20 occurrences :

- Côté lieux : il y en a 38 au lieu de 12, mais 8 posent problème
- Côté personnages : sur un total de 72 (au lieu de 42) 7 erreurs de plus soit 15 en tout, ce qui ne diminue pas l'intérêt

---

#### Remarques sur le document : graphe des déplacements

Il permet au romancier de se faire une idée des déplacements de ses personnages, mais aussi des retours sur les mêmes lieux à travers tout le corpus. Un tel outil pourrait être fait manuellement mais jamais avec cette précision.

Dans le graphe à 50 occurrences :

- Côté lieux : il n'y a qu'une seule erreur sur 12 lieux principaux, le dernier : Prétendante est un personnage
- Côté personnages : il y en a 42 :
- 15 sont des personnages à point de vue (il en manque donc 2)
- 8 apparaissent dans la nomenclature comme personnages

Pour en faire un outil encore plus utile au romancier, je suggère :

- Qu'on puisse exclure les noms de lieux ou de personnages qui n'en sont pas pour ne considérer que les déplacements réels. Le romancier pourrait vouloir vérifier que dans son corpus chaque tome a ou n'a pas le même genre de pattern. Et pouvoir ainsi comprendre qui sont ses personnages mobiles et lesquels sont les plus « tranquilles ».
- Une représentation géographique en plaçant les lieux et les parcours sur une mappemonde serait appréciée. (cf. annexe Bourdieu sur l'Education sentimentale de Flaubert)

Autre vœux : faire le même exercice sur un corpus numérisé comprenant de nombreux personnages, genre Balzac.

# 5 Introducing a Novel Editor for Authors: *Saga+*

## 5.1 State of the Art

### 5.1.1 Text Visualization for Literary Scholars

Graphical interfaces visualizing features of textual corpora have been used for many different purposes, ranging from the analysis of consumer reviews from *TripAdvisor*<sup>1</sup> [Wu et al.(2010)] to exploring scientific publications [Erten et al.(2004)]. Set apart the plethora of tag cloud generators and their derivatives, several earlier works explored ways of empowering literary scholars with easy access to artificial intelligence, statistics and visualization tools, for instance for documents classification purposes [Plaisant et al.(2006)]. Closer to us, *VarifocalReader* was developed with the aim of easing large text documents analyses through a combination of automatic topic segmentation and different visualizations strategies [Koch et al.(2014)].

Several online tools aiming to bring books visualization to literary scholars have also been made available recently. Among them, *WordSeer*<sup>2</sup> features navigation through large text collections with the help of similarities lookup and pattern extraction [Muralidharan and Hearst(2011), Muralidharan and Hearst(2013)]. In 2016, Bilenko et al. released a simple tool combining sentiment analysis and a chord diagram showing cooccurrences of the 20 most common characters<sup>3</sup>, and John et al. [John et al.(2016)] presented a very complete web application<sup>4</sup>, including tools similar to the ones we previously described, such as automatic named entity extraction, word clouds, characters networks graphs and storyline visualization.

---

<sup>1</sup>TripAdvisor - <https://www.tripadvisor.com>

<sup>2</sup>WordSeer | A Text Analysis Environment for Humanities Scholars - <http://wordseer.berkeley.edu>

<sup>3</sup>[https://github.com/nbilenko/narrative\\_explorer](https://github.com/nbilenko/narrative_explorer)

<sup>4</sup>Visual Analytics for Narrative Text - <http://textvis.visualdataweb.org>

### 5.1.2 Connected Editing and SaaS

The soaring development of Internet made possible a wide range of connected tools within the last decade [Turner et al.(2003)], which in turn opened the door to new business models, such as *Software as a service (SaaS)*, a mode in which computer software are licensed in the form of a subscription and made accessible only through centralized services. Those tools are widely appreciated today and often preferred to their offline counterparts, mainly for convenience and costs reasons [Gilleo and Lind(2016)]. Blogging platforms such as Wordpress.com<sup>5</sup>, but also more specialized web content management tools such as Squarespace<sup>6</sup>, Weebly<sup>7</sup>, Wix.com<sup>8</sup> or Jimdo<sup>9</sup>, to name only a few, all claim dozens of millions of active sites and content publishing users.

Within the last decade, the SaaS model quickly spread to almost all software businesses, and in particular to any domain where users need to process digital data. In those areas, its advantages are obvious not only to access information but also as the tools used for visualization more and more benefit from a web-based approach, allowing for lower maintenance costs, wider compatibility ranges and faster deployment cycles [Mwalongo et al.(2016)]. Closer to traditional book publishing, one can for instance notably name *PressBook*<sup>10</sup> as an innovative web service that allows to edit or import chapters of a book, and export them as a formatted PDF, or immediately publish them online. Following that report, making our solution fully connected was a primer.

## 5.2 Overview

The main purpose of *Saga+* is to edit possibly very long novels or sagas, around the *fragment organization* level. Regular text editors follow a logic that operates at the syntactic level with at most a couple of tools revealing the overall structure (as for instance, automatic summarization or generation of table of contents). On the contrary, *Saga+* is in concept closer related to video editing, focusing on leveraging interaction at a degree of abstraction where finding the right order to present the different parts of a story becomes a core task of the narrative strategy.

---

<sup>5</sup><http://www.wordpress.com>

<sup>6</sup><http://www.squarespace.com>

<sup>7</sup><http://www.weebly.com>

<sup>8</sup><http://www.wix.com>

<sup>9</sup><http://www.jimdo.com>

<sup>10</sup><https://www.pressbooks.com>

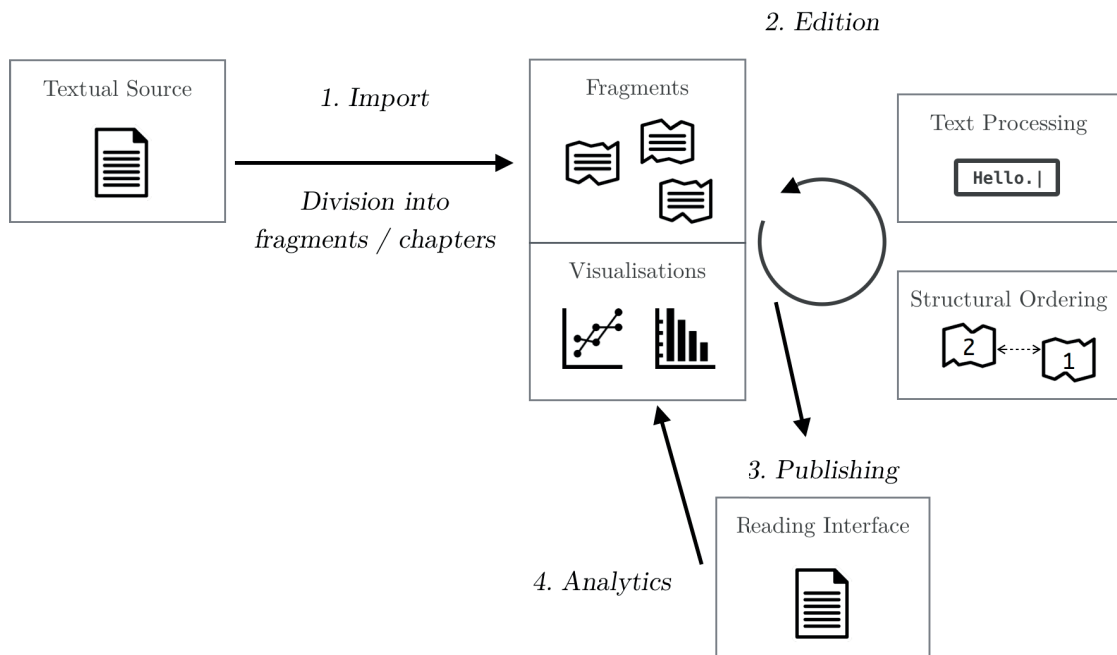


Figure 5.1 – Text editing interface

The tool proposes the following workflow (illustrated in fig. 5.1):

1. **Import** (optional) (see section 5.3.1). At the beginning of the typical workflow, *Saga+* allows to import one's own preexisting books or to load precomputed samples from famous public domain books (this step is of course optional since it is also possible to start a new book from scratch). When used, the import procedure performs a first automated analysis of the text and creates fragments from its chapters.
2. **Edition** (see sections 5.3.2 and 5.3.4). *Saga+* fosters experimenting with structural reorganization through several real time graphical feedbacks. The actual text can at in the same process be edited on a per-fragment basis to make details fit the adjusted context.
3. **Publishing** on the web reader (see section 5.3.6). *Saga+* requires little effort to make the texts available to the public. The produced digital books are listed in an online catalog and can be read directly from the website, on a computer, tablet or mobile phone.
4. **Statistics** collection and analysis (see section 5.3.7). Like on any website, detailed information about the visitors get logged in order to pull usage statistics. However, in *Saga+* the reading analytics for each published books are made visible to their author, for them to know how their works get read.

## 5.3 Main Features

### 5.3.1 Importing Existing Books

For efficiency and simplicity reasons, we implemented an import routine accepting an input of books or series of books of any size in the form of plain text files. The system uses an heuristic process to separate the chapters, provided the text is reasonably well formatted.

Given a plain text file with or without formatting line breaks, each line is read and screened for clues hinting at a chapter separation mark. A line is considered a chapter separation, if it was preceded by at least two carriage returns, and one of the following conditions is true:

- it contains the mention "*chapitre*" followed by a number, or
- it starts with a number, followed or no by a separating period mark, and a title

To be considered valid, a chapter number must be either written in Arabic or Roman numerals, and have a plausible nominal value.

Once we found the candidate splitting points, we run a consistency check to remove as many mistakes as possible. The rules that need to be ensured are:

- if numbered, the chapters need to be numbered consecutively
- a chapter must not be empty

This additional process allows to detect possible parsing errors, like peculiarly formatted quotations or tables of contents that might be mistaken for chapter titles.

### 5.3.2 Entities Manager

The named entities extraction is done using a method we designed to get a good compromise between efficiency and robustness on fictional French corpora (section 4.3). Following this method, characters and places are automatically tagged, extracted and classified at the first time a new book is loaded. This process takes between a few seconds to a few minutes for very large chunks of texts, but only needs to be run once.

To account for detection errors and the possible will of the user to customize this structure, the classified entities are stored away in a database table tracking the initial results and possible updates. For the subsequent graphical representations, each entity is assigned a



default color chosen among a palette of 25 colors predefined in order to provide maximal visual differentiation, which has been shown to be enough for immediate and unambiguous identification [Carter and Carter(1982), Green-Armytage(2010)]. Of course, the user is always given the possibility to reassign different colors, or to turn off coloration in order to focus on the more important entities.

All these settings are grouped in one single interface, depicted in fig. 5.2. We also added an histogram visualization and textual excerpts of the sentences containing the mentions, to help the identification process when manually creating new entities.

The screenshot shows a window titled "Détails sur l'entité" for the character "Fumika". At the top, there are tabs for "Personnage" (selected), "Lieu", and "Autre", with "ou" between them. A "Supprimer" button is in the top right. Below the tabs is a "Couleur" palette with 25 color swatches. To the right is a "Mémo" text input field. Further right is a "Coréférences" section with a list containing "Mme Fumika" and a "+ Ajouter..." button with an "OK" button. Below this is an histogram showing the frequency of mentions across a range of values. At the bottom is a "Citations" table with three rows.

Chapitre	Extrait
1.1	Depuis qu'elle a rencontré Wolfgang, Fumika ne peut plus se l'arracher du cœur.
1.1	Au réveil, son œil droit reste plus bridé que le gauche. Toi, Fumika, se dit-elle, tu es trop seule. La Californie et Berkeley ne te réussissent pas.
1.1	un jour, quand cette nouvelle guerre mondiale aura pris fin, d'autres jeunes Japonaises - et pas seulement Fumika - veulent émigrer aux États-Unis. Il faut qu'elles cachent dans ce pays, tout est trop grand.

Figure 5.2 – *Saga+* entities manager (here displaying the character *Fumika* in *La simulation humaine*).

### 5.3.3 Montage and Reorganization Interface

Most interfaces dedicated to writing books allow to split the text into chapters, and to reorder them. However, they usually assume the authors know their contents in enough details to do this without external help, or are following an external plan in order to do so. When designing *Saga+*, we wanted the reorganization interface to be at the core of the editing process. In order to do so, each chapter is represented inside the story as a tile featuring the colors of its characters, and the main location it is set in (fig. 5.3). The tiles can be freely sorted and organized in a sequence of volumes, or moved aside and stored for further use.



Figure 5.3 – Structure montage interface (left) and layout of a chapter tile (right)

Along with this control we displayed dynamic web implementations of the visualization tools we described in the previous chapter, namely the social network of characters (section 4.4.2), the entries-exit graph (section 4.4.4), and of course the timeline graph of the story (section 4.4.5). Acting as a *distant editing* dashboard, the whole interface has been designed to provide real time feedback on any change the user may want to experiment with. Any update in the text is immediately reflected on the corresponding tile, and any structural reordering is propagated to the visualizations.

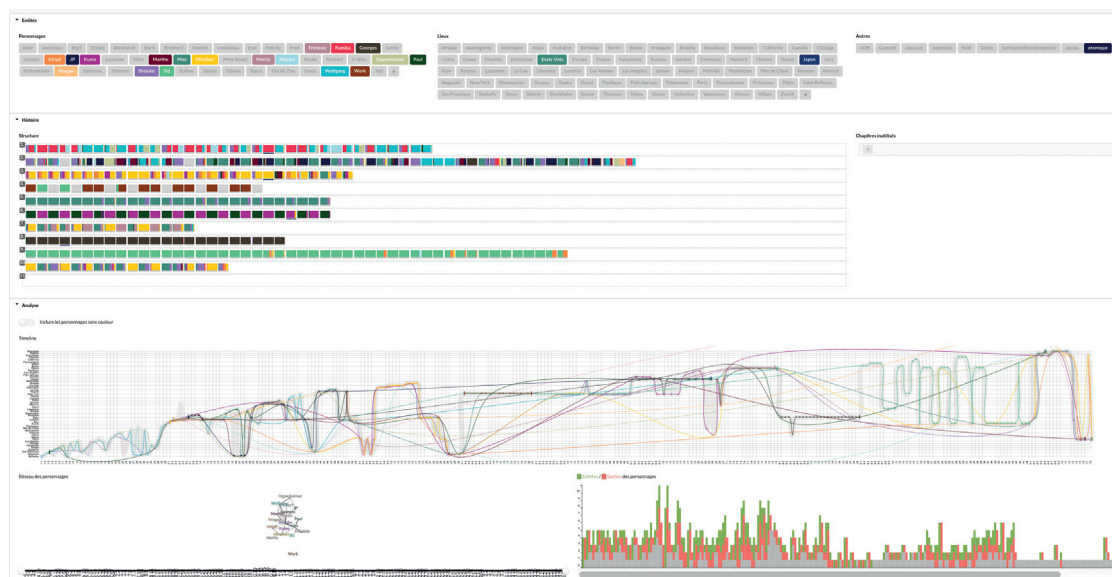


Figure 5.4 – *Saga+* montage and reorganization interface (featuring *La simulation humaine*)

### 5.3.4 Editing Features

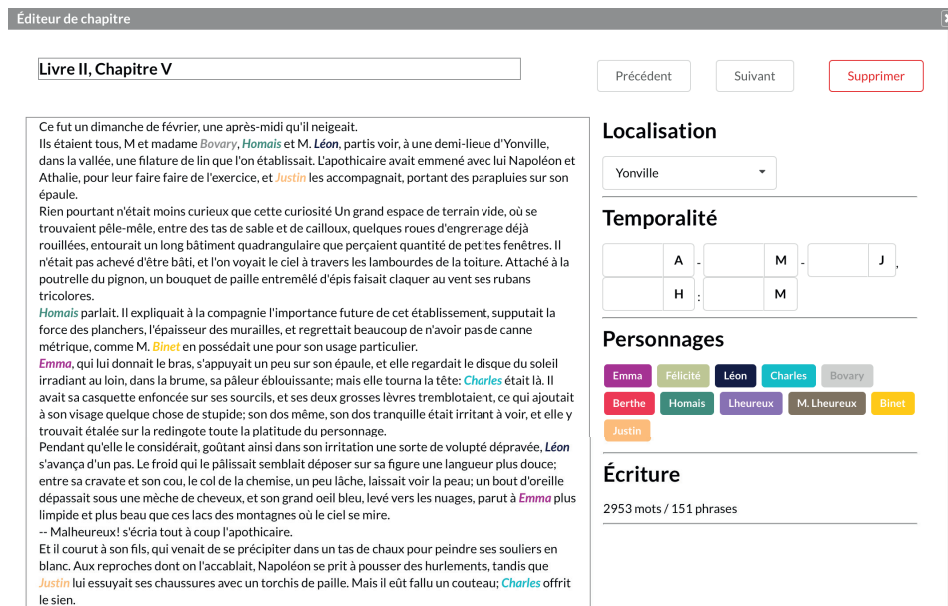


Figure 5.5 – Saga+ text editing interface

Saga+ provides a simple text processor interface that can be reached by a click on any chapter. Shown in fig. 5.5, this editor features two main parts:

- The left part of the screen, display the title and full text of the chapter, with all named entities labelled in colors automatically
- On the right, a couple of fields allow to fill in metadata about the chapter, such as its localization and a time marker. Below, the all relevant characters are shown with their colors, and a small space is reserved for writing statistics, such as the count of words and sentences. At the top of the screen, three buttons allow to navigate to the preceding and next chapters, or to delete this part altogether.

To account for possible network errors unavoidable with web applications, any change made in this interface is automatically sent back to the server and saved, without the need for the user having to worry about it.

### 5.3.5 Versioning

One couldn't imagine a creative process involving a lot of trial and error without an easy way to keep safe different stages of the work, in order to come back later to an earlier version

## Chapter 5. Introducing a Novel Editor for Authors: *Saga+*

or to be able to compare them before deciding on the best one. *Saga+* was designed with versioning in mind and can thus gracefully handle as many simultaneous versions as wanted. The versioning manager interface (shown in fig. 5.6) conveniently lists all existing versions in the same window, allowing to delete or to reopen any of them. A new version is created by saving the work state of the currently opened variant, and automatically numbered by increasing the number of the original by one unit.

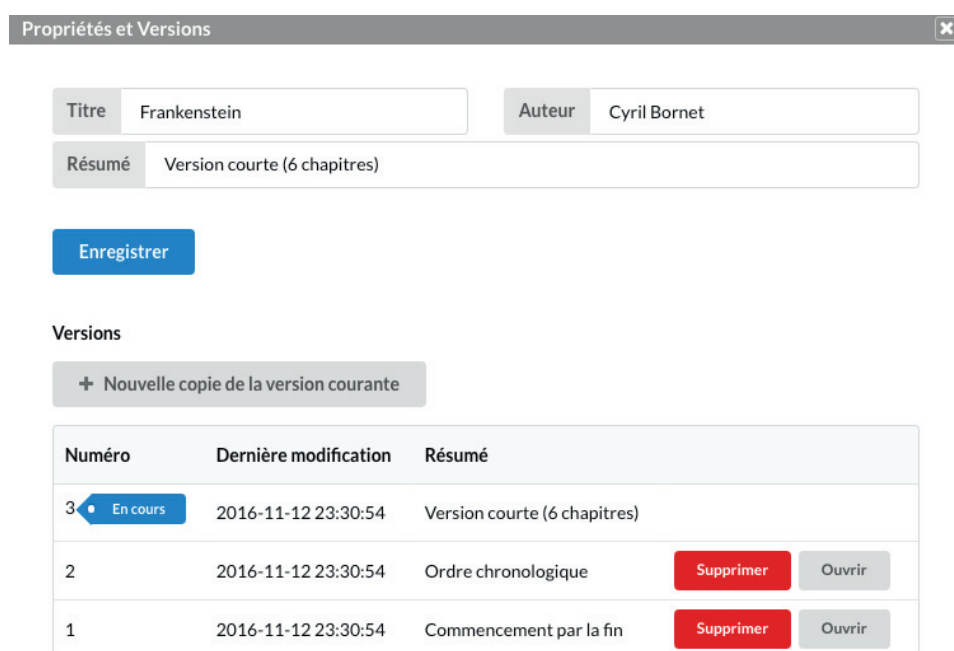


Figure 5.6 – *Saga+* versioning manager window

### 5.3.6 Reading Interface

The produced works can be made readily available on the Internet through an interface tailored for reading online or on a mobile device, including everything we learnt from building our first digital saga reading interface. Here too, the best possible immersion experience is guaranteed by intendedly keeping the page very plain, featuring no additional documents or links to external resources, nor advertisement.

To render a homepage effect following the one of *simulationhumaine.com*, a publication assistant module asks need to fill in a short description and an optional icon. Navigation principles are similar as those described in section 3.4.2. If comments were left in the Entities' cards, they are exported and displayed in the glossary mode.

Optimal reading conditions are achieved by using:

- a serifed typeface [Merriam-Webster, Inc(1998)] with a middle to large font size;
- lines layout with lengths set to around 60 characters [Craig et al.(2006)], which are achieved with the help of on the fly automatic hyphenation<sup>11</sup>.

### 5.3.7 Feedback Loop and Optimization

Having shown the importance of reading analytics to understand the effectiveness of composition choices, we added this feedback information directly into *Saga+*. Collected reading statistics can immediately be displayed in the editing interface, and show insights about reading, such as how many times each chapter has been displayed or the mean time spent on each of them.

## 5.4 Technical specifications and software licensing

**Backend application:** custom code written in PHP<sup>12</sup> with partial support of the following frameworks: Laravel<sup>13</sup>, *\_db*<sup>14</sup> (GNU Lesser General Public License).

**Frontend application:** custom code written HTML5<sup>15</sup> and Javascript<sup>16</sup>, making use of the following frameworks and libraries: *jQuery*<sup>17</sup> and *jQuery UI*<sup>18</sup> (both MIT license under copyright from the jQuery Foundation), *jQuery Viewport*<sup>19</sup> by Mika Tuupola (MIT license), *jQuery ScrollTo*<sup>20</sup> by Ariel Flesler (MIT license), *jQuery throttle / debounce*<sup>21</sup> by "Cowboy" Ben Alman (MIT and GPL license), *jQuery appear*<sup>22</sup> by Andrey Sidorov (MIT license), *Waypoints*<sup>23</sup> by Caleb Troughton (MIT and GPL license), *d3.js*<sup>24</sup> by Mike Bostock, *d3 Fisheye Distortion*<sup>25</sup>, *AlertifyJS*<sup>26</sup>

---

<sup>11</sup>Hyphenator - <https://github.com/mnater/Hyphenator>

<sup>12</sup><http://www.php.net>

<sup>13</sup><http://www.laravel.com>

<sup>14</sup>[https://github.com/cyrilbornet/\\_db](https://github.com/cyrilbornet/_db)

<sup>15</sup><https://www.w3.org/TR/html5/>

<sup>16</sup><https://www.w3.org/standards/webdesign/script>

<sup>17</sup><http://jquery.com>

<sup>18</sup><http://jqueryui.com>

<sup>19</sup><http://www.appelsiini.net/projects/viewport>

<sup>20</sup><http://flesler.blogspot.com>

<sup>21</sup><http://benalman.com/projects/jquery-throttle-debounce-plugin>

<sup>22</sup><https://github.com/morr/jquery.appear>

<sup>23</sup><https://github.com/imakewebthings/waypoints>

<sup>24</sup><https://d3js.org>

<sup>25</sup><http://bost.ocks.org/mike/fisheye>

<sup>26</sup><http://alertifyjs.com>

## Chapter 5. Introducing a Novel Editor for Authors: *Saga+*

---

by Mohammad Younes (MIT license) and *Hyphenator*<sup>27</sup> by Mathias Nater (MIT license). The GUI was implemented using custom SCSS<sup>28</sup> and the Semantic UI framework<sup>29</sup>. The *Saga+* icon was designed from a quill feather image by an unknown author, given permission of use for non-commercial projects by cliparts.co<sup>30</sup>, and with the *Silver BellyButton Ring* font by Emily Spadoni.

The complete code source of *Saga+* has been released on GitHub<sup>31</sup>.

---

<sup>27</sup><https://github.com/mnater/Hyphenator>

<sup>28</sup><http://sass-lang.org>

<sup>29</sup><http://semantic-ui.com>

<sup>30</sup><http://cliparts.co/clipart/3593416>

<sup>31</sup><https://github.com/dhlab-epfl/saga>

## 6 Results and Discussion

### 6.1 Influence on the Author(s)

We already presented throughout the preceding chapters scattered bits of mutual influences between researcher and novelist. Focusing on Daniel de Roulet, the reported influences can be summarized as following:

- **Distantiation from the text:** the author several time mentioned the benefits of visualizing his texts using new methods, as a way to get a new perspective on his own works, which in turn allowed for the discovery of literary features that were different from expected. As from appendix A.3.1:

*Grâce à l'outil de recomposition et de lecture mis à disposition par le DHLAB, l'auteur a pu se mettre dans la situation d'un lecteur zappeur, extérieur à ses propres textes. C'est ainsi que des épisodes narratifs qui paraissaient fondamentaux ont pu être considérés comme secondaires et vice versa.*

Similarly, he admitted that this process gave new values to old critics whose relevance may have been misunderstood in the first place (Document 1):

*En devenant le lecteur distancé de mes textes, j'en ai vu les défauts. Je comprenais seulement maintenant les critiques intelligentes de mes lecteurs d'alors, remarques aimables ou dures que j'avais balayées alors par le mépris que je croyais dû aux lecteurs inattentifs aux profondeurs de mes phrases.*

The parallels with *distant editing* are far from merely semantic. Of course, distantiation can be considered in two complementary dimensions, those being emotional, where the one wants to consider the text as an external observer to let go of preconceived ideas,

or knowledge-wise, zooming out to have a better oversight of the global structuration and gather new insights. Yet both directions can indeed be explored with *distant editing*, with the tools we designed centered around transversal and fact-driven approaches to the text.

- **Methodological:** as the produced visualization tools were released to the author, he needed to find a way to incorporate this information in the writing process, changing his way of working with text. He describes it (appendix A.3.1) as a bilateral process:

*L'auteur s'est familiarisé avec toutes sortes de d'outils numériques en les mettant directement en œuvre après y avoir été initié par le DHLAB.*

and:

*Le DHLAB a pu, à chaque pas, se confronter au jugement de l'auteur sur la pertinence littéraire des propositions faites.*

As one may expect from our research methodology, both phenomena were indeed interdependent. For instance, the visualizations that we presented originated and were perfected within the continuous feedback results gathered from working with the author. As a result, even if the released tools could definitely be perfected in many ways, they show that the participative action research methodology is a valid approach to tackle problems of this nature, and could probably be taken on as is for further improvements or generalization works.

- **Writings:** The released digital saga and *Le démantèlement du cœur* both are witnesses that carry marks of this research. As a very notable example, our early attempts in focalization modeling influenced the composition strategies the author used in *Le démantèlement du cœur* (several steps of the structural evolution of this novel before its paper publication are documented in appendix A.3.1). About that, he commented:

*À chaque fois qu'un cul-de-sac fécond a été exploré, des outils utilisés et des résultats partiels ont pu être récupérés comme bénéfices collatéraux. Ainsi l'auteur a utilisé les mesures de la tension narrative pour le tome 10 et le DHLAB a mis au point une méthode d'analyse du passage du témoin narratif d'un personnage à l'autre.*

Before discussing the limitations and perspectives of this thesis, let us consider one last look back at our initial research questions, by asking Daniel de Roulet to directly answer them according to his own experience of our common work (December 2016):



### **Is it possible for an author to edit better novels out of already published ones, given the access to adapted tools?**

*Usually the tools available to writers are basic. I use, for example, the one that allows you to count characters when I must format or reformat a text for a newspaper that requires a text of a very precise length because of the requirements of the publication. Or the automatic spell checker that I cannot do without. In the case of more sophisticated programs, syntactic analysis and stylistic (Antidote<sup>1</sup> for example) I use up to 50% of the proposed features and already then I realize that some features are useless for my way of writing. For example, the tool that detects passive turns is of no help to me.*

*The tools that make it possible to detect for example the main character of a chapter or the place where the action happens will allow me to reorder the narration. But when the narrative voice is concealed, I realized that the tool should not be used in the same way. It seems to me therefore important that the new tools can be parameterized for the writing of each one. I see great potential in this direction.*

*The authors of literary works, unlike those who write scientific or technical works, all have a precise idea of what an optimal text is, but at the same time they are often unable to say what will work. Everyone will tell you that there is no recipe.*

*My participation in the project made me discover that the author had to agree to consider his own text with a certain distance, even if he could never be outside the text. When the author agrees to take distance, and the tools offered to him are adapted to the questions he has for himself, then his writing process can be greatly facilitated.*

*In Anglo-Saxon and Germanic countries there exists a profession in publishing which consists in re-reading and improving literary texts. In German, he is called "der Lektor". It is perhaps more to this profession than to writers too fond of their texts that the new tools will be the most useful.*

*There exists in every literary text a nucleus irreducible to analysis, which is therefore not susceptible of optimization. On the other hand, each author makes clumsiness, thinks one thing while he says another. In this field, these digital tools will be of great help. To those who are suspicious of these tools, it must be said: all tools to help creation are ambiguous. A rhyme dictionary does not make a poet but a poet can use a rhyming dictionary to specify what he wants to say. Thus, writers are waiting for the tools that digital humanities are beginning to offer them.*

---

<sup>1</sup>Druide Antidote - <http://www.antidote.info>

## Chapter 6. Results and Discussion

---

In this conclusive answer, we get once again underlined the core issue of the writing tools and their relation to novelists. The common tools here mentioned focus on enhancing the editing process at the level of words or sentences, usually with a syntactic approach (spelling and grammar). They are of course important in their own way, but do not provide a more distanced view on the text, which limits the level of assistance they can provide to a novelist. On the contrary, there is plenty of room for improvement with more structural approaches, where we merely seem to have started scratching the surface. It may also well be that the tools and methods we introduced could be more profitable to someone looking at improving texts rather than writing new ones, such as consulting editors or even more generally, as part of the workflow of any literary editors.

Considering texts from a distance was indeed one of the core ideas that kept coming on and on in most aspects of our work. In the case of someone external to the text, looking at it from a distance is hardly a choice as opposed to the alternative of reading it whole. But to the author, this meaning is very different, as he will immediately know which features the tools exhibit, and thus might be tempted to have an interpretative reading of the mechanics leading to the shown results, rather than their true pertinence. However, given an admittedly more external reading, this artificial distance imposed by visualization tools might also exhibit features that are closer to a reader's perception, and thus help mitigate between the idea a writer has of his text, and the actual message it conveys.

The idea of an irreducible literary nucleus is an interesting concept we didn't explore before this point. One could argue that this definition could be very versatile from work to work. A very intricate short novel could be almost impossible to summarize, whereas one could get a good idea of very long classic works within half a page. But in any case, there is indeed one red wire inside any story, that makes it the original creation of its author and that indeed composition tools will not be suitable to alter. As in a movie where impertinent critics would state how it should have ended in to their view, changing a story in its core necessarily involves a matter of personal taste, and alters the nature of the work. However, as we have shown in our reorganization experiences, there are endless ways of telling the same story, sometimes with better results or different possible perceptions for our readers. This dimension seems to be the one where the potential for optimisation is the highest, when one leaves aside the purely syntactic features of the text.

As a matter of fact, we will discuss in section 6.3.1 an experiment done with *Saga+* that precisely resulted in an attempt of stripping down the text of a novel to its nucleus.

Let us now turn to our second research question:

### Will authors change their way of writing when they know how they are being read?

*The literary writer has an ambition. He wants to write something new, something his readers have not yet read, something that must surprise them or move them despite them. There is therefore a double requirement, one of being read and one of shaking his readers.*

*As he writes, every writer imagines a reader or series of readers: a mother, a friend, a publisher, a teacher, usually someone he trusts. When a writer reads his text in public, he quickly feels places that captivate his audience and those who leave him indifferent. It is difficult for him to identify the uninteresting passages when he is in phase of writing. Sometimes he has great difficulty in suppressing a passage which he considers essential.*

*A tool that measures the response of readers to the text is the dream of every writer. The writer does not want at any cost to please his readers. He wants to move them. Too often he moves only himself by writing, so a tool which allows him to know, paragraph by paragraph, how the reader reacts would be of great help. If one can measure pauses or interruptions of the reader it is already a very useful return.*

This contrasted answer summarizes well the conflicting and interdependent interests a novelist faces when reaching to the public. Being read is indeed a must, since it is arguably the most immediate reward and first indication of success of any piece of writing. The requirement of shaking the public may seem less obvious and maybe rather linked to long-term goals, related to notoriety, distinction gaining or responding to specific expectations from a public. Of course as an author with a cause, Daniel may have a vision for his works that wouldn't be the one of any romance writer, for whom the first requirement, read and please, could be contenting enough.

The idea that the removal of a part that seemed indispensable to the author at some point is slightly contradictory with the concept of nucleus we discussed in the first question. Indeed, what if the said part was part of the nucleus imagined by the author, and readers don't see its value? What if this result had been predicted by any method of analysis beforehand? On the other hand, the subjectivity is high when addressing those questions, and if an author doesn't know the answer, the opinion of the public or of its majority might not be convincing, or even converging.

Finally, on the topic of reading analytics, Daniel seems more enthusiastic than many of his colleagues. Maybe made confident by our long collaboration, or due to his natural inclination

to the topic, he estimates measures of reading pauses and interrupts to be helpful to his writing. The nature and practical relevance of information gathered this way is of course an open question, but we will discuss the technicality of this matter through some concrete examples in section 6.2.2.

## 6.2 Public reception of *simulationhumaine.com*

### 6.2.1 Medias and Scholars

The digital release of the saga was publicly announced at the publication time of the tenth paper volume, both as an inlay card and on the publisher's website<sup>2</sup>.

The site and mobile app sparked interest and debates among independent journalists and enthusiasts of digital reading [Bienvault(2014)]. It was presented in several interviews and public debates [de Roulet(2016)] as well as in various local and national medias [Barraud(2014), Passer(2014), Délétroz(2014)]. As an interesting anecdote, one would also mind later interviews that focused on the political concerns of the contents of the series [Peyret(2016)], for which it is however difficult to say if the digital dimension helped to gain in popularity.

The existence of the digital saga was also relayed on international news channels<sup>3</sup> and newspapers[Catinchi(2014)], as well as state libraries [de Roulet(2014b)]. The project was also very recently cited by literary scholars as an example of noteworthy modern writing study [Lumbroso and Pagès(2016)].

### 6.2.2 Collected Usage Statistics

Over a period of eight months<sup>4</sup>, a total of 7 000 chapters were read, out of which around half were not relatable to regular readers (be it search-engines queries or users refusing to be tracked). Furthermore, some users seemed to have tried the interface per se, looking into one chapter but stopping thereafter. For the needs of our research, we thus kept the subset of data where at least two chapters were related to the same person. We also left out the paths presenting clear test or demonstration patterns (for instance, paths consisting of multiple repetitions of the same chapters) or no overall significant reading time. In all, we considered 310 unique *reading paths*.

---

<sup>2</sup>Buchet Chastel | La Simulation Humaine de Daniel de Roulet : une fin de saga et une expérimentation numérique inédite - <http://www.buchetchastel.fr/actualites/actualite-1716/la-simulation-humaine-de-daniel-de-roulet-une-fin-de-saga-et-une-experimentation-numerique-inedite>

<sup>3</sup>Tendencias 21 | Primeros pasos hacia la Literatura Experimental 2.0 - [http://www.tendencias21.net/Primeros-pasos-hacia-la-Literatura-Experimental-2-0\\_a33358.html](http://www.tendencias21.net/Primeros-pasos-hacia-la-Literatura-Experimental-2-0_a33358.html)

<sup>4</sup>From November 2014 to June 2015

Concretely, each *path* is represented by a chain of tuples  $(c_i, t_i)$ , with  $c_i$  a chapter and  $t_i$  the time the reader spent on it (that is, the interval in seconds between start of  $c_i$  and  $c_i + 1$ ). For instance, if an user reads the first three chapters of a novel made out of five chapters  $\{1, 2, 3, 4, 5\}$  and spending ten minutes on each of them, the resulting *path* will be  $\{(1, 10), (2, 10), (3, 10)\}$ .

Given the average reading speed of an educated adult [Kershner(1964)] and the relative homogeneity of our corpus, we marked transitions as “skipped” when the said time interval was less than 60 seconds, thus considering the chapter as not having been read by that person. We then aggregated the total number of read chapters and plotted them according to their respective stories’ orderings, resulting in one reading curve for each story.

### 6.2.3 Identifying Reading Patterns

Not all readers finish the books they start reading. A curve showing how many people have read each chapter of a book is likely to be progressively going down: the further we look in the book, the lower the number of readers that got to this point. Actually, all reading curves are expected to start with a strong drop-off effect, corresponding to the considerable number of readers who merely peek at the start of the story, but don’t finish it. Of course, this effect is especially noticeable in the context of digital and mobile content<sup>5</sup>, possibly due to the lower acquisition cost and commitment from the end-user (one may expect that buying a paper book implies a higher commitment to fully use the said item compared to downloading a mobile app or even just opening a web page). This phenomenon has been well known in web publishing for several years, with sites such as Slate.com openly claiming only a small minority of their readers made it to the end of the articles<sup>6</sup>.

The classic “drop-off” reading curve is characterized by an exponential drop-off as illustrated in fig. 6.1. In our samples, this phase was shown to typically last for the first three or four chapters of the story. Then, the rate of the monotonic drop-off can vary depending on the attractiveness of the texts. Figure 6.1 shows two examples, one with a very high drop-off rate and another one where this effect is present but less important. Notice that in the second case, not only is the drop lower but the effect also wears off earlier.

---

<sup>5</sup>Localytics Indexes, Q3 2014 <http://www.localytics.com/resources/app-indexes-q3-2014/>

<sup>6</sup>You Won’t Finish This Article - [http://www.slate.com/articles/technology/technology/2013/06/how\\_people\\_read\\_online\\_why\\_you\\_won\\_t\\_finish\\_this\\_article.html](http://www.slate.com/articles/technology/technology/2013/06/how_people_read_online_why_you_won_t_finish_this_article.html)

## Chapter 6. Results and Discussion

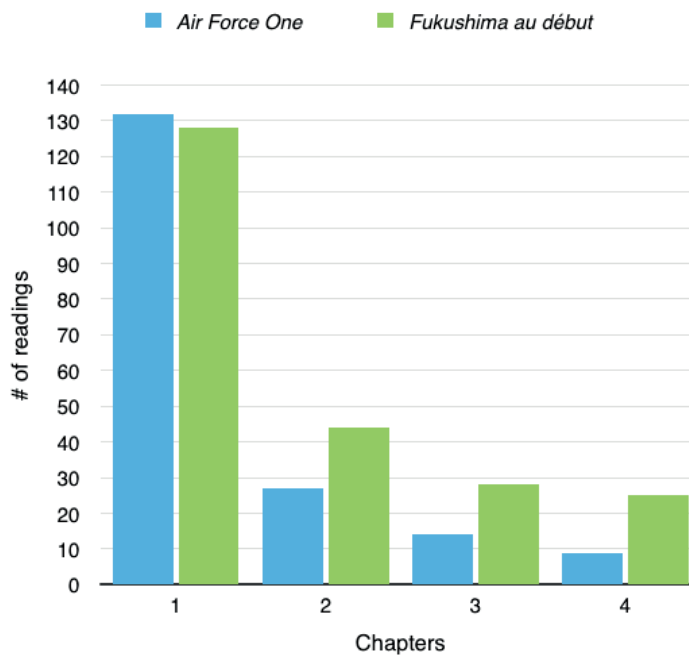


Figure 6.1 – Two examples of exponential drop-offs with different rates / slopes

More interestingly, some “drop-off reading curves” are also characterised by a plateau that occurs typically once the first chapters are passed. This could be because readers who reach a certain point in the book are convinced to go on till the end. Figure 6.2 shows an example of this plateau characterised by an almost null drop-off rate once chapter 6 is passed.

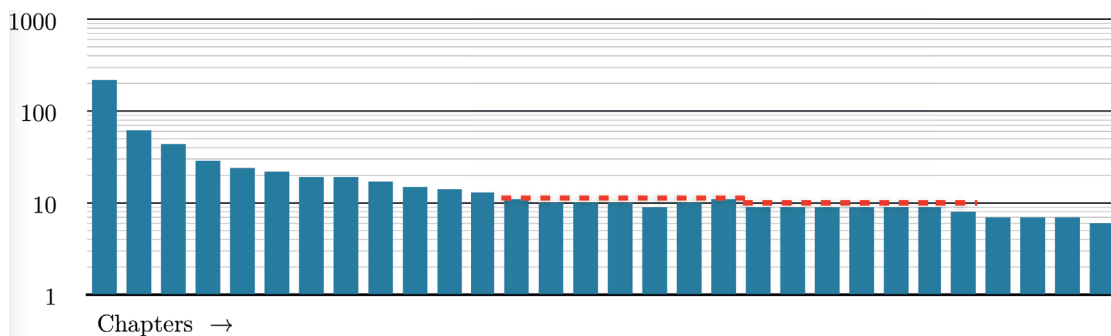
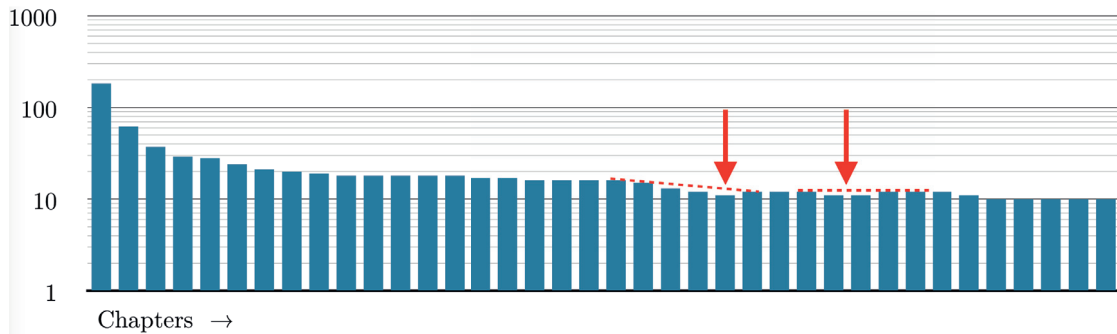


Figure 6.2 – An example of a curve with a plateau in *La réalité, mais digitale*

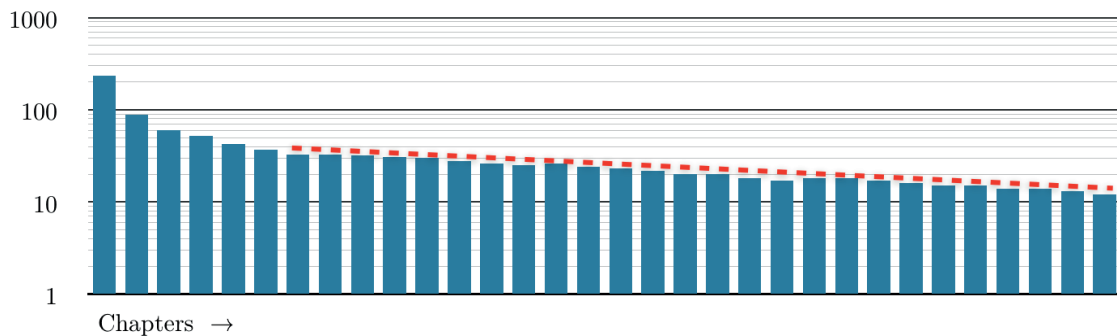
In some cases, the “drop-off reading curves” show unexpected gaps in the reading plateau. This case could typically be attributed to readers who skipped one or several chapters (hence resulting in low reading times that were filtered out of our data), but then resumed reading later on in the same story. This “skimming gap” could thus be interpreted as a mild sign of boredom, not sufficiently strong to stop the reading but significant enough to speed it up. This

phenomenon can be seen both in figs. 6.2 and 6.3.



**Figure 6.3** – Two skimming gaps in the number of readings per chapter in *L'ingénieur et la fillette*.

Moreover, some other stories never achieve the plateau stage, and continuously lose readers as the chapters go by. In the example shown in fig. 6.4, this decrease also features a couple of stop points which could possibly be interesting starting points to investigate why the subsequent drops occur.



**Figure 6.4** – Continuously decreasing readers in *Fissions*

#### 6.2.4 Qualitative Books and Chapters Classification

These simple examples show the potential richness of “drop-off” curves among the various reading analytics curves. Two reading regimes can be identified from this initial study: the immersion mode, characterized by very small drop-off rate, and the critical mode, characterized by dropping and skimming behavior.

We segmented the chapters of our corpus using these two categories and trained a machine learning classifier to try to identify key features to predict whether a chapter is immersive or leading to a potential drop-off. We decided to use an implementation of a J48 pruned tree, given the generally good scores obtained by this method in computational linguistics [Pedersen(2001), Youn and McLeod(2007)] and general purposes classification [Omid(2011)].

## Chapter 6. Results and Discussion

A visual representation of the resulting tree is shown in fig. 6.5.

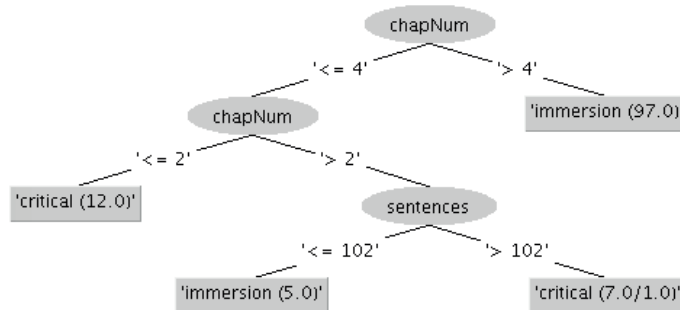


Figure 6.5 – J48 tree predictor for critical and immersion reading modes

This structure yields a 95% correct classification in a 10-fold cross-validation on our dataset, which makes it an overall pretty accurate predictor. It basically confirms our empirical observations on the general shape of the reading curves, putting the start of the immersion phase latest after the fourth chapter. Interestingly, it also hints that using shorter chapters in the critical phase could hasten the immersion process. Furthermore, this result shows that the other features characterizing the chapters, such as their position in the saga, length in words or presence of main characters, had no significant influence on the drop rate prediction.

### 6.3 Workshop with Authors

On November 13th 2016, a public event<sup>7</sup> for writers was organized at the *Fondation Jan Michalski pour l'écriture et la littérature*, in Montricher (Switzerland). During a one-day workshop, authors were introduced to *Saga+* and given a chance to try out the tool by composing a new version of the novel *Frankenstein*.

In addition to the team leading this research (prof. Frédéric Kaplan, Daniel de Roulet and myself), three professional writers and one student from the DHLAB took part to the workshop. Moreover, Mrs. Vera Michalski-Hoffmann was present at the debriefing and closing of the event.

Before getting introduced to *Saga+*, the writers were asked to comment about their usual writing process, specifically on matters related to structural organization. Unsurprisingly, it turned out their concerns and the strategies they used up to now were very similar to the case studies we had done so far. All of them would write in fragments, using non-specialized tools

<sup>7</sup>Actualités EPFL | Frankenstein recomposé - <http://actu.epfl.ch/news/frankenstein-recompose>



*Frankenstein* Reçomposé

Le 13 novembre 2016, 10h - 17h  
Fondation Jan Michalski pour l'écriture et la littérature

Toute histoire est une suite d'évènements dont l'enchaînement n'est jamais laissé au hasard. Mais comment trouver la meilleure organisation narrative, ou celle qui maximisera l'effet désiré ? À quoi ressemblerait un logiciel de "montage" littéraire, à la même façon que l'on travaille avec des vidéos ?

Le DHLAB présente avec Saga un logiciel d'édition permettant de réorganiser n'importe quel long texte divisé en chapitres, que ce soit un roman en cours d'écriture ou une "saga" déjà écrite, afin d'en évaluer différentes possibilités de "montages". À l'aide de technologies permettant automatiquement de détecter personnages et lieux au sein du texte, ce système propose une stratégie de visualisations intuitives simulant les conséquences narratives de chaque changement d'organisation.

Cet outil sera testé en avant-première avec un groupe d'écrivains sur le roman *Frankenstein*. Le dimanche 13 novembre 2016, ceux-ci se réuniront le temps d'une journée à la *Fondation Jan Michalski pour l'écriture et la littérature*, à Montricher. Ils se livreront à la composition des nouvelles plus ou moins longues, reprenant comme base le texte de Mary Shelley, en utilisant Saga pour manipuler facilement le roman et en extraire des sous-ensembles dotés d'un intérêt narratif particulier.

Les romanciers, écrivains professionnels ou amateurs, souhaitant se joindre à l'expérience sont cordialement invités.

Programme de la journée :

10:00	Présentation de l'outil, café & croissants
Dès 11:00	Atelier d'écriture, individuel ou en groupes
12:30	Pause déjeuner
14:00	Reprise des activités d'écriture
16:00	Partage des œuvres créées, remarques conclusives et verre de l'amitié

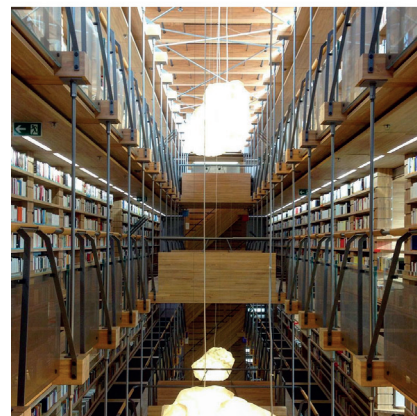
Inscriptions et informations : par e-mail à [cyril.bornet@epfl.ch](mailto:cyril.bornet@epfl.ch), d'ici au jeudi 10 novembre.



**DIGITAL  
HUMANITIES**  
- LABORATORY



FONDATION  
JAN MICHALSKI  
POUR  
L'ÉCRITURE  
ET LA  
LITTÉRATURE



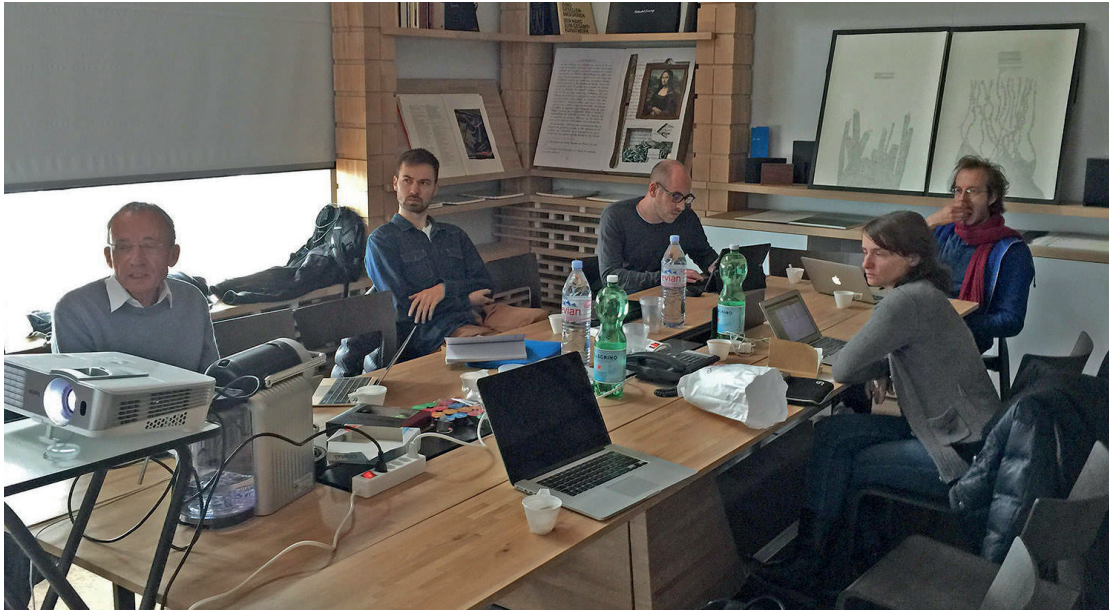
(a) Display

(b) Fondation Jan Michalski

Figure 6.6 – Advertising material for the workshop

(or *Scrivener* for one of them), and use material strategies such as post-its to organize the overall structure before getting to a stage where they would edit the full text, following the final structure. One author also mentioned the use of Excel spreadsheets to sketch plans with the various scenes beforehand, both aiming to list features and details to include in the story, and to shape the organization of the different parts, in particular the case of flashbacks. Another author mentioned the use of codes, such as symbols and color schemes, to remember more precisely the contents of each part and to ease the structuration process.

The introduction of *Saga+* sparked lots of interesting discussions that made this presentation last much longer than first expected. The chapters organization interface immediately reminded the authors of features similar to those they devised to manipulate text fragments by hand. Actually, although the immediate drag-and-drop reordering of chapters was new to them, the color schemes reminding of the contents and hinting at contiguous segments through location continuity seemed a very appreciated feature in itself.



**Figure 6.7** – Presentation of *Saga+* to participants of the workshop. Photo: Frédéric Kaplan.



**Figure 6.8** – Collective writing time at the workshop. Photo: Frédéric Kaplan.

The three other main visualization tools received mixed returns:

- The timeline graph was considered the main and most explicit feature for visualization of full stories, and some minor improvements were suggested. For instance, one of the authors wished for more interactivity, such as the possibility to vertically reorder the list of places, or highlight one specific character by clicking on the corresponding line.

- The characters network graph was discussed in relation to the meaning of the displayed information. One author objected that it wasn't really useful for writing per se, as it was displaying information that was already known beforehand. However, all agreed that it constituted a precious information in a rereading context as it could hint at a reader's perception of the text, and to which extent it matched the view of the world the author intended to convey.
- The entries-exists graph needed more explanation than expected to be fully understood. Two people asked for clarifications about how to read it, which called for a complete illustration using one detailed example with a real story. It thus seemed this visualization to be less obvious to understand and interpret, but it is unclear if this was linked to a weak introduction on our side or an unclear graphical layout. Another explanation could be related to a mismatch between the presented model and the thought process of younger writers that are not driven by the following of classic theatrical composition rules.

As for the editing interface, the feature that sparked most interest was the entities view and its histogram visualization. The authors came up with very prolific ideas about how to use this information, and possible improvements. Amongst other things, there was a clear request for visualizing thematics that wouldn't be tied to specific keywords, as for instance stylistic features (dialogs, poetry), structural ones (flashbacks) or very vague thematics (e.g. all parts where some character is dreaming). Of course not all of these features would be practical to use or even for some of them possible to do without asking for some manual tagging from the user, which goes against the philosophy of our tools, but the mere fact that so many ideas were suggested before the tool was even put to use shows once again the potential demand, and the lack of innovative solutions currently available.

#### 6.3.1 Produced Novel

The authors quickly got used to the graphical interface of *Saga+* and found their way easily, sometimes with minimal guidance. By looking at the various ways to explore the text offered by our tools, they came up with very creative ideas that could be tried out within the very limited time at their disposal. Some of those included to change the structure of the novel to remove its characteristic double embedding of reported speech, composing a new story from the perspective of the monster instead of Victor's, or one in the form of a detective story, following the deaths and revealing the existence of the monster only at the end.

After some time they decided to put the system to the test by collectively composing a short-

## Chapter 6. Results and Discussion

ened version of *Frankenstein*. The challenge here was to identify core parts which would retain the main narrative features of the original story, and compose a new novel only by removing parts of the text and reorganizing the remaining bits. The final novel, produced in less than two hours, retained 8500 words (hence about 10% of the original text) and can thus be read in about half an hour. It can be read online <sup>8</sup> and has been reproduced for convenience in appendix A.7.1. The structuration interface of *Saga+* for this novel in its final stage is shown in fig. 6.9.

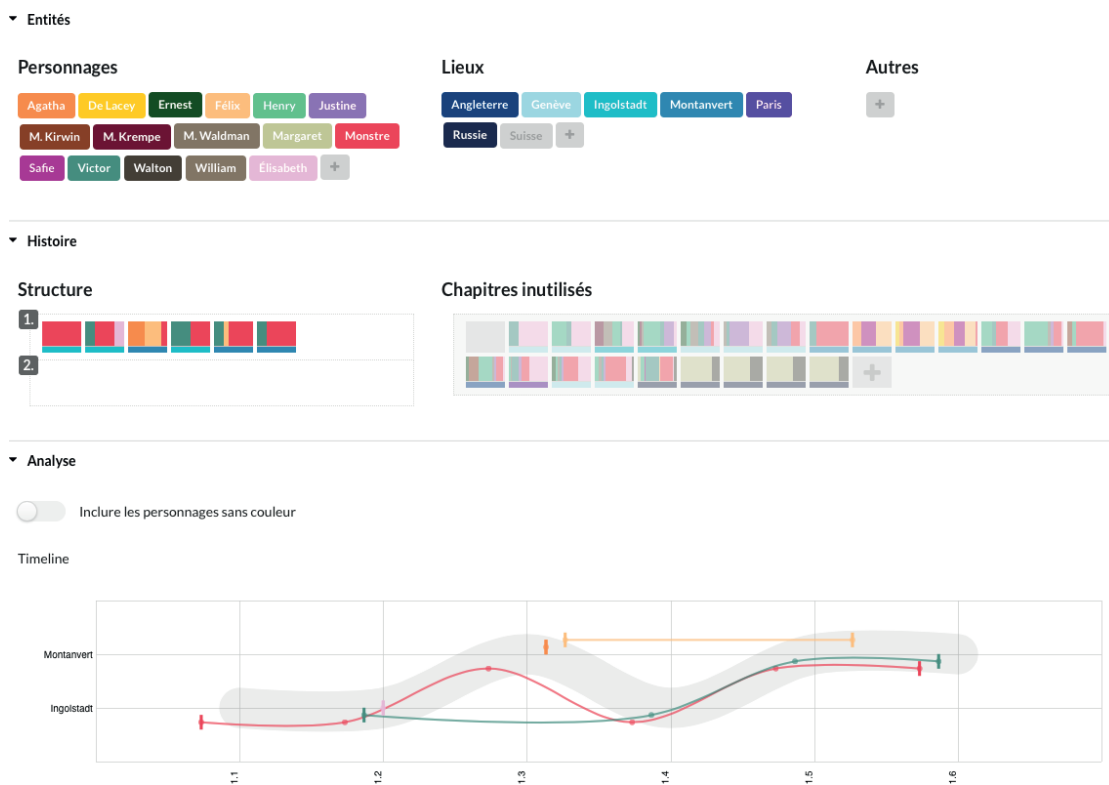


Figure 6.9 – Structure visualization of *Frankenstein (version courte)* in *Saga+*

### 6.3.2 Usability Feedbacks

The three writers who tried *Saga+* kindly answered a couple of questions about the usability and their perception of the tool. We will shortly discuss their answers in this section (the full text compiling them all can be found in appendix A.6.1).

In some way all authors spontaneously agreed that the use of such tools may lead to interesting findings and allow to write different, if not better texts. One author praised the ease of use of the graphical interface, which he considered an essential quality of the platform, and hence

<sup>8</sup><http://saga.plus/reader.php?id=168>

towards which further improvements must not be detrimental:

*« [...] la prise en main me semble pour le moment remarquablement simple (le risque de contre productivité doit être envisagé dans le cas d'ajouts de nombreuses fonctions et réglages)»*

As for the nature of the aspect to be optimized however, no clear target seemed to emerge. One author honestly admitted that the meaning of "optimality" was hard to define given the numerous dimensions that may be involved. The overall main use that the authors saw in this first context was rather structural and linked to solving coherence issues relative to the complexity of long and rich texts. One author stated very clearly:

*[...] je suis plus à la recherche de fonctionnalités qui m'aident à maîtriser le monde fictif que je suis en train de construire. J'ai certains besoins en particulier dans le roman, ou dans certaines pièces de théâtre, disons chaque fois que le travail implique d'édifier un cosmos fictif complet, qui doit se tenir.*

About the reading analytics and their possible influence on writing, opinions were rather contrasted. All stated in some way that the interpretative value one may get only by looking at usage information was limited due to the lack of model about the behavior of readers. Two authors were rather inclined to accept this idea, whereas the third one argued that writing was an innate process whose keys aren't to be found in the readers' opinions:

*« [...] les types de lecture et les voies de l'interprétation sont innombrables; et c'est en soi que se trouvent les clés de l'écriture plutôt que dans les avis des lecteurs. »*

Furthermore, two authors sharply mentioned limitations of reading statistics to understand the emotional and aesthetics dimensions that are core to literature:

*[...] l'essentiel va se passer dans la sphère émotionnelle et sensitive du lecteur, et à ces sphères, je ne vois pas comment l'outil de "Reading analytics" va accéder??*

*[...] cette question, fondamentale, de goût et d'esthétique ne me semble pas pouvoir être résolue par une application informatique [...]*

Finally, one author also mentioned ethical concerns, stressing an uneasy feeling at the idea of being observed during reading:

*« [...] mon plaisir quand je lis c'est l'intimité stricte avec l'oeuvre, la pensée de l'auteur, et me savoir observée [...] me serait désagréable. »*

Regarding possible improvements of the tools, many good ideas were cornered. Some were purely linked to improve or expand the current visualization tools, as for instance the manual labelling of specific entities, or the simultaneous representation of storytelling and chronological orderings. Other ideas echoed to natural further developments that would indeed benefit a lot to the current tools, such as the automatic identification of thematics or resolution of pronouns. Two authors also made the wish to be allowed to work on smaller units than chapters, such as pages or paragraphs.

When asked about their interest for using the tool in the future, all authors showed their enthusiasm by considering to recommend it to colleagues, or having already done so. One author also mentioned the interest to take over an older collaborative project that was abandoned due to its complexity:

*[...] je pense ici à un roman collectif [...] finalement abandonné pour des raisons qui relèvent de la structure de l'intrigue et d'un nombre de personnage difficilement envisageable seulement par une relecture « humaine », même attentive. [...] Saga+ pourrait sans doute montrer, grâce à ses représentations visuelles du récit et des relations entre les personnages – un point fort – les endroits du texte qu'il faut retravailler en priorité.*

### 6.4 The Case of Dramaturgy

The workshop event drove interest of some people that weren't available that day but still wanted to try *Saga+* and see for themselves what the tool could bring to them. Among them, we met a dramaturge that showed a strong interest in our research. Even if this intervention came in very late in the project, we conducted a small mutual presentation allowing to exchange ideas around our tools.

One of the main findings was that, perhaps not surprisingly, several graphs we devised were quite similar to the ones the dramaturge devised by hand when composing theater works. For instance, one very central tool featured a graphical representation showing the presence and number of characters in each scene, with goals such as reaching an equilibrium in play time of the different actors and the narrative rhythm with respect to the story line. The dynamic social network representation was also found interesting, as the strength of the bounds in that case get correlated to the quantity of dialogs between the characters.

#### 6.4. The Case of Dramaturgy

---

Following this interview, we started to evaluate the use of the tool for personal and teaching purposes in the frame of dramaturgy. One piece of work *Solène et le Fils de Zeus*, is currently being written using *Saga+*.





## 7 Limitations and Perspectives

### 7.1 The Present

As already stated, this thesis comes with obvious limitations. Our single subject case study does not allow nor aim at generalization, and the released tools would need many more users and subsequent improvements to get truly disruptive, even if at the end of this study, Daniel de Roulet predicated that our literary classics wouldn't be the same if Honoré de Balzac or Roger Martin du Gard would have benefitted from our tools. As researchers, we can only be flattered and are comforted that an author would be in the best position to evaluate the veracity of this statement, but we need to stay skeptical whilst admitting that a proof or disproof could hardly be brought to it. Probably the only possible methodology that could answer that question would imply an excruciatingly long process to dig results, as well as decades of writing and the luck of a widely used commercial product for generalization.

That being said, we have shown that text *distant editing* responded to a known issue that anyone gets to struggle with when working on lengthy documents, and can be eased through rather simple means in the context of novels. We also know that writing fiction is widely admitted to be hard, and current tools don't really help (as no specialized tools are widely adopted, and many authors' testimonies point to think general purpose tools limit their creative abilities). Besides, despite only being a few more examples, all authors we showed our results to were enthusiast and came up with more ideas and complementary returns of experience that all backed our work hypotheses. This as a whole is a strong indicator for replicability, in the sense that our findings wouldn't be much different if this work had been done with other novelists in a similar context.

Perspectives on the implications of our findings can be very varied, depending on the possibility of generalized adoption of the kind of tools we developed, or the lack of it. One has to

admit that the potential for disruption exists, as cases where a paradigm change in the media translated to a transformation in the way contents are produced were already documented. One very recent example was the on-demand television service Netflix, that sparked interesting debates about the market power it got from analyzing consumer data <sup>1</sup> and how its business model might be changing the way TV series are written <sup>2</sup>.

### 7.2 A Thought Experiment

Another way to look at the perspectives opened by this thesis is to project ourselves in a (not so far) future, where some of the changes we explored may, or may not, have been adopted by a representative share of novelists. Two features summarize the axes explored by our initial research questions: the first one relates to a potential wide use of innovative composition tools fostering *distant editing* strategies, whereas the second one reflects the systematic use of reading statistics as a mean to produce better selling novels. Let us consider and discuss the possible outcomes of those potential futures:

	No Reading Statistics	Massive Use of Statistics
No Composition Tools	<b>Status Quo</b> (1)	<b>Style Dictatorship</b> (2)
Massive Use of Tools	<b>Unbounded Creativity</b> (3)	<b>Chase for Optimality</b> (4)

1. **Status Quo:** this future is similar to our present. Authors may continue using the current text processors and tools that aren't tailored for their work, if the publishing and software industry don't see an opportunity to make this model evolve. Reading statistics might be collected by digital books publishers but not necessarily used for the production of texts, as it is already the case today. This world still is full of opportunities for talented and probably technology-savvy authors and publishers. It doesn't necessarily prevent interesting and disrupting innovations, but they are likely to stay experimental or exceptions to the vast majority of conventional books.
2. **Style Dictatorship:** this future is very similar to the present situation, but where reading statistics get of massive relevance without significant changes in the editing tools. Once the reading statistics get omnipresent in the writing process of enough authors, commercial and business pressures will make it hard for the others to ignore them, and readers may start to dictate what will be written tomorrow by what they're reading today.

---

<sup>1</sup>How Netflix Reverse Engineered Hollywood - <http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/>

<sup>2</sup>Comment Netflix est en train de transformer le processus créatif derrière les séries - <http://www.presse-citron.net/comment-netflix-est-en-train-de-transformer-le-processus-creatif-derriere-les-series/>

One could imagine the reading figures to be provided by digital publishing actors by some mean that would necessarily be distinct from the text processing tools, such as detailed periodical reports.

However, since editing tools do not allow to work efficiently at the organization level of the texts, authors could only stay within traditional styles and structures. Those styles might even get reduced in variety, as the outselling ones get copied and only vary at the thematics and micro-features level. This kind of situation would be very prone to see again big dominating genres alternately take over representative parts of the literary production, as it used to be the case before the 20th century.

3. **Unbounded Creativity:** in this future, authors massively use powerful composition tools that allow them to achieve the desired level of optimality in their works, as well as to explore new kinds of literatures around the concepts linked to narrative recomposition (and possibly many others). Their styles might probably diverge as the tools impose less formatting and call for a larger freedom. Optimization strategies using those tools might get used to their full potential, in a very experimental way, with some authors possibly coming up with new literary genres or storytelling strategies unexplored to this day.

The general disregard of reading analytics, which could for instance be a consequence of ethical restrictions, can be seen as a chance fostering creativity, but also as a risk, potentially leading to a few actors nevertheless using this information or its derivatives to get an unfair advantage.

4. **Chase for Optimality:** this future combines the positive answer to both of our research questions. New tools fostering *distant editing* are widely made available and massively used, as they provide at the same time insights about the various dimensions of the texts but also allow to peek at how they are read. Many interesting behaviors might be expected from there on, as this would probably redefine stylistic and composition rules, as well as the power relation between literary contents producers, publishers and readers. Authors may try to structurally improve their works in a very individual way, according to how their readership reacted to them. Or conversely, they might try to imitate best selling works by reproducing their structural organization, or applying it to other literary genres.

Like in the preceding case and in any world where the consumer is given a central place, a form of convergence would have to be expected. Yet, the use of tools that at the same time allow to easily explore different propositions and give transversal notions of optimality that are not only liked to end-user preferences may see the rise of new genres and original compositions that could in turn get popularized by reading analytics. That

is, provided the said tools don't end up being another limiting technocratic black box that would then push authors to a convergence of styles driven by common optimization strategies.

### 7.3 Conclusion

The contributions of this thesis are multiple. From a methodological point of view, we showed that an highly interdisciplinary study involving actors with very different backgrounds could be conducted using a participatory action research protocol, and uncover results that could be immediately validated by different actors that weren't part of the project. By conceptualizing and formalizing *distant editing*, we showed the validity of an alternative way of looking at narrative composition, that could possibly resonate with future digital technologies to tackle the idea of optimality in literature and hopefully to produce fiction novels of a new kind. We also showed which forms those technologies could take, by prototyping a set of tools that enhanced narrative reorganization with a concrete *distant editing* approach. Even if the potential effects we discussed can only be imagined in the frame of a projection to a semi-distant future and thus lack refutability, we paved the way for future research to investigate them with the right tools.

# **A** Appendices

## A.1 Appendix A

### A.1.1 Exploratory Modeling of Gamebooks

In the early stages of this thesis, we evaluated the possibility of producing a digital saga following the life paths of the protagonists, where the reader would be given a choice of moving to the story of another character every time they co-occur. This led us to explore the formalism behind the composition of gamebooks, and so we set to evaluate the complexity of several works by modeling their sequencing and the possible transitions offered to the player.

Figure A.1 shows such a model on Raymond Queneau's *Un conte à votre façon* [Queneau et al.(1982)], which quickly allows to pull enumeration statistics on the possible stories the game can generate. For instance, the total number of different paths one may run through is 190, whereas the shortest, average and longest length of a story are respectively 4, 12 and 17 segments.

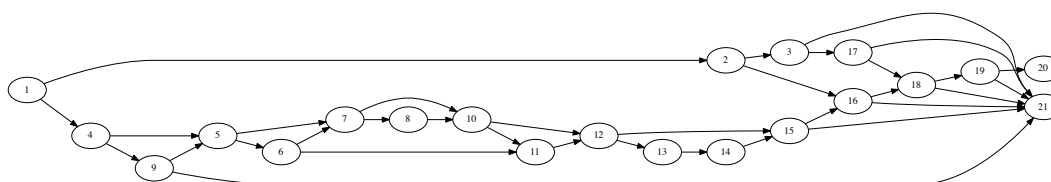


Figure A.1 – Possible readings of Raymond Queneau's *Un conte à votre façon*

For the sake of comparison, we include below two former models, related to short gamebooks aimed to a younger public, pulled from the French bimonthly magazine *Mickey Jeux*. Both contain loops that make the number of possible iterations infinite, which we cut out for the quantitative evaluation. The first and most complex one, *Pierrick remonte le temps* [Kluytmans(1998)] (fig. A.2), features 2128 possible readings with an average story length of 20 segments, whereas the second one, *Yoël-Krahn le magnifique* [Lelièvre(1999)] (fig. A.3), features 92 different stories with an average length of 13 segments.

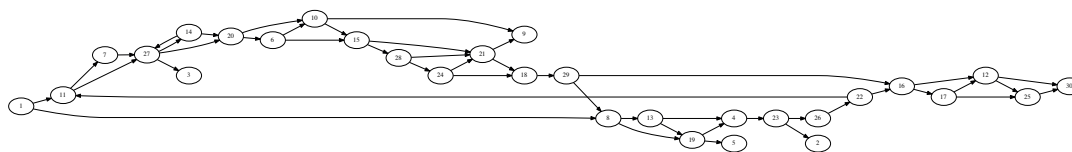


Figure A.2 – Possible readings of *Pierrick remonte le temps*

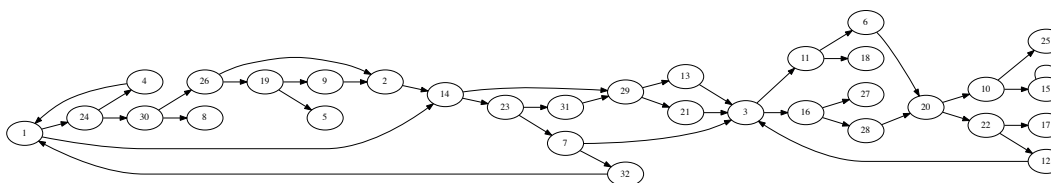


Figure A.3 – Possible readings of *Yoël-Krahn le magnifique*

## A.2 Appendix B

### A.2.1 La succession des chapitres dans trois versions différentes (Daniel de Roulet, 18.11.2013)

Il y a trois personnages dont je reprends la désignation comme indiquée dans les 9 premiers tomes :

Shizuko	D
Max	H
Mirafiori	F

#### **SH.10 version 5, Décembre 2010**

12 chapitres, 2 personnages

DHDHDHDDHDHD

#### **SH.10 version 11, Août 2012**

19 chapitres, 3 personnages

FHHDDFFDHHFHDDFFHFF

#### **SH.10 version 24, septembre 2013**

18 chapitres, 3 personnages

FHDHFHDHFHFHDFDFD

## A.3 Appendix C

### A.3.1 Nouvelles narrations numériques Le projet 3N reconstruit la SH

#### Esquisse d'un papier, avancement du travail début juillet 2013

**Résumé :** Etant donné un corpus de textes homogènes soit par les personnages mis en scène (issus de deux familles et de leurs connaissances), soit par le thème (le développement du nucléaire militaire et ci-vil), soit par l'auteur (le même écrivain pendant vingt ans), une méthode est développée pour proposer de nouveaux parcours de lecture à l'intérieur de ce corpus. Ceux-ci s'appuient sur quelques hypothèses concernant d'une part le modus operandi de l'auteur et d'autre part les nouvelles possibilités offertes par la numérisation des contenus. Le résultat de cette méthode est la création d'un nombre quasi infini de parcours qui tous épuisent dans un ordre différent l'ensemble des textes de ce corpus.

#### Phase 1 : la déconstruction du corpus

Dans la première phase du travail, commencée trois ans avant le début de la collaboration entre le labo DH et l'auteur, ce dernier a rendu aussi homogène que possible l'ensemble du corpus. Le corpus appelé La Simulation humaine (SH) est composé de 9 tomes publiés entre 1993 et 2012 chez différents éditeurs. Et dans un ordre non chronologique. Les 9 tomes (SH1 à SH9) sont tous composés de chapitres (entre une dizaine et une cinquantaine par tome) qui sont en général écrits à partir du point de vue d'un seul personnage par chapitre. Dans la plupart des cas, la focalisation interne est rendue par une troisième personne du singulier (il ou elle) et le temps principal de la narration est l'indicatif présent. Il existait cependant plusieurs chapitres qui ne répondaient pas à ce schéma majoritaire parce qu'ils étaient écrits à la première personne du singulier, la première du pluriel (nous dans SH6), la deuxième du singulier, la deuxième du pluriel (vous dans SH2). Certains chapitres qui mêlaient les voix de deux personnages s'éloignaient aussi du standard

(SH3, SH6, SH7) Afin de pouvoir travailler sur un corpus homogène, il a paru nécessaire à l'auteur de récrire ces chapitres. Cette tâche est terminée à trois exceptions près qui restent à la première personne du singulier. Les trois cas présentent en eux-mêmes une cohérence qui, à ce stade, ne justifient pas une réécriture (le je du narrateur unique de SH8, le je auteur des lettres du kamikaze dans SH1, le je de Reagan dans SH2). La translation à la troisième personne, si elle devait s'avérer nécessaire, peut se faire plus tard. Le produit de cette déconstruction offre un corpus de plus de 280 chapitres de longueurs inégales, tous écrits à la troisième du singulier (à part les exceptions signalées) et présentant le point de vue d'un seul personnage par chapitre.

#### Phase 2 : la segmentation

Pour vérifier la cohérence du travail accompli par l'auteur, le texte de chaque chapitre a été soumis à une analyse lexicale des noms de personnes et de lieux (mots commençant par une majuscule à l'intérieur des phrases) pour s'assurer des résultats de la décomposition. Y avait-il des personnages « trop » présents ou absents là où ils étaient attendus ? Et si oui, y avait-il une explication ou s'agissait-il d'une incohérence à l'intérieur du corpus ? A part deux ou trois cas d'inattention ainsi corrigés, un autre résultat paradoxal de cette analyse a été la mise en évidence d'une possibilité de décomposer encore d'avantage la plupart des chapitres. En analysant le déroulement narratif de chaque chapitre, on s'est aperçu qu'il était possible de retrouver à travers tout le corpus des portions de texte d'environ 5000 signes qui constituaient un découpage cohérent, séquence narrative présentant en général une unité temporelle. Chaque chapitre a été alors divisé en séquences significatives entre 2000 et 6000 signes (entre 1 à 7 par chapitre). Le corpus compte donc 9 tomes, plus de 280 chapitres et plus de 680 séquences. Cette dernière sous-division a été vérifiée et affinée par l'analyse lexicale des noms pro-



pres. Pour le moment les séquences n'ont pas été utilisées dans les phases suivantes (sauf dans l'analyse de SH5). Elles sont pourtant riches de possibilités qui pourront être mises en œuvre par la suite.

### Phase 3 : hypothèse de reconstruction

A partir de chapitres dont chacun porte l'étiquette (le tag) d'un personnage, il est possible de reconstruire des récits homogènes qui regroupent à chaque fois le parcours d'un seul personnage. On obtient ainsi 17 récits (14 à la troisième personne et 3 à la première) de longueur fort variable (de 4 à 52 chapitres). Comme chaque chapitre possède en outre une étiquette temporelle (time stamp), il est possible de remettre le récit concernant un personnage dans l'ordre chronologique. Ainsi est obtenu un premier résultat : la reconstruction de 17 récits retraçant chacun les aventures d'un seul personnage, en général depuis son enfance jusqu'au point où l'auteur de la SH l'a conduit, éventuellement sa mort ou un autre événement important, mais parfois aussi sans point final ni dénouement du récit. La lecture de ces récits, même par un lecteur de bonne volonté, ne présente cependant que rarement un vrai intérêt littéraire. Exception fait des trois récits à la première personne (Tsutsui, George, Reagan) qui justifient ainsi la cohérence que l'auteur leur a donné dès le début. Dans le cas des autres personnages importants (Max, Shizuko, Mirafiori) la mise à plat de leur récit présente une intéressante source d'information, mais appauvrit la tension narrative du récit. On passe d'un épisode à l'autre comme dans une biographie, mais non seulement on rate certains développements mais encore l'empathie du lecteur diminue parce que rien ne vient le contrarier. En retirant au personnage la possibilité de se confronter à d'autres et en supprimant le point de vue des autres sur lui, on en fait une figure qui a perdu en relief. Bref, le plaisir de lecture s'en va peu à peu.

### Phase 4 : analyse du modus operandi de l'auteur

Après l'échec de la reconstruction linéaire, il a bien fallu se demander si d'autres parcours

de lecture, parcours plus paradoxaux, pouvaient faire renaître la tension narrative que l'auteur avait mise dans la construction de chaque tome publié (SH1 à SH9) avec plus ou moins de bonheur. La réflexion de l'auteur sur son propre travail, appuyée sur une lecture attentive de la littérature décrivant l'état de l'art en narratologie (cf. biblio) a permis de rebondir en étudiant les procédés utilisés aussi bien dans la SH que dans les sagas littéraires où se pose en permanence cette double question : qu'est-ce qui intrigue le lecteur et provoque son plaisir de lecture ? Dans SH5, deux récits sont présents simultanément dans chaque chapitre (d'où l'intérêt du séquençage des chapitres qui les séparent). Le premier récit suit le parcours d'un marathonien qui va du départ de la course, (chapitre 0) à l'arrivée de la course 42 km plus loin, soit 27 miles (chapitres 27). Le second récit décrit aussi une course de 27 miles, à raison d'un mile par chapitre, courue par le même personnage mais quinze ans plus tôt. Il s'agit du souvenir de cette course écrite à l'imparfait. La particularité de cette seconde course : elle est écrite à l'envers, de l'arrivée au chapitre 0 au départ au chapitre 27. La première course est publique à New York, la deuxième à rebours est clandestine, nocturne et champêtre. Le lecteur va bien sûr du chapitre 0 au chapitre 27 et est intrigué, voire tenu en haleine, parce que d'une part il se demande si le personnage, Max, va réussir son marathon, mais aussi il comprend, au fur et à mesure que la seconde course se déroule à l'envers que le véritable enjeu narratif est l'attentat que Max a commis dans sa jeunesse. Le procédé est maintenu sans un écart d'un bout à l'autre du tome, la structure de la double course est assez vite comprise par le lecteur qui sait qu'il n'y aura pas de surprise dans la construction. Sa curiosité est maintenue chapitre après chapitre. La fin du livre constitue donc un double accomplissement : Max réussit son marathon et Max réussit son attentat contre une centrale nucléaire. Pour le lecteur, c'est aussi une double récompense à sa curiosité. Evidemment ces deux récits s'ils étaient déconstruits et racontés l'un après l'autre dans l'ordre chronologique ne présenteraient pas le même intérêt. L'auteur en changeant l'ordre

de la narration aiguillonne le lecteur selon un procédé qui relève simplement de la curiosité. Dans SH2, un autre procédé de construction est utilisé. Les 54 chapitres sont racontés selon le point de vue de toute une série de personnages. (cf. graphique en couleur). Certains chapitres après avoir exposés le point de vue d'un personnage ne sont pas suivis par le chapitre qu'on attendrait. Le récit est interrompu, parfois même abruptement, par l'intrusion d'un chapitre donnant le point de vue d'un autre personnage qui va introduire des éléments nouveaux dans la narration, mais suspendre (laisser en suspense) des éléments du récit en cours. Le lecteur est sollicité pendant sa lecture pour faire des pronostics. Que va-t-il se passer dans le récit de ce personnage ? Mais le suspens n'est pas seulement produit par ces interruptions et le croisement des fils narratifs, la temporalité, elle aussi, provoque un effet de suspens. Tout SH2 se déroule en un seul jour (2 juin 1988) où plusieurs grands événements se déroulent en parallèle (un accord international, une fusion d'entreprise, les retrouvailles d'un couple qui s'était formé dans SH1. Pour expliquer ce qui se passe ce jour-là, plusieurs chapitres se situent en flash-back dans les années qui précèdent l'événement (entre 1968 et 1988). Ces retours de temporalité eux aussi retardent le dénouement de la journée. Ils suspendent le jugement que le lecteur pourrait avoir sur le grand récit. La déconstruction de SH2 permet de regrouper les informations, sa reconstruction s'appuie sur le suspense. Ces deux manières (curiosité et suspense) sont abondamment documentées dans les études de narratologie (cf. biblio). Pour la curiosité, on prendra l'exemple de l'opposition entre roman policier et roman noir faite par Todorov : « Il existe deux formes d'intérêt tout à fait différentes. La première peut être appelée la curiosité, sa marche va de l'effet à la cause : à partir d'un certain effet (un cadavre et certains indices), il faut trouver sa cause (le coupable et ce qui l'a poussé au crime). La deuxième forme est le suspense et on va ici de la cause à l'effet : on nous montre d'abord les causes, les données initiales (des gangsters qui préparent de mauvais coups) et notre intérêt est soutenu par l'attente de ce qui va ar-

river, c'est-à-dire des effets (cadavres, crimes, accrochages). » Baroni p. 101. Pour le suspens à ne pas confondre avec la surprise, on donnera l'exemple du maître Hitchcock lors d'une interview à François Truffaut : « La différence entre suspense et surprise est très simple ... Nous sommes en train de parler, il y a peut-être une bombe sous cette table et notre conversation est très ordinaire, il ne se passe rien de spécial, et tout d'un coup : boum, explosion. Le public est surpris, mais, avant qu'il ne l'ait été, on lui a montré une scène absolument ordinaire, dénuée d'intérêt. Maintenant, examinons le suspense. La bombe est sous la table et le public le sait, probablement parce qu'il a vu l'anarchiste la déposer. Le public sait que la bombe explosera à une heure et il sait qu'il est une heure moins quart - il y a une horloge dans le décor - ; la même conversation anodine devient tout à coup très intéressante parce que le public participe à la scène. Il a envie de dire aux personnages qui sont sur l'écran : « Vous ne devriez pas raconter des choses si banales, il y a une bombe sous la table, et elle va bientôt exploser. » Dans le premier cas, on a offert au public quinze secondes de surprise au moment de l'explosion. Dans le deuxième cas, nous lui offrons quinze minutes de suspense. »

### **Phase 5 : remise en intrigue.**

Dans tout le corpus de la SH, on constate l'utilisation, entre autres, de deux procédés (suspense et curiosité). Il s'agit donc de les mettre en œuvre pour une nouvelle tentative de recomposition. Voici comment nous procéderons : La curiosité étant provoquée par une inversion temporelle initiale (la découverte du cadavre au début du roman policier) il s'agira de trouver dans le parcours de chaque personnage un épisode particulier (invraisemblable, ou mystérieux ou seulement bizarre). Cet épisode étant communiqué au lecteur comme toute première observation, il s'agira ensuite de reprendre le récit dans l'ordre chronologique jusqu'au point non éclairci qui s'expliquera alors. La curiosité du lecteur sera aiguillonnée par l'attente de la résolution de la situation initialement décrite. Encore faudra-t-il trouver de tels épisodes clé. Toutes les vies des

personnages de la SH ne possèdent pas un tel moment. A ce point, on peut envisager que l'auteur soit sollicité pour reformuler une telle situation en réécrivant éventuellement le chapitre à mettre au début. Pour le suspens en revanche, la chronologie ne gagne pas à être bousculée. Au contraire, son strict rétablissement (lors de la déconstruction) n'aura pas été inutile. On utilisera ici une technique dont il est fait abondamment usage à l'intérieur des tomes de la SH : la coupure brusque dans le flux du récit, provoqué par le changement de focalisation d'un chapitre à l'autre. Il s'agira de remixer deux ou plusieurs trames narratives pour qu'elles se gênent l'une l'autre, interrompant le récit au moment où le lecteur attend le plus intensément sa suite. Comme souvent les chapitres du corpus se terminent sur un tel point d'interrogation, on utilise ici une propriété intrinsèque de l'écriture de l'auteur. Il suffira de prendre une paire de personnages et de mêler deux voix, provoquant des interrogations perturbatrices à chaque fois que cela est possible, c'est-à-dire aux endroits où les questions ouvertes sur chaque personnage sont maximales. On peut ensuite multiplier le nombre de voix interrompues (pas seulement deux mais trois ou plus).

#### **Phase 6 : présentation numérique**

Le lecteur est mis en face d'une représentation graphique du corpus où chaque chapitre est représenté par une case (il y en a 280) qui porte l'étiquette d'un personnage. Au fur et à mesure de sa lecture, les cases lues sont marquées jusqu'à ce que toute aient été marquées et donc lues. Etant donné les deux techniques de remise en intrigue choisies, il s'agira de les combiner et d'en tester et évaluer les résultats. Au départ, le lecteur choisira dans le double arbre généalogique qui lui est présenté (familles Vom Pokk et Tsutsui) le personnage à partir duquel il veut entrer dans le corpus. Ce sera son seul choix, ensuite tout lui est imposé. S'il ne veut pas choisir, un personnage initial lui

est attribué de manière aléatoire. A partir de là le lecteur est introduit dans le labyrinthe de la recomposition de manière aléatoire. Reste à trouver un arrangement entre les parties recomposées selon la méthode de la curiosité et du suspense. Au début, on les alternera en essayant de reconstituer des patterns aussi proches que possibles de ceux qui sont donnés par l'analyse des 9 tomes publiés. On produira des représentations colorées des parcours proposés à travers les 280 chapitres. Le graphisme constituera une aide précieuse et directe. Parallèlement seront mis en place des outils permettant de vérifier l'intérêt que le lecteur a pris à une remise en intrigue plutôt qu'à une autre. On pourra même définir, par un système de scores, la probabilité que cette technique produise un ou plusieurs parcours procurant au lecteur un plaisir plus intense que celui qui résulte de la lecture des tomes 1 à 9 de la SH.

#### **Discussion :**

Dans un corpus musical comme dans un corpus littéraire, on peut essayer de déconstruire puis reconstruire. Il existe un cas d'école, celui de Beethoven. Après sa mort, tous les conservatoires où s'enseignait la composition musicale ont cru pouvoir analyser et déconstruire l'œuvre de Beethoven pour en tirer des règles de composition. On disait : voilà comment est construite un sonate, une symphonie, on pourra donc créer à volonté des sonates de Beethoven. Mais le dix-neuvième siècle n'était pas terminé que déjà toutes ces règles de recomposition étaient obsolètes. Il s'agit donc de ne pas s'emballer. La déconstruction puis reconstruction d'un corpus et la recherche des règles qui sous-tendent ces processus ne sont sans doute possibles que quand il s'agit d'un corpus homogène. Avant de commencer à énoncer des règles universelles, d'autres corpus devraient être soumis à ces processus. Alors seulement on pourra croire s'approcher des mécanismes de la création.

### A.3.2 Les étapes de *La simulation humaine* (Daniel de Roulet, 28.04.2014)

#### L'existant

Au départ, il y a un auteur qui écrit depuis un certain temps sur le même thème, avec les mêmes personnages, neuf romans et un dernier en cours qui forment un corpus. Chacun de ces romans peut se lire indépendamment de tous les autres. Chaque chapitre du corpus est écrit à partir du point de vue d'un seul personnage. Par là, c'est un corpus homogène. Mais il est aussi hétérogène à cause de ses styles différents (l'écriture de l'auteur change sur deux décennies), à cause du point de vue qui change souvent d'un chapitre à l'autre (en tout 17 points de vue différents) et à cause du public auquel il s'adresse (certains volumes ont été écrits pour un public moins lettré que d'autres).

#### Le projet de l'auteur en 2010

1. Publier la saga telle quelle en remettant dans un ordre chronologique les dix tomes qui vont de 1938 à aujourd'hui (ce n'était pas l'ordre de la publication)
2. Publier plusieurs histoires qui suivent la vie d'un seul personnage (en tout 17 livres nouveaux). Pour cela, réécrire les 9 premiers tomes en réduisant les hétérogénéités, mettre tout à la troisième personne du singulier, tout au temps présent, avec le moins de retours en arrière possibles. (Cette déconstruction et réécriture a été menée à terme en trois ans.)
3. Utiliser éventuellement l'informatique afin d'enrichir la saga par des documents multimédia collectés au fur et à mesure de l'écriture (textes historiques, photos, vidéos, archives de journaux, revues scientifiques) et afin de permettre de lire sur d'autres supports (tablettes, Smartphones)

#### La rencontre avec l'EPFL en 2012

Le DHLAB propose d'emblée d'utiliser le numérique non pas pour enrichir le texte par le multimédia (objectif de l'auteur au début) mais en se concentrant sur le texte seulement. Différentes méthodes d'analyse du corpus sont essayées puis abandonnées, puis récupérées.

Etude sur la succession des points de vue d'un chapitre à l'autre. Selon quels patterns à l'intérieur d'un tome passe-t-on d'un personnage à un autre ?

1. Une généralisation est-elle possible ?

(présentation d'un papier de linguistique computationnelle)

2. Etude de la succession des phases de la narration : augmentation de la tension, climax et résolution (étude détaillée de deux tomes).
3. Analyse des fréquences de l'apparition de noms propres dans chaque chapitre, noms répertoriés dans une nomenclature fournie par l'auteur (utilisé pour comparer les intentions de l'auteur et la réalité du texte)
4. Proposition pour découper le texte en séquences plus courtes que le chapitre, apparition de 600 séquences comprenant entre 4 et 6000 signes (utilisé dans la reconstruction pour éliminer les retours en arrière).
5. Analyse du cas d'autres corpus qui proposent des modes de lecture multiple (Cortazar, Queneau, etc.).

Toutes ces méthodes sont des culs-de-sac féconds, c'est-à-dire qu'elles permettent de préciser les limites de la recherche et d'ouvrir d'autres voies.

#### La remise en cause (novembre 2013)

La recomposition de la saga en histoires qui ne suivent qu'un personnage a produit des récits très pauvres en tension narrative (ennui de lecture). Les analyses numériques appliquées au corpus, même sans résultats probants, sont désormais disponibles pour la suite. A ce stade, et vu la contrainte d'avoir une application disponible pour la sortie papier du tome 10, on a décidé que l'auteur proposerait lui-même une série de parcours alternatifs cohérents. D'où une application numérique comprenant :

- 3 chapitres isolés
- 3 nouvelles entre 4 à 10 chapitres
- 3 romans entre 30 et 38 chapitres
- un corpus.

Les contraintes posées à cet exercice sont qu'

- aucun des chapitres proposés dans l'une des dix combinaisons de chapitres ne devait réapparaître dans la lecture.
- aucun des trois romans ne pouvait être identique à un roman déjà publié.

Il s'agit donc de trouver pour tout le corpus une combinaison qui reste significative du point de vue du lecteur alors qu'on a pu lui soustraire

jusqu'à trois chapitres isolés, plus trois nouvelles plus trois romans. Si une telle combinaison existe, l'auteur se trouvera avoir construit un nouvel objet littéraire inédit et cohérent. S'il échoue, il aura prouvé que seule une agrégation préétablie (les tomes tels que publiés ou la saga telle que recomposée par dix tomes mis en ordre chronologique) est possible.

#### **Premiers résultats (avril 2014)**

Le pari a été tenu. Dix sous-ensembles du corpus, tous testés et assumés par l'auteur, ont été réunis pour l'application et proposés au public selon le planning prévu. Il existe donc au moins un deuxième corpus qui est un objet littéraire nouveau puisque susceptible d'amputations répétées tout en restant cohérent, ce qui n'est pas le cas du corpus de départ.

Il est donc maintenant nécessaire de comprendre comment l'auteur, s'appuyant en grande partie sur son intuition, mais utilisant les résultats des diverses autres manipulations sur le texte, a réussi à produire ce résultat. Existe-t-il d'autres corpus de cette nature, peuvent-ils être le résultat d'algorithmes ?

#### **L'atelier de l'écrivain**

La collaboration du DHLAB avec l'auteur permet déjà certaines conclusions :

- Les outils numériques testés le sont tous sur ce corpus très particulier. Leur application sur d'autres corpus n'est pas envisageable dans l'état actuel du travail.
- À chaque fois qu'un cul-de-sac fécond a été exploré, des outils utilisés et des résultats partiels ont pu être récupérés comme bénéfiques collatéraux. Ainsi l'auteur a utilisé les mesures de la tension narrative pour le tome 10 et le DHLAB a mis au point une méthode d'analyse du passage du témoin narratif d'un personnage à l'autre.
- L'auteur s'est familiarisé avec toutes sortes de d'outils numériques en les mettant directement en œuvre après y avoir été initié par le DHLAB (production d'un site de l'auteur, contrôle de la nomenclature, etc.)
- Le DHLAB a pu, à chaque pas, se confronter au jugement de l'auteur sur la pertinence littéraire des propositions faites.
- Grâce à l'outil de recomposition et de lecture mis à disposition par le DHLAB, l'auteur a pu se mettre dans la situation d'un lecteur zappeur, extérieur à ses propres textes. C'est ainsi que des épisodes narratifs qui paraissaient fondamentaux ont pu être considérés comme secondaires et vice versa.
- Le DHLAB a pu tester les limites d'un exercice abstrait sur un corpus : la complexité des combinaisons ne peut être augmentée que jusqu'à un certain point qui dépend de la bienveillance du lecteur.

## **A.4 Appendix D**

### **A.4.1 Numeric Results of Named Entities Classifiers**

	1. Obvious Context			2. Naive Position			3. Semantics			4. Grammar			5. Wikipedia			6. Quotes		
	P	R	$F_1$	P	R	$F_1$	P	R	$F_1$	P	R	$F_1$	P	R	$F_1$	P	R	$F_1$
20000 Lieues sous les mers	<b>0.667</b>	0.269	0.383	0.806	0.763	0.784	0.736	0.609	0.667	0.800	0.857	0.828	0.627	0.652	0.639	0.682	0.633	0.657
Au bonheur des dames	0.947	0.354	0.515	0.737	0.864	0.796	0.839	0.695	0.761	0.778	0.840	0.808	0.592	0.701	0.642	0.643	0.568	0.603
Belle rose	0.980	0.419	0.587	0.764	0.859	0.809	0.813	0.799	0.806	0.748	0.831	0.788	0.337	<b>0.363</b>	0.350	0.675	0.609	0.640
Bijou	0.905	0.266	0.411	<b>0.375</b>	<b>0.578</b>	0.455	0.792	0.766	0.778	0.550	0.859	0.671	0.251	0.516	0.338	0.500	0.719	0.590
Blanche et Bleue	1.000	0.312	0.476	0.761	0.823	0.791	0.659	0.542	0.595	0.802	0.885	0.842	0.634	0.708	0.669	0.643	<b>0.229</b>	<b>0.338</b>
Boule de Suif	1.000	0.154	0.267	0.543	0.808	0.649	0.857	0.442	0.584	0.621	0.885	0.730	0.407	0.442	0.424	0.595	0.731	0.656
Candide	0.917	0.208	0.340	<b>0.906</b>	<b>0.875</b>	0.890	0.944	0.487	0.643	0.906	0.875	0.890	0.783	0.583	0.668	0.909	0.513	0.656
Colomba	1.000	0.167	0.286	0.637	0.750	0.689	0.662	0.556	0.604	0.802	0.917	0.856	0.577	0.639	0.606	0.746	0.778	0.762
La Comedie Humaine	0.879	0.451	<b>0.596</b>	0.631	0.811	0.710	0.715	0.645	0.678	0.648	0.857	0.738	0.536	0.633	0.581	0.622	0.675	0.648
Consuelo	0.940	0.239	0.381	0.657	0.728	0.691	0.802	0.696	0.745	0.682	0.787	0.731	0.576	0.650	0.611	0.758	0.704	0.730
David Copperfield	<b>1.000</b>	0.333	0.500	0.593	0.688	0.637	0.787	0.785	0.786	0.763	0.942	0.843	0.635	0.812	0.713	0.622	0.591	0.606
Don Quichotte	0.974	0.313	0.473	0.576	0.792	0.667	0.929	0.596	0.726	0.559	0.811	0.662	0.480	0.602	0.534	0.609	0.650	0.628
Germaine	0.889	0.337	0.489	0.536	0.576	0.555	0.608	0.549	0.577	0.685	<b>0.742</b>	0.713	0.604	0.655	0.629	0.767	0.527	0.624
Germinal	1.000	0.128	0.227	0.731	0.789	0.759	0.795	0.644	0.712	0.821	0.889	0.854	0.637	0.694	0.665	0.462	0.394	0.426
Germinie Lacerteux	0.833	0.200	0.323	0.438	1.000	0.609	0.857	0.500	0.632	<b>0.475</b>	0.800	<b>0.596</b>	<b>0.171</b>	0.600	<b>0.267</b>	<b>0.417</b>	0.600	0.492
Jane Eyre	1.000	0.205	0.340	0.705	0.864	0.776	<b>0.896</b>	<b>0.920</b>	0.908	0.846	0.966	0.902	0.613	0.706	0.656	0.565	0.439	0.494
La Bête Humaine	1.000	0.321	0.486	0.833	0.917	0.873	0.837	0.775	0.805	0.857	0.938	0.896	0.559	0.600	0.579	0.644	0.588	0.615
La Femme du mort	0.964	0.397	0.562	0.501	0.788	0.612	0.716	0.694	0.705	0.721	0.905	0.803	0.304	0.636	0.411	0.518	0.592	0.553
L'Assommoir	0.971	0.242	0.388	0.743	0.864	0.799	0.815	0.594	0.687	0.794	0.879	0.834	0.656	0.742	0.696	0.644	0.588	0.615
Le Mort vivant	1.000	0.207	0.343	0.667	0.707	0.686	0.810	0.567	0.667	0.766	0.858	0.809	0.496	0.502	0.499	0.680	0.605	0.640
Le Parfum de la dame en noir	1.000	0.274	0.430	0.689	0.609	0.647	0.703	0.639	0.669	0.903	0.919	0.911	0.655	0.683	0.669	0.761	0.566	0.649
Les Malheurs de Sophie	0.875	0.188	0.309	0.617	0.812	0.701	<b>0.438</b>	<b>0.438</b>	0.438	0.833	1.000	0.909	0.542	0.625	0.580	0.833	0.875	0.854
Les Miserables	0.933	0.350	0.509	0.734	0.800	0.766	0.841	0.808	0.824	0.774	0.833	0.802	0.409	0.433	0.421	0.694	0.633	0.662
Les Tribulations d'un Chinois	1.000	0.132	0.234	0.710	0.808	0.756	0.792	0.739	0.765	0.757	0.857	0.804	0.506	0.603	0.550	0.692	0.585	0.634
Les trois mousquetaires	0.923	0.275	0.424	0.657	0.762	0.706	0.852	0.752	0.799	0.751	0.875	0.809	0.449	0.530	0.486	0.675	0.656	0.665
Les Vacances	0.889	0.175	0.292	0.562	0.683	0.617	0.807	0.783	0.795	0.667	0.875	0.757	0.531	0.583	0.556	0.628	0.633	0.631
Le Tour du Monde en 80 jours	1.000	0.190	0.320	0.716	0.701	0.709	0.764	0.633	0.693	0.883	0.939	0.910	0.642	0.691	0.665	0.750	0.728	0.739
Le voluptueux voyage	1.000	0.042	0.080	0.650	0.701	0.674	0.875	0.466	0.608	0.893	0.958	0.924	0.362	0.390	0.376	0.634	0.617	0.626
L'Iliade	0.950	0.117	0.208	0.612	0.747	0.673	0.715	0.548	0.620	0.644	0.797	0.712	0.513	0.619	0.561	0.592	0.312	0.408
L'Odyssée	0.929	0.111	0.199	0.627	0.807	0.706	0.566	0.466	0.511	0.642	0.770	0.700	0.522	0.704	0.600	0.604	0.336	0.432
Madame Bovary	1.000	0.241	0.388	0.758	0.816	0.786	0.731	0.680	0.705	0.893	0.963	0.927	0.611	0.660	0.634	0.659	0.606	0.632
Les Mystères de Paris	0.961	0.345	0.508	0.565	0.671	0.613	0.877	0.835	0.856	0.660	0.841	0.740	0.472	0.537	0.502	0.648	0.610	0.629
Nana	0.972	0.197	0.328	0.591	0.841	0.694	0.701	0.687	0.694	0.663	0.913	0.768	0.430	0.534	0.477	0.525	0.544	0.534
Notre-Dame de Paris	0.857	0.266	0.406	0.704	0.720	0.712	0.767	0.624	0.688	0.776	0.837	0.806	0.742	0.833	0.785	0.798	0.728	0.761
Récits d'un soldat	1.000	<b>0.000</b>	<b>0.000</b>	0.667	0.705	0.685	0.625	0.477	0.541	0.958	1.000	0.979	0.458	0.523	0.488	0.617	0.727	0.667
Tarass Boulba	1.000	0.167	0.286	0.626	0.844	0.719	0.821	0.700	0.756	0.703	1.000	0.826	0.586	0.844	0.692	0.650	0.622	0.636
Thérèse Raquin	0.833	0.278	0.417	0.838	0.833	0.835	0.875	0.778	0.824	0.875	0.833	0.854	0.889	0.889	0.889	0.646	0.500	0.564
<b>Average</b>	0.949	0.239	0.369	0.656	0.774	0.706	0.771	0.643	0.695	0.751	0.881	0.808	0.525	0.617	0.562	0.652	0.598	0.615
<b>Std. Deviation</b>	0.070	0.101	0.133	0.110	0.091	0.088	0.105	0.124	0.102	0.111	0.065	0.085	0.143	0.119	0.130	0.099	0.129	0.103

Table A.1 – Overall precision, recall and  $F_1$  scores. Lowest and highest scores are underlined in bold.

## Appendix A. Appendices

	Opener			Simple Vote			Self-Assessment			Fixed Weights			Meta-Recognition		
	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>
20000 Lieues sous les mers	0.696	0.858	0.769	0.760	0.776	0.768	0.732	0.757	0.744	0.732	0.757	0.744	0.732	<b>0.757</b>	0.744
Au bonheur des dames	0.572	0.795	0.665	0.862	0.945	0.902	0.862	0.945	0.902	0.862	0.945	0.902	0.862	0.945	0.902
Belle rose	0.645	0.638	0.641	0.832	0.936	0.881	0.832	0.936	<b>0.881</b>	0.826	0.921	0.871	0.892	0.972	0.931
Bijou	0.569	0.788	0.661	0.454	0.766	<b>0.570</b>	<b>0.508</b>	0.812	<b>0.825</b>	0.532	0.828	0.648	0.872	0.875	0.874
Blanc et Bleu	0.616	0.714	0.661	0.833	0.917	0.873	0.795	0.885	0.838	0.795	0.885	0.838	0.814	0.906	0.858
Boule de Suif	0.524	0.857	0.650	0.543	0.808	0.649	0.673	0.923	0.779	0.576	0.846	0.685	<b>0.576</b>	0.846	0.685
Candide	0.895	0.778	0.832	0.933	0.917	0.925	0.933	0.917	0.925	0.933	0.917	0.925	0.920	0.920	0.920
Colomba	0.632	0.864	0.730	0.739	0.889	0.807	0.769	0.917	0.836	0.769	0.917	0.836	0.769	0.917	0.836
La Comedie Humaine	0.545	0.721	0.621	0.737	0.889	0.806	0.745	0.908	0.819	0.745	0.908	0.819	0.746	0.844	0.792
Consuelo	0.679	0.787	0.729	0.791	0.896	0.840	0.791	0.896	0.840	0.791	0.896	0.840	0.820	0.901	0.859
David Copperfield	0.638	0.877	0.739	0.794	0.957	0.868	0.833	0.971	0.897	0.833	0.971	0.897	0.885	0.986	0.932
Don Quichotte	0.477	0.782	0.593	0.566	<b>0.744</b>	0.643	0.566	<b>0.744</b>	0.643	0.566	<b>0.744</b>	0.643	0.598	0.772	<b>0.674</b>
Germaine	0.610	0.694	0.649	0.721	0.784	0.751	0.721	0.784	0.751	0.721	0.784	0.751	0.788	0.822	0.805
Germinal	0.683	0.692	0.687	0.794	0.850	0.821	0.812	0.878	0.844	0.812	0.878	0.844	0.812	0.878	0.844
Germinie Lacerteux	0.321	0.600	0.418	<b>0.438</b>	1.000	0.609	0.528	1.000	0.691	<b>0.524</b>	1.000	0.688	0.857	1.000	0.923
Jane Eyre	0.619	0.891	0.731	0.875	0.977	0.923	0.909	0.989	0.947	0.934	0.979	0.956	0.950	1.000	0.974
La Bête Humaine	0.683	0.711	0.697	0.934	0.979	0.956	0.934	0.979	0.956	0.934	0.979	0.956	<b>0.980</b>	1.000	<b>0.990</b>
La Femme du mort	0.713	0.822	0.764	0.587	0.939	0.723	0.670	0.970	0.793	0.682	0.970	0.801	0.669	0.920	0.775
L'Assommoir	0.462	0.623	0.531	0.816	0.924	0.867	0.816	0.924	0.867	0.816	0.924	0.867	0.814	0.889	0.850
Le Mort vivant	0.718	0.872	0.788	0.766	0.858	0.809	0.816	0.854	0.835	0.816	0.854	0.835	0.856	0.872	0.864
Le Parfum de la dame en noir	0.811	0.878	0.843	0.885	0.919	0.902	0.864	0.899	0.882	0.864	0.899	0.882	0.864	0.899	0.882
Les Malheurs de Sophie	<b>0.267</b>	<b>0.571</b>	<b>0.364</b>	0.771	0.938	0.846	0.944	1.000	0.971	0.944	1.000	0.971	0.944	1.000	0.971
Les Miserables	0.693	0.800	0.743	0.821	0.900	0.859	0.831	0.908	0.868	0.866	0.950	0.906	0.856	0.933	0.893
Les Tribulations d'un Chinois	0.569	0.654	0.609	0.721	0.846	0.779	0.776	0.923	0.843	0.776	0.923	0.843	0.776	0.923	0.843
Les trois mousquetaires	0.510	0.758	0.610	0.745	0.875	0.805	0.773	0.900	0.832	0.773	0.900	0.832	0.828	0.924	0.873
Les Vacances	0.430	0.711	0.536	0.774	0.950	0.853	0.774	0.950	0.853	0.774	0.950	0.853	0.850	0.975	0.908
Le Tour du Monde en 80 jours	0.802	0.903	0.850	0.800	0.820	0.810	0.845	0.892	0.868	0.845	0.892	0.868	0.845	0.892	0.868
Le voluptueux voyage	0.646	0.625	0.635	0.801	0.871	0.835	0.817	0.875	0.845	0.817	0.875	0.845	0.867	0.875	0.871
L'Illiade	0.494	0.583	0.535	0.616	0.762	0.681	0.653	0.801	0.720	0.653	0.801	0.720	0.694	0.833	0.757
L'Odyssée	0.604	0.631	0.617	0.612	0.807	0.696	0.597	0.762	0.669	0.597	0.762	0.669	0.622	0.790	0.696
Madame Bovary	0.471	0.667	0.552	0.838	0.872	0.855	0.898	0.936	0.917	0.898	0.936	0.917	0.936	0.936	0.936
Les Mystères de Paris	0.542	0.800	0.646	0.703	0.879	0.781	0.730	0.891	0.803	0.782	0.916	0.844	0.885	0.946	0.915
Nana	0.524	0.723	0.608	0.676	0.929	0.782	0.665	0.921	0.772	0.693	0.890	0.779	0.667	0.827	0.738
Notre-Dame de Paris	0.721	0.772	0.746	0.798	0.890	0.842	0.798	0.890	0.842	0.798	0.890	0.842	0.796	0.893	0.842
Récits d'un soldat	<b>0.923</b>	<b>1.000</b>	<b>0.960</b>	<b>0.958</b>	<b>1.000</b>	<b>0.979</b>	<b>0.958</b>	<b>1.000</b>	<b>0.979</b>	<b>0.958</b>	<b>1.000</b>	<b>0.979</b>	0.958	1.000	0.979
Tarass Boulba	0.571	0.778	0.659	0.687	1.000	0.814	0.738	1.000	0.849	0.738	1.000	0.849	0.738	1.000	0.849
Therèse Raquin	0.675	0.667	0.671	0.950	0.944	0.947	0.950	0.944	0.947	0.950	0.944	0.947	0.950	0.944	0.947
<b>Average</b>	0.608	0.756	0.669	0.749	0.889	0.809	0.775	0.904	0.831	0.776	0.903	0.831	0.815	0.907	0.857
<b>Std. Deviation</b>	0.137	0.103	0.116	0.127	0.071	0.096	0.115	0.069	0.088	0.117	0.070	0.090	0.103	0.067	0.082

Table A.2 – Overall precision, recall and F<sub>1</sub> scores. Lowest and highest scores are underlined in bold.



## A.4.2 Semantic Classifier Words

### A.4.2.1 Characters

**titres** monsieur, homme, mâle, mec, quidam, seigneur, sieur, madame, dame, mademoiselle, aimée, amie, amoureuse, bien-aimée, commandante, compagne, concubine, copine, dirigeante, dulcinée, employeuse, favorite, femme, fiancée, initiatrice, négrière, pédagogue, primordiale, reine, supérieure, prince, altesse, émir, archiduc, évêque, célébrité, cardinal, dauphin, diadoque, empereur, gloire, grand-duc, hospodar, infant, kronprinz, landgrave, maharadjah, maharajah, maharaja, margrave, monarque, monseigneur, principicule, rajah, rhingrave, roi, sire, souverain, autocrate, César, chah, despote, dynaste, mogol, kaiser, khan, magnat, majesté, mikado, phénix, pharaon, ponte, potentat, raja, roitelet, sultan, trésor, abbé, capelan, corbeau, curé, curiaillon, cureton, dignitaire, ecclésiastique, inquisiteur, inquisitrice, père, pontife, prédicant, prélat, prestolet, prieur, raticchon, réxerend, religieux, pape, agrée, collecteur, décimateur, exacteur, fisc, impôt, baderne, belliqueux, caporal, colonel, conscrit, lieutenant, maréchal, martial, polémologique, rengagé, sergent, soldatesque, sous-officier, tactique, troupe, appelé, bachi-bouzouk, bidasse, biffin, bleu-saille, bras, cavalier, champion, cipaye, combattant, conquérant, deuxième, pompe, drille, engagé, estafette, evzone, factionnaire, fantassin, franc-tireur, fusilier, garde-voie, goumier, grenadier, griveton, grognard, guérillero, guerrier, jalonneur, janissaire, légionnaire, méhariste, maquisard, mercenaire, morte-payé, palikare, pandour, papalin, partisan, patrouilleur, pertuisanier, pierrot, pionnier, pioupiou, piquier, planton, poilu, pourvoyeur, réquisitionnaire, résistant, reître, recrue, sentinelle, serviteur, soudard, soudrille, spahi, territorial, tirailleur, tommy, tourlourou, tringlot, trouffion, troupier, vétéran, fils, aîné, élève, citoyen, descendant, disciple, fieu, fiston, fruit, géniture, garçon, garçonnet, gars, gosse, grand, jeune, jouvenceau, postérité, progéniture, race, rejeton, sang, fille, adolescente, bachelette, bambine, belle-fille, blondinette, bouchon, brebis, brunette, célibataire, catherinette, colombe, coureuse, courtisane, demoiselle, descendante, donzelle, fille, fillasse, fillette, frangine, gamine, garçonne, gazille, gonzesse, héritière, hétaïre, jeunesse, jouvencelle, louloute, midinette, mignonne, nana, nénette, nonne, nymphette, petite, poulette, prostituée, quille, religieuse, rosière, soeur, soeurette, souris, suivante, tendron, trottin, typesse, bébé, baby, bambin, chiard, enfanton, enfant, gamin, môme, marmaille, marmot, mioche, mouffette, moutard, moutatchou, nourrisson, nouveau-né, petit, petit-salé, poupée, poupard, poupon, têtard, mari, époux, conjoint, épouse, bourgeoise, conjointe, déesse, doudou, ménagère, mariée, mousmé, m, mr, mme, mrs, miss **parole** accent, marcher, écorcher, émouvoir, énoncer, apprendre, argot, articulation, articuler, avouer, babiller, bafouiller, bégayer, balbutier, baragouiner, bavarder, blaguer, bredouiller, calomnier, causer, chevroter, chuchoter, citer, communiquer, confabuler, conférer, consulter, converser, débattre, débit, débiter, déblatérer, déclamer, déroiser, dénoncer, déviser, dialecte, dialoguer, diction, dire, discourir, discuter, dissenter, enregistrer, entretenir, exprimer, fasciner, franc-parler, frapper, giberner, gueuler, haranguer, idiome, intervenir, jaboter, jacter, jargon, jargonner, jaser, jaspiner, jobelin, langage, langue, médire, marmotter, marmotter, mentionner, murmurer, nasiller, nommer, palabrer, parlementer, parler, parole, pérorer, parler, patois, phonation, plaie, pratiquer, prononcer, prononciation, rabâcher, raconter, radoter, révéler, relater, retracer, rognonner, abandonner, adresser, épancher, confier, expliquer, style, usage, usance, verbe, vociférer, aboyer, acclamer, accuser, éclater, égossiller, annoncer, apostropher, appeler, attraper, avertir, bêler, beugler, brailler, braire, bramer, bruire, chahuter, chanter, clabauder, clamer, conspuer, couiner, craquer, criailler, croasser, feuler, gémir, glapir, grincer, grogner, gronder, hêler, houpper, hucher, hul-

uler, hurler, implorer, interpellier, invectiver, invoquer, jurer, manifester, meugler, mugir, pépier, piailler, plaindre, pleurer, pousser, prévenir, prier, proclamer, préférer, protester, publier, réclamer, répandre, réprimander, retentir, rouspéter, rugir, écrier, époumoner, fâcher, récrier, sermonner, exclamer, signaler, indigner, supplier, tancer, tempêter, tonitruer, tonner, vagir **pensée** faire, convaincu, avis, persuadé, admettre, élaborer, étudier, aviser, évoquer, cogiter, combiner, concevoir, conjecturer, considérer, contempler, délibérer, devoir, envisager, espérer, estimer, examiner, faillir, attention, gamberger, idée, imaginer, juger, méditer, manquer, mûrir, ordonnancer, pensée, peser, présumer, prévoir, prendre, projeter, rêvasser, rêver, réfléchir, raisonner, rappeler, recueillir, regarder, repenser, rouler, saisir, scruter, concentrer, douter, figurer, sentir, préoccuper, proposer, représenter, souvenir, targuer, intéresser, occuper, songer, souhaiter, soupçonner, spéculer, supposer, trouver **nutrition** absorber, alimentation, attaquer, avaler, bâfrer, becuquer, bouffe, bouffer, boulotter, boustifailleur, brichetonner, briffer, brouter, chipoter, claquer, collationner, consommer, consumer, corroder, croûte, croûter, croquer, croustiller, cuisine, débrider, décolorer, déguster, déjeuner, dîner, dépenser, déteindre, dévorer, dilapider, dissiper, engloutir, engouffrer, entamer, disparaître, faner, festoyer, fricasser, fritoter, friper, gaspiller, goûter, gober, gobichonner, godailler, goinfrer, grailleur, grappiller, grignoter, gruger, gueuletonner, ingérer, ingurgiter, mâcher, manger, mastiquer, mets, mettre, mordre, nourrir, nourriture, oublier, pâître, pâture, picorer, pignocher, pitance, prodiguer, repas, ripailler, ronger, ruiner, alimenter, sauter, savourer, bourrer, décarêmer, gaver, goberger, gorger, lester, empiffrer, emplir, piffrer, régaler, rassasier, remplir, repaître, restaurer, sustenter, souper, sucer, tâter, tordre, tortiller, tortorer, toucher, transgresser **émotions** amour, ébullition, agité, agitation, agressif, agressivité, aigreur, algarade, animosité, éréthisme, atrabilaire, atrabile, bile, bilieux, bouffée, bourrasque, chagrin, colérique, courroucé, courroux, crise, déchaînement, effervescence, emportement, exaspéré, exaspération, excitation, explosion, fâcherie, foudre, fulminant, fureur, furie, furieux, grogne, haine, hargne, hargneux, humeur, impatience, impatient, indignation, irascibilité, ire, irrité, irritabilité, irritable, mécontentement, monté, péché, querelle, rage, rageur, rancoeur, représsailles, rogne, scène, surexcitation, susceptibilité, tempête, vengeance, violence, accouplement, adoration, adultère, affect, affection, altruisme, amativité, amitié, amourette, ange, aphrodite, archerot, éros, association, attachement, attraction, aventure, babiole, badinage, bagatelle, béguin, baibatifolage, biquet, biquette, bluette, bouillonnement, bricole, caprice, chaleur, charité, coït, coeur, concubinage, concupiscence, conquête, copulation, coquetterie, cupidon, débauche, délicatesse, désir, dévotion, dévouement, dilection, engouement, entente, enthousiasme, estime, faible, fanatisme, fantaisie, ferveur, fièvre, flamme, fleurette, flirt, folie, fréquentation, fraternité, galanterie, goût, grâce, hyménée, hymen, idolâtrie, inceste, inclination, intérêt, intrigue, ivresse, lascivité, libertinage, luxure, maladie, mariage, marivaudage, mouvement, mysticisme, passade, passion, passionnette, penchant, philanthropie, piété, plaisir, pulsion, relation, rut, sens, sensibilité, sentiment, tendance, tendresse, toquade, touche, union, vénération, vénus, agrément, allégresse, amusement, épanouissement, ardeur, avantage, béatitude, bien-être, bienfait, bonheur, consolation, contentement, délice, douceur, enchantement, enjouement, entrain, euphorie, exaltation, extase, exultation, félicité, fierté, folichonnerie, gaieté, gaité, griserie, hilarité, jouissance, lies-plaisir, régal, réjouissance, ravissement, rayonnement, rigolade, sourire, abandon, abattement, accablement, affliction, amertume, éploement, épreuve, assombrissement, asthénie, bourdon, cafard, calamité, découragement, dépression, désabusement, désenchantement, désespérance, désespérer, désespoir, désolation, deuil, douleur, grisaille, laideur, lassi-

## Appendix A. Appendices

tude, lypémanie, mal, malaise, maussaderie, mocheté, monotomie, navrement, neurasthénie, noir, noirceur, nostalgie, nuage, pauvreté, peine, platitude, sévérité, sombreur, souffrance, spleen, suspens, uniformité, affres, alarme, attente, crainte, désarroi, incertitude, inquiétude, peur, souci, transe, abattu, adjudicataire, égotant, aliéné, alité, altéré, anéanti, anormal, atteint, avarié, cacochyme, chétif, cinglé, client, démoli, déprimé, dérangé, désaxé, détraqué, dingue, dolent, fada, fanatique, fatigué, fiévreux, fou, galeux, gâté, gâteux, grabataire, incommode, indisposé, infirme, ladre, lépreuse, lépreux, maboul, maladié, malsain, mauvais, morbide, pâle, patient, patraque, perturbé, piqué, rachitique, raide, révolutionné, scrofuleux, secoué, sonné, souffrant, souffreteux, tifosi, timbré, toqué, traumatisé, valétudinaire, abhorrer, abominer, éructer, évacuer, chasser, cracher, débêcher, dégobiller, dégorger, dégueuler, détester, excréter, expectorer, expulser, fulminer, gerber, haïr, honnir, huer, lancer, mépriser, maudire, réurgiter, répudier, rejeter, rendre, gorge, restituer, siffler, souffler, apeurer, cajoler, déconcerter, désespérer, effaroucher, effrayer, embarrasser, gêner, geler, glacer, impressionner, inhiber, inquiéter, menacer, paralyser, terroriser, ébauché, angoissé, approximatif, capon, complexe, confus, craintif, démerdard, discret, douteux, effarouché, effarouchable, effrayé, embarrassé, farouche, flottant, flou, frileux, froussard, fumeux, gêné, gauche, hésitant, honteux, humble, imperceptible, incertain, indéci, indéfini, indéterminé, indistinct, intimidé, lâche, modeste, nébuleux, nuageux, obscur, peureux, poltron, pudibond, pusillanime, réservé, rougisant, sauvage, subtil, timoré, transi, vague, émoi, appréhensif, appréhension, circonspection, confusion, effacement, effarouchement, gêne, gaucherie, hésitation, honte, humilité, indécision, modestie, prémonition, prudence, pudeur, réserve, trac, ébranlé, affecté, affolé, alarmé, éméché, émotionné, éperdu, apitoyé, éploré, attendri, attristé, blessé, bondissant, bouleversé, bouleversant, déchiré, déchirant, empoigné, enflammé, excité, frappé, impressionné, inquiet, intéressé, ivre, pantelant, pathétique, remué, retourné, saisi, suffoqué, surexcité, touché, touchant, tremblant, écoeurément, affadissement, allergie, éloignement, accident, affligé, aigre, élégiaque, assombri, bidoche, bougon, boulet, bourru, cafardeux, chagriné, colère, coléreux, consterné, consternation, content, contrarié, contrariété, contrit, couenne, cuir, déboire, déception, déchirement, dégoût, dépit, dépité, déplaisir, désagrément, désappointement, désolé, détresse, douloureux, ennui, ennuyé, gémissant, grimaud, grincheux, grognon, hypocondriaque, hypocondrie, inconsolable, inquiet, larme, larmoyant, lugubre, mélancolie, malheur, maussade, misanthrope, misère, morne, morose, morosité, mortifié, navré, neurasthénique, pataqués, peau, peiné, pelage, pisse-vinaigre, plaintif, pleurard, pleureur, rabat-joie, rechigné, regret, revêche, sinistre, sombre, téguement, tourment, tracasserie, tribulation, triste, tristesse, anorexie, antipathie, aversion, satiété, blasement, démoralisation, déplaisance, excréation, haut-le-cœur, horreur, humiliation, inappétence, indigestion, mépris, mortification, nausée, répugnance, répulsion, ressentiment, affolement, affres, alerte, émotion, angoisse, anxiété, cauchemar, couardise, dédain, effroi, foire, frayeur, frisson, frousse, intimidation, lâcheté, malepeur, panique, pétotoche, phobie, pleurerie, poltronnerie, pusillanimité, saisissement, souleur, spectre, stupéfaction, terreur, timidité, trouille, venette, vertige, vigilance, préoccupation, ébahissement, abasourdissement, admiration, ahurissement, émerveillement, épatement, épouvante, étonnure, effarement, impression, pétrification, scandale, stupeur, surprise **foi** assurance, autorité, évangile, catéchisme, certitude, confession, confiance, conviction, créance, credo, croyance, déloyauté, dogme, droiture, espérance, exactitude, fausseté, fidélité, fourberie, franchise, hommage, honnêteté, honneur, loyauté, mystique, opinion, perfidie, promesse, religion, serment, sincérité, spiritualité, témoignage, vérité **tromper** éblouir, abuser, échauder, égarer, amuser, apaiser, arnaquer, étriller, éviter, bafouer, baisier, balancer, bercer, berner, blouser, bluffer, capter, caroter, circonvenir, cocuifier, coiffer, confondre, consoler, couil-

lonner, croire, décevoir, déguiser, déjouer, dépiter, désappointer, dissimuler, duper, embabouiner, emberlificoter, embobeliner, embobiner, empaumer, empiler, encorner, endormir, enfler, engeigner, enjôler, enquinauder, entôler, entortiller, entourlouper, entuber, envelopper, errer, escroquer, estamper, exploiter, cocu, cornard, marcher, miroiter, feindre, feinter, fichier, flatter, flouer, foutre, frauder, frustrer, gonfler, gourer, imposer, induire, jobarder, jouer, leurrer, méprendre, mentir, moquer, mystifier, pigeonner, piper, promener, refaire, renarder, repasser, resquiller, ruser, séduire, surprendre, trahir, tricher, truquer **mort** tuer, abattre, achever, échiner, écraser, égorger, allonger, anéantir, épuiser, éreinter, assassiner, assommer, étendre, étouffer, étrangler, étripier, bousiller, buter, casser, chagriner, chouriner, crever, décapiter, décimer, dégeler, dégringoler, démolir, dépêcher, détruire, descendre, empoisonner, emporter, estourbir, exécuter, excéder, exténuer, exterminer, cesser, mourir, fatiguer, faucher, flinguer, foudroyer, fusiller, gâcher, guillotiner, harasser, immoler, juguler, lapider, lasser, liquider, lyncher, massacrer, meurtrir, miner, moissonner, navrer, nettoyer, noyer, occire, peiner, pendre, percer, poignarder, pourfendre, ratiboiser, refroidir, sacrifier, saigner, défaire, servir, stériliser, suicider, supplicier, supprimer, suriner, trucidier, user, vanner, mort, écroulement, agonie, anéantissement, ankylosé, apathique, épuisé, éreinté, assassinat, éteint, évanoui, brisé, cadavre, camarde, chute, claqué, condamné, corps, crève, crevé, croupissant, décédé, décès, décomposition, défunt, délavé, dépouille, désert, détruit, cujus, destruction, disparition, disparu, dormant, effondrement, enterrement, enveloppe, esquinté, euthanasie, exécution, excédé, exténué, extinction, fade, fantôme, feu, fichu, figé, fin, fini, flapi, foutu, voyage, harassé, immobile, inanimé, inerte, inhabité, insensible, intimider, irrécupérable, languide, lessivé, macchab, macchabée, malemort, mânes, meurtre, mourant, néant, nécrosé, nuit, éternelle, ombre, parque, passé, perdu, perte, plat, plongeon, rétamé, recru, rendu, repos, éternel, restes, rompu, ruine, silencieux, sommeil, somnolent, stagnant, supplice, tombe, tombeau, torture, tranquille, trépas, trépassé, tué, usé, victime, vidé **guérir** échapper, adoucir, mieux, améliorer, calmer, châtier, cicatriser, corriger, couper, débarrasser, délivrer, désintoxiquer, estomper, tomber, pallier, réchapper, rambliner, relever, remédier, remettre, renaître, ressusciter, revivre, sauver, fermer, rétablir, soigner, soulager, traiter **dormir** coucher, dormir, dormailler, lambiner, lanterner, négliger, pioncer, ronfler, roupiller, sommeiller, somnoler **salutations** bonjour, adieu, bienvenue, bonsoir, salamalec, salut, ciao, bye, salutations **vouloir** accepter, accorder, acquiescer, affirmer, aimer, alléguer, ambitionner, appéter, aspirer, attendre, envie, briguer, commander, concéder, consentir, convoiter, craindre, décider, défendre, daigner, désirer, demander, disposition, entendre, envier, escompter, essayer, exiger, forme, gouverner, guigner, intention, interdire, lorgner, loucher, manière, nécessiter, objectif, ordonner, ordre, oser, permettre, pouvoir, prétendre, prescrire, résolution, recevoir, requérir, revendiquer, s'efforcer, s'essayer, soif, solliciter, soupirer, subir, tenir, tolérer, viser, vouloir **soupir, sanglot** soupir, plainte, sanglot, silence, remord, cri, gémissement, geignement, hoquet, jérémiade, larmoiement, pleur, pleurnichement, spasme **raison** allégation, équité, argumenter, argumentation, base, but, cause, cerveau, cervelle, compréhension, conception, connaissance, conscience, considérant, considération, couleur, dédommagement, démonstration, discernement, droit, entendement, esprit, excuse, explication, facultés, fondement, indice, intellect, intelligence, judiciaire, jugeote, justesse, justice, justification, lieu, logique, lucidité, lumière, méthode, mobile, modération, motif, occasion, office, origine, philosophie, pondération, pourquoi, prétexte, preuve, principe, principes, probabilité, probité, quoi, réfutation, raison, raisonnement, rapport, rationnel, rectitude, sagesse, satisfaction, tête, tact **métiers** abatteur, aboyeur, accessoiriste, accisien, acconier, accordéoniste, accordeur, accordeuse, accoucheur, accouveur, accouveuse, accréditeur, accréditeuse, accréditrice, accrocheur, acériculteur, acéricultrice, acheteur, acheteuse,

aciériste, aconier, acousticienne, acousticien, acrobate, acteur, actrice, actuaire, acuponcteur, acuponctrice, acupuncteur, acupunctrice, adaptateur, adaptatrice, adjudicateur, adjudicatrice, administrateur, administratrice, aède, aérodynamicien, aéronaute, aérostier, afficheur, affichiste, affileur, affineur, affineuse, affranchisseur, affréteur, affuteur, affuteuse, agencier, agencière, agioteur, agréateur, agréuse, agriculteur, agricultrice, agrobiologiste, agronome, aiguilleur, aiguillier, aiguiseur, aiguiseuse, ajusteur, ajusteuse, albatrier, alcade, alchimiste, alcoolier, alcoolologue, alderman, alem, aleseur, aléseuse, alfater, algébriste, algologue, alguazil, aliéniste, alisier, allergologiste, allergologue, allopathe, allumettier, allumettière, alphabétiseur, alphabétiseuse, alpiniste, alviste, amareyeur, amareyeuse, ambassadeur, ambassadrice, ambulancier, ambulancière, aménagiste, amidonnier, amidonnière, amodiateur, amphibie, amuseur, amuseuse, analyste, anatomiste, andrologue, anesthésiste, ânier, ânière, animalier, animalière, animateur, animatrice, annaliste, annonceur, annonceuse, annoncière, annoncier, annotateur, annotatrice, antenniste, anthropologue, antiquaire, aouteron, apiculteur, apicultrice, apiculateur, apicéuse, apatisseur, apothicaire, appareilleur, appareilleuse, appareteur, apponteur, apprentie, apprenti, apprêteur, apprêteuse, aquaculteur, aquacultrice, aquafortiste, aquarelliste, aquatintiste, aquiculteur, aquicultrice, arbalétrier, arbitrage, arbitre, arboriculteur, arboricultrice, archéologue, archer, archetière, archetier, archiâtre, architecte, archiviste, ardoisier, ardoisière, argenteur, argenteuse, argentier, argotiste, arithméticien, armallis, armateur, armoriste, armurier, aromaticienne, aromaticien, arpenteur, arpenteuse, arpète, arpette, arquebusier, arracheur, arrangeur, arrangeuse, arrêstetier, arrimeur, arrondissement, arroseur, arroseuse, artificier, artilleur, assainisseur, assemblagiste, assesseuse, assesseur, assistant, assistante, assureur, assyriologue, astaculteur, astacultrice, astrologue, astronaute, astronauticien, astronome, astrophysicien, atomiste, attorney, aubergiste, audencier, audit, auditeur, auditrice, aumônier, aurige, auriste, auteur, auteure, autocariste, automatique, automatique, autoursier, autrice, avaleur, avaleuse, aventurier, aventurière, aviateur, aviatrice, aviculteur, avicultrice, avionneur, avitailleur, avocat, avocate, avoué, avouée, bachoteur, badigeonneur, baes, baesine, bagagiste, baigneur, baigneuse, baillie, bailli, baladin, balancier, balayeur, balayeuse, baleinier, baliseur, ballerine, bandagiste, banquier, banquière, barbier, barbière, barde, barmaid, barman, barman, barragiste, barreur, barreuse, baryton, basculeur, basketteur, basketteuse, bassiste, bateleur, bateleuse, batelier, batelière, bâtisseur, bâtisseuse, bâtonnier, bâtonnière, batteur, batteuse, baudrier, bêcheur, bedeau, berger, bergère, bibliographe, bibliothécaire, bijoutier, bijoutière, billetiste, bimbeltière, bimbeltier, bineur, biothicienne, biothicien, biogéographe, biographe, biologiste, biométricienne, biométricien, biophysicienne, biophysicien, biscuitier, biscuitière, blanchisseur, blanchisseuse, bobineur, bobineuse, bobinière, bobinier, boiseur, boisselier, boitier, bombagiste, bombeur, bonne, bonnetier, bonnetière, bordier, bordière, botaniste, botteleur, botteleuse, bottier, bottière, boucanier, boucher, bouchère, boucheur, boucholeur, bouchonnier, bouchoteur, bouclier, boueur, bougnat, bouilleur, boulanger, boulangère, boulier, boulinier, bouquetier, bouquiniste, bourgmestre, bourreau, bourrelrière, bourrelleur, bourrier, boursier, boursière, bousilleur, bouteiller, boteur, boutilier, boutonner, boutonnière, bouvier, bouvière, boyeur, boyeuse, boyaudier, boyaudière, braconnier, brancardier, brancardière, brandevinière, brandevinier, brasseur, brasseuse, brillanteur, briqueteur, briquetier, brocanteur, brocanteuse, brocheur, brodeur, brodeuse, broker, bronzeur, bronzieuse, bronzière, bronzier, brosier, brosière, brouetteur, brouettier, broyeur, bruiteur, bruiteuse, brûleur, brunisseur, brunisseuse, buandier, buandière, bucheron, bucheronne, bucheur, buffetier, buffetière, bulbiculteur, bulbicultrice, buraliste, bureaucratienne, bureaucratien, burineur, buriniste, bustier, bustière, buvetier, cabaretière, cabaretier, cabinier, cabinotier, cableur, cableuse, cabliste, caboteur, cadreur, cadreuse, cafetier,

cafetière, caid, caissier, caissière, calandreur, calandreuse, califat, calicot, calier, calleur, calleuse, calligraphe, cambiste, cambreur, cambusier, camelot, cameraman, camerier, camerière, cameriste, camionneur, camionneuse, campaniste, canalisateur, canalisatrice, cancerologue, canetière, canissier, canetière, canneur, canneuse, cannier, cannière, cannissier, canonnier, canotier, canotière, cantatrice, cantinier, cantinière, cantonnier, cantonnière, cantor, canuse, canut, capilliculteur, capitaine, carabinier, cardeur, cardeuse, cardiologue, caricaturiste, carillonneur, carillonneuse, cariste, carpettier, carreleur, carreleuse, carrier, carrossier, cartier, cartographe, cartonier, cartonnière, cartons, cartooniste, cascadeur, cascadeuse, casernier, casquettier, casseur, casseuse, catcheur, catcheuse, catéchiste, caviste, cellière, celliste, censeur, censure, censière, censier, ceramiste, cerealier, chainetière, chainetier, chaîneur, chaîneuse, chainier, chainiste, chaisier, chaisière, chalcographe, chaloupier, chalutier, chambellan, chambrière, chambrier, chamelier, chamoiseur, chancelière, chandelier, changeur, changeuse, chansonnière, chansonnier, chanteur, chanteuse, chantre, chanvrière, chanvrier, chapelier, chapelière, charbonnier, charcutier, charcutière, charpentier, charretier, charron, chartrier, chasseur, chasublier, chasublière, chaudronnier, chaudronnière, chauffagiste, chauffeur, chafournier, chaumier, chasseur, chaussonnier, chef, cheminot, cheminote, chemisier, chemisière, chercheur, chercheuse, chevillard, chevilleur, chevière, chevrier, chiffonnière, chiffonnier, chiffreur, chiffreuse, chimiste, chiropracteur, chiropraticien, chirurgien, chirurgienne, chocolatier, chocolatière, chorégraphe, choriste, chromeur, chromiste, chroniqueur, cicérone, cigarettière, cigarettier, cigariier, cigarière, cimentier, cineaste, cireur, cireuse, cirier, cirière, ciseleur, ciseleuse, cithariste, clapman, clarinetiste, claveciniste, clavieriste, claviste, clerc, clicheur, clicheuse, climatologue, clinicienne, clinicien, cloutier, cloutière, clown, coache, coach, cocher, codeur, codeuse, codirecteur, codirectrice, coéditeur, coffreur, cogérante, cogérant, cogniticienne, coiffeur, coiffeuse, colleur, colleuse, coloriste, colporteur, coltineur, colzatier, comédien, comédienne, comique, commandant, commerçante, commerçant, commis, commise, commissaire, communarde, communard, communicateur, communicatrice, compagnon, compositeur, compositrice, comptable, comptaie, concepteur, conceptrice, concierge, conducteur, conductrice, confectionneur, confédération, confédencier, confiseur, confisseuse, confiture, confiturier, connecticienne, connecticien, conseiller, conseilère, conservateur, conservatrice, conserveur, conserveuse, consignataire, constable, consul, consule, consultant, consultante, contactologue, conteur, conteuse, contrebassiste, contremaitre, contrôleur, contrôleuse, convoyeur, convoyeuse, coolie, coordinateur, coordinatrice, coordonnateur, coordonnatrice, copilote, copiste, coprésidente, coprésident, coqueleuse, coquette, coquetier, corailleur, corailleuse, cordeur, cordeuse, cordier, cordière, cordonnier, cordonnière, cornac, cornemuseur, cornettiste, corneur, corneuse, corniste, coroner, correcteur, correctrice, correspondante, correspondant, corroyeur, corsaire, corsetier, corsetière, coryphée, cosmétologue, cosmologiste, cosmologue, cosmonaute, costumier, costumière, coteur, cotonnier, cotonnière, coupeur, coupeuse, courriériste, courrier, coursier, coursière, courtier, courtière, couseur, couseuse, coutelier, coutelière, couturier, couturière, couvreur, crassier, créancière, créancier, créateur, créatif, créatrice, crémier, crémère, créoliste, créprier, crêpière, crieur, crieuse, criminologue, crinier, cristallière, crocheteur, croupier, croupière, cryptographe, cueilleur, cueilleuse, cuiseur, cuisinier, cuisinière, cuisiniste, culotière, culottier, cultivateur, cultivatrice, curateur, curatrice, cytologiste, dactylographe, dalleur, damasquineur, dameur, danseur, danseuse, débardeur, débardeuse, débateur, débiteur, décapeur, décatisseur, décatisseuse, déchiqueteur, déchiqueteuse, décideur, décideuse, déclamateur, déclamatrice, décolleteur, décolleteuse, décorateur, décoratrice, découpeur, découpeuse, découvreur, découvreuse, décrocheur, décrocheuse, défendeur, défibreur, défibreuse, défricheur,

## Appendix A. Appendices

défricheuse, dégraisseur, dégraisseuse, dégustateur, dégustatrice, démarcheur, démarcheuse, déménageur, demandeuse, démineur, démographe, démolisseur, démonstrateur, démonstratrice, dendrologue, dentellière, dentellier, dentiste, déontologue, dépanneur, dépeceur, dépeceuse, dépollueur, dépoussiéreur, dermatologue, dermato, designer, dessinateur, dessinatrice, déstockeur, détacheur, detacheuse, détaillant, détaillante, détective, développeur, développeuse, devideur, diabolotin, diagnosticien, diagnostiqueur, dialectologue, dialoguiste, diamantaire, dictionnariste, diéséliste, diététicienne, diététicien, dinandier, dinandière, dindonnière, dindonnier, diplomate, directeur, directrice, discothécaire, discounter, disconteur, dispatcher, disquaire, distillateur, distillatrice, divette, docimologue, docker, docteur, doctoresse, documentaliste, documentariste, domestique, dominotière, dominotier, dompteur, dompteuse, donneur, doreur, doreuse, douanier, douanière, doubleuse, doubleur, doucheur, doucheuse, dramaturge, drapier, drapière, draveur, draveuse, dresseur, dresseuse, droguiste, duègne, duumvir, dynamiteur, dynamiteuse, ébarbeur, ébaucheur, ébéniste, éboueur, éboueuse, écailleur, écailleuse, écangueur, écangueuse, échanson, échenilleur, échevin, échetotier, échetotière, éclairagiste, éclairateur, éclairieuse, éclateur, éclusier, éclusière, écordeur, écologue, économètre, économétricien, économe, écorceur, écorcheur, écorcheuse, écrivain, écrivaine, écuireur, écuyer, écuyère, édile, éditeur, éditorialiste, éditrice, éducateur, éducatrice, effileur, effilocheur, effilocheuse, égoutier, égratigneur, élagueur, élagueuse, électricien, électricienne, électronicien, éleveur, éleveuse, émailleur, émailleuse, emballleur, embaumeur, embaumeuse, embouteilleur, embouteilleuse, embryologiste, embryologue, émondeur, émondeuse, émotteur, émotteuse, émouleur, émouleuse, empaillleur, empaillieuse, empaqueteur, empaqueteuse, empileur, empileuse, employé, employée, encadreur, encadreuse, encaisseur, encaveur, encaveuse, encodeur, encodeuse, encolleur, encolleuse, endocrinologue, endosseur, enfileur, enfileuse, enfourneur, engraisseur, enlumineur, enlumineuse, enqueteur, enquêtrice, enrouleur, ensacheur, ensacheuse, enseignant, enseignante, ensemblier, entraîneur, entraîneuse, entrepreneur, envoyeur, épaviste, épicier, épicière, épigraphiste, épilleur, épileuse, épinceteur, épinceteuse, épinceteur, épinceuse, épinglière, épinglier, épistolier, éplucheur, éplucheuse, équilibriste, équipier, équipière, ergonome, ergotherapeute, espion, espionne, essayeur, essayeuse, essayiste, estafier, estampeur, estampeuse, esthétique, esthéticien, estimateur, étalagiste, étalière, étalier, étalonneur, étalonneuse, étameur, étameuse, étampeur, étampeuse, ethnobotaniste, ethnographe, ethnologue, ethologue, étiqueteur, étiqueteuse, étireur, étireuse, expéditeur, expéditrice, expert, experte, fabuliste, façadier, facilitateur, facilitateur, façonneur, façonner, façonnière, facteur, factrice, fagoteur, fagoteuse, fagotier, fagotière, faïencier, faïencière, faneur, faneuse, farinier, farinière, faucheur, faucheuse, fauconnier, fendeur, fendeuse, ferblantier, fermier, fermière, ferrailleur, ferrailleuse, ferreur, ferronnier, feudiste, feuillagiste, ficelier, figurant, figurante, figuriniste, figuriste, finlandier, filandière, filateur, fildeferiste, fileur, fileuse, filmeur, filmeuse, financier, finisseur, finisseuse, fiscaliste, fleuriste, floriculteur, floricultrice, flotteur, flutiste, fonctionnaire, fondeur, fondeuse, fontainier, footballeur, forain, foraine, forceur, forestier, foreur, foreuse, forfaitiste, forgeron, formateur, formatrice, fossoyeur, fossoyeuse, fouacier, foudrier, fouteur, foulon, foulonnier, fourbisser, fournier, fourreur, fourreuse, fraiseur, fraiseuse, frappeur, frappeuse, freineur, fresquiste, freteur, frigoriste, fripier, fripière, friteuse, friturier, fromager, fromagère, fructiculteur, fruitière, fruitier, fumiste, funambule, gabarier, gabarrier, gabier, gâcheur, gâcheuse, gagiste, gainier, gainière, galeniste, galeriste, galibot, gallonier, galonnière, gambiste, gantier, gantière, garagiste, garanceur, garanceuse, garde, gardeur, gardeuse, gardien, gardienne, gargotière, gargotier, garnisseur, garnisseuse, gaucher, gaufreur, gaufreuse, gaveur, gaveuse, gazetier, gazetière, gazier, gazière, geisha, gommeur, gommeuse, gemmologue,

gendarme, généalogiste, généraliste, généticienne, généticien, genévrier, géodésienne, géodésien, géographe, géolier, géolière, géologue, géomètre, géophysicienne, géophysicien, gériatre, gérontologue, gestionnaire, giletier, giletière, gitologue, glaceur, glaceuse, glacier, glacière, glaciologue, glaneuse, gobeletière, gobeletier, goémonier, goémonière, golfeur, golfeuse, gondolier, gondolière, gonfalonier, gonfanonier, goudronneur, goudronneuse, goudronnier, gouteur, gouteuse, gouvernante, grainetier, grainetière, graineur, grainier, grainière, graisseur, grammairienne, grammairien, grammaticien, granitier, graphiste, graphologue, gratteur, gratteuse, gravatier, graveur, graveuse, gravier, gréeur, greffeur, greffeuse, greffier, greffière, greneur, grenier, grignoteur, grillageur, groom, grossiste, groupeur, groupeuse, grutier, grutière, guetteur, guichetière, guichetier, guide, guillotineur, guitariste, gymnaste, gynécologue, gypsier, habilleur, habilleuse, hagiographe, haleur, haleuse, handicapé, harengère, harmoniste, harmoniste, harnacheur, harpiste, haubergeon, haveur, heaumier, héliculteur, hélicultrice, hélicier, héliographe, héliographeuse, hépatologue, héraur, herbager, herbiste, hercheur, herscheur, hippologue, historien, historienne, hockeyeur, hockeyeuse, hongreur, hongroyeur, horloger, horlogère, horticulteur, horticultrice, hortillon, hote, hôtelier, hôtelière, hotesse, hotteur, hotteuse, houppier, huchet, huchier, huillier, huissier, huissière, humoriste, hybrideur, hybrideuse, hypnologue, hypnotiseur, hypnotiseuse, ichtyologiste, ichtyologue, iconographe, iconologiste, iconologue, illusionniste, illustrateur, illustratrice, ilotier, ilotière, imageur, imagier, imagière, imitateur, imitatrice, imposeur, impresario, imprimeur, imprimeuse, incrusteur, incrusteuse, indexeur, indexeuse, indianiste, industrielle, industriel, infectiologue, infirmier, infirmière, infographiste, informaticien, ingénieur, ingénieur, inséminateur, inspecteur, inspectrice, installateur, installatrice, instituteur, institutrice, instructeur, instructrice, instrumentiste, intendante, intendat, intermittent, interprète, introducteur, ivoirier, ivoirière, jardinier, jardinière, jardinst, jaugeur, joaillière, joaillier, jockey, jongleur, jongleuse, journalier, journaliste, juge, junior, jupier, jupière, jurat, juré, juriste, kinesiologue, kinésithérapeute, kiosquière, kiosquier, laborantin, laborantine, laboureur, laceur, laceuse, lâcheur, lad, laineur, laineuse, lainier, lainière, laitier, laitière, lamaneur, lamier, lamineur, lamineuse, lampier, lampiste, lancier, lanternier, lapicide, lapidaire, laquais, laqueur, laqueuse, latiniste, lavandière, laveur, laveuse, layetier, layeur, lecteur, lectrice, legionnaire, légiste, légumier, lessiveur, lessivier, lettré, lettré, lettré, leveur, leveuse, levurie, levurier, lexicographe, lexicologue, libraire, librettiste, licier, liègeur, liègeuse, lieu, lieuse, liftier, liftière, limeur, limeuse, limonadier, limonadière, linger, lingère, linguiste, linier, linière, linotypiste, liquoriste, liseur, liseuse, lisseur, lisseuse, lissier, lithographe, livreur, livreuse, logisticien, logisticienne, logographe, logopède, logothète, lotisseur, loueur, loueuse, loutetier, ludologue, ludothécaire, lunetier, lunetière, lunetier, lunetière, lustreur, lustreuse, luthier, luthière, luthiste, lutteur, lutteuse, machiniste, maçon, maçonne, magasinier, magasinière, magicien, magicienne, magistrat, magistrate, magnanarelle, magnanière, magnanier, magnanelle, magnétiseur, magnétiseuse, maieure, maieur, maieuticien, maître, maîtresse, malletière, malletier, malteur, maltotier, mammalogiste, manadier, manager, manageur, manageuse, manieur, manipulateur, manipulatrice, mannequin, manoeuvre, manoeuvrière, manoeuvrier, manucure, manufacturier, manufacturière, maquettiste, maquignon, maquilleur, maquilleuse, maraicher, maraichère, marathonnienne, marathonnier, marbreur, marbreuse, marbrier, marcaire, marchand, marchande, marchandiseur, marcheur, mareyeur, mareyeuse, margarinier, margeur, margeuse, margoulin, marieur, marieuse, marin, marinier, marinière, marionnettiste, marketeur, marketeuse, manieur, maroquinier, maroquinère, marqueteur, marqueur, marqueuse, marteur, masseur, masseuse, massier, mastologue, matelassier, matelassière, matelot, mateur, mathématicien, matrone, mayeure, mayeur, mécanicien, mécanicienne, mécanographe, médaillier, médecin, médiateur, médiatrice,

mégissier, mélangeur, mélangeuse, mélodiste, menager, menagère, ménagiste, ménestrel, ménétrier, menuisier, menuisière, mercaticien, mercaticienne, marchandiser, mercier, mercièr, messenger, mesureur, mesureuse, métallier, métallière, metallographe, métallurgiste, métayère, métayer, météorologiste, météorologue, mètreur, mètreuse, métrologiste, métrologue, meulier, meunier, meunière, militaire, mime, minerviste, mineur, miniaturiste, ministre, minotier, minotière, mireur, mireuse, miroitière, miroitier, missilier, mitron, mixeur, modèle, modeleur, modeleuse, modélisateur, modélisatrice, modéliste, modiste, moireur, moireuse, moissonneur, moissonneuse, moniteur, monitrice, monnayer, monteur, monteuse, mortaiseur, morutier, morutièr, mosaïste, motociste, motoriste, mouleur, mouleuse, moulineur, moulineuse, moulinier, moulinière, mouliste, moulurier, moulurière, mousse, moutardier, moutonnier, muezzin, mulassier, muletier, muletière, muséographe, musicien, musicienne, musicologue, mycologue, myrmecologue, mytilculteur, mytilcultrice, nageur, nageuse, naisseur, natier, natière, naturaliste, nautonier, nautonière, navigateur, navigatrice, négociant, négociant, négociateur, neoniste, néphrologue, netsurfeur, netsurfeuse, nettoyeur, nettoyeuse, neurologue, nez, nielleur, nielleuse, nivologue, nocher, notaire, notateur, notatrice, noueur, nourrice, nourrisseur, novice, noyauteur, nuitouse, nurse, nutritionniste, obstétricienne, obstétricien, océanote, oculariste, oculiste, odontologiste, oenologue, officier, officière, oiseleur, oiseuse, oiselier, oiselière, oleiculteur, oleicultrice, oliveur, oliveuse, ombudsman, omnipraticien, oncologiste, oncologue, onirologue, opérateur, opératrice, ophtalmologue, opticien, opticienne, optométriste, orchestrateur, orchestratrice, ordinatricienne, ordinatricien, orfèvre, organicien, organicienne, organier, organière, organisateur, organisatrice, organiste, orienteur, orienteuse, ornemaniste, ornithologue, orpailleur, orpailleuse, orthodontiste, orthopédiste, orthophoniste, orthoptiste, ostéopathe, ostéopraticien, ostréiculteur, ostréicultrice, otorhino, outilleur, outplacer, outplaceur, outplaceuse, ouvreur, ouvreuse, ouvrier, ouvrière, oxycoupeur, packager, packeur, pailler, pailleur, pailleuse, palefrenière, palefrenier, paléographe, paléologue, paléontologue, palettiseur, palissonneur, paloteur, paludier, paludière, paludologie, palynologue, panetier, panseur, panseuse, pantomime, pantouffière, pantoufflier, papetier, papetière, papillonneur, parachutiste, parasitologue, pareur, pareuse, parfumeur, parfumeuse, parolier, parolière, parqueteur, parqueur, parqueuse, parquier, parurier, parurière, passementier, passementière, passeur, pastelliste, pasteur, pasteur, pasteurisateur, pastilleur, pastilleuse, pastourelle, patineur, patineuse, pâtissier, pâtissière, pâte, patron, patronne, patronnier, patronnière, pattière, pattier, paveur, paveuse, paysagiste, paysan, paysanne, péager, péagère, péagiste, peaussier, pêcheur, pêcheuse, pédiatre, pédicure, pédologue, pédopsychiatre, peigneur, peigneuse, peignier, peintre, pelletier, pelletière, pendulier, peon, pepinieriste, percepateur, perceptrice, perceur, perceuse, perchiste, perchman, perforateur, perforatrice, perliculteur, perlicultrice, perlier, perlière, perruquière, perruquier, personnage, peseur, pétrisseur, pétrisseuse, pétrographe, pétrolier, pétrolière, pharmacienne, pharmacien, philologue, phlébologue, phoniatre, photographe, photographeur, physicienne, physicien, physiologiste, physiologiste, phytographeur, pianiste, picador, pigiste, pileur, pileuse, pilote, pipier, pipière, piqueur, piqueuse, piroguier, pisciculteur, pisteur, pisteuse, pistoleur, pizzaiolo, placeur, placeuse, placier, placière, plagiste, planétologue, planeur, planteur, planteuse, plaquiste, plasticien, plasticienne, plasturgiste, plâtrier, plieur, plieuse, plisseur, plisseuse, plombier, plombier, plongeur, plongeuse, plumassière, plumassier, plumeur, plumeuse, pneumologue, pocheuse, podiatre, podologue, poëlier, poëlière, poète, poétesse, poinçonneur, poinçonneuse, pointeur, pointeur, pointeuse, poissonnier, poissonnière, policeman, policier, policière, polisseur, polisseuse, pompeur, pompier, pompière, pompiste, ponceur, ponceuse, pontier, pontière, pon-

tonnier, pope, populiculteur, populicultrice, porcelainier, porcelainière, porchèr, porcher, porion, portefaix, porteur, porteuse, portier, portière, portraitiste, poseur, poseuse, posticheur, posticheuse, postier, postière, potier, potière, praticien, praticienne, précepteur, préceptrice, préfet, préfète, préparateur, préparatrice, préposé, préposée, présentateur, présentatrice, président, présidente, presseur, presseuse, pressier, prêteur, prêtre, prévisionniste, prévôt, primeur, primeuriste, principal, principale, priseur, priseuse, privée, privé, procureur, procureuse, producteur, productrice, professeur, professeuse, profleur, programmeur, programmatrice, programmeur, programmeuse, projeteur, projeteuse, promoteur, promotrice, prospecteur, prospectrice, prote, prothésiste, prototypiste, proviseur, psy, psychanalyste, psychiatre, psychologue, publicitaire, puériculteur, puéricultrice, puisatier, pupitreur, pupitreuse, pyrotechnicien, pyrotechnicienne, qualicien, qualicienne, questeur, quincailleur, quincaillière, rabatteur, rabatteuse, rabbin, raboteur, raboteuse, raccommodeur, raccommodeuse, racleur, radariste, radio, radioastronome, radiologue, radionavigant, raffineur, raffineuse, ramasseur, ramasseuse, ramendeur, ramendeuse, rameur, ramoneur, rapsode, raseur, raseuse, rateleur, rateuse, rauteur, ravaleur, ravaudeur, ravaudeuse, ravitailleur, rayonneur, réalisateur, réalisatrice, réassureur, rebouteuse, recenseur, recenseuse, réceptionnaire, réceptionniste, receveur, receveuse, recherchiste, récolteur, récolteuse, recors, recruteur, recruteuse, recteur, rectifieur, rectifieuse, rectrice, récupérateur, récupératrice, rédacteur, rédactrice, redresseur, redresseuse, régent, régente, régisseur, régisseuse, registraire, régleur, régleuse, regrattière, regrattier, régulateur, régulatrice, rejointoyeur, rejointoyeuse, relationniste, relecteur, relectrice, releveur, relieur, relieuse, remailleur, remailleuse, remisier, remisière, remmailleur, remmailleuse, remonteur, remouleur, remouleuse, rempailleur, rempailleuse, remplisseur, remplisseuse, rentoileur, rentoileuse, rentrayeur, rentrayeuse, renvideur, renvideuse, réparateur, réparatrice, repasseur, repasseuse, répétiteur, répétitrice, reporter, reporteur, reportrice, représentante, représentant, repriseuse, réserviste, résinière, résinier, responsable, restaurateur, restauratrice, rétamateur, retordeur, retoucheur, retoucheuse, réviser, réviser, rhabilleur, rhabilleuse, rhéteur, rhumatologue, rinceuse, riveur, riveuse, riziculteur, rizicultrice, robeuse, robinetier, roboticien, roboticienne, rocailler, rôdeur, rogneur, rogneuse, romancier, romancière, romaniste, rosier, rosier, rôtisseur, rôtisseuse, rouennier, rouennière, rouleur, roulier, routeur, routeuse, routier, routière, rubanier, rubanière, rugbyman, sableur, sableuse, sabotier, sabotièr, sabreur, sacristain, sagard, saigneur, saigneuse, saleur, saleuse, saliculteur, salicultrice, salinier, salinière, salmoniculteur, salonnier, salonnière, salpêtrière, salpêtrier, sandalier, sandalière, santonnier, santonnière, sapeur, sapiteur, sardinier, sardinière, sarodiste, sasseur, sasseuse, satineur, satineuse, saucier, saucière, saunier, saunière, saurisseur, saurisseuse, sauveteur, savetier, savetièr, savonnier, savonnière, saxophoniste, sbire, scaphandrier, scénariste, scénographe, scieur, scieuse, scoliaste, scribe, scripte, sculpteur, secouriste, secrétaire, sellier, semeur, semeuse, séranceur, séranceuse, serriste, serrurier, serrurière, sertisseur, servant, servante, serveur, serveuse, sexologue, champouineur, champouineuse, sherif, sherpa, signaleur, sismologue, sitologue, sociologue, soignante, soignant, soigneur, soigneuse, solier, soliste, sommelier, sommelière, sondeur, sondeuse, sonneur, soudeur, soudeuse, souffleur, soufreur, soufreuse, souscripteur, souscriptrice, soutier, spadassin, spationaute, speakerine, speaker, spéléo, spéléologue, spéléonote, stadier, stadière, staffeur, standardiste, stationnaire, statisticienne, statisticien, stéarinier, stenciliste, steno, sténodactylo, sténographe, sténotypiste, steward, stomatologiste, stomatologue, storiste, stripteaseur, stripteaseuse, stucateur, stylicien, stylicienne, styliste, sujet, superviseur, surveillant, surveillante, sylviculteur, sylvicultrice, tabaculteur, tabacultrice, tabletier, tabletière, tacheron, taillandier, tailleur, tailleuse, talonneur, tamiseur, tamisier, tamisière, tanneur, tanneuse, tapis-



dériveur, embarcation, felouque, jonque, navire, paquebot, péniche, périssoire, pirogue, radeau, rafiote, steamer, steamboat, skiff, submersible, transatlantique, trirème, vaisseau, vapeur, voilier, aérodrome, aéronef, avion, aéroscaphe, airbus, appareil, avion-taxi, bac, aérien, bombardier, canadien, charter, chasseur, avion, giravion, hydravion, jet, tacot, taxi, transport, U.L.M., voiture, attelage, auto, berline, cabriolet,

carriole, charrette, fardier, fourgon, train, véhicule, locomotive, automotrice, coucou, locomotrice, machine, motrice, bagnole, chignole, coche, fiacre, limousine, roadster, tire, tramway, vélo, deux-roues, bécanne, bicyclette, moto, cyclomoteur, enduro, engin, motocycle, motocyclisme, scooter, solex, motocyclette, vélomoteur, embarquement

### A.4.3 OpeNER classification results

Entity	<i>OpeNER</i> result	Reference
Amérique	place	place
AUBERT	character	N/A
Aubert	place	N/A
Beau-Minon	place	character
Camille	place	character
Dieu	N/A	other
Élisabeth	place	character
LAMBERT	character	character
Lambert	place	character
Louise	N/A	character
M. de Réan	N/A	character
Madeleine	place	character
Mme d'Aubert	N/A	character
Mme de Réan	N/A	character
Paris	place	place
Paul	character	character
Réan	place	N/A
Sophie	place	character

**Table A.3** – Benchmarked *OpeNER* results on “*Les Malheurs de Sophie*”

## Appendix A. Appendices

---

Entity	<i>OpeNER</i>	Reference
Adèle	N/A	character
Afrique	N/A	place
Dieu	character	other
Gautruche	place	character
Germinie	place	character
Joseph	N/A	character
Jupillon	place	character
M. de Varandeuil	N/A	character
Mlle	place	N/A
Mlle de Varandeuil	N/A	character
Mme Jupillon	character	character
Montmartre	N/A	place
Paris	place	place
Révolution	N/A	other
Sempronie	place	character
Varandeuil	place	N/A

**Table A.4** – Benchmarked *OpeNER* results on “*Germinie Lacerteux*”



## A.5 Appendix E

### A.5.1 Attribution of chapters to originating books with LIWC features

=== Summary ===

Correctly Classified Instances	196	65.9933 %
Incorrectly Classified Instances	101	34.0067 %
Kappa statistic	0.6154	
Mean absolute error	0.0764	
Root mean squared error	0.2237	
Relative absolute error	43.1794 %	
Root relative squared error	75.231 %	
Total Number of Instances	297	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.722	0.05	0.667	0.722	0.693	0.958	1
	0.685	0.058	0.725	0.685	0.705	0.935	2
	0.759	0.026	0.759	0.759	0.759	0.967	3
	0.524	0.036	0.524	0.524	0.524	0.898	4
	0.667	0.033	0.667	0.667	0.667	0.922	5
	0.741	0.044	0.625	0.741	0.678	0.956	6
	0.467	0.018	0.583	0.467	0.519	0.946	7
	0.739	0.022	0.739	0.739	0.739	0.97	8
	0.617	0.068	0.63	0.617	0.624	0.925	9
	0.5	0.029	0.529	0.5	0.514	0.925	10
Weighted Avg.	0.66	0.044	0.66	0.66	0.659	0.94	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
26	0	0	2	0	1	2	2	3	0	a = 1
1	37	2	2	1	3	0	0	5	3	b = 2
2	3	22	0	0	0	0	0	1	1	c = 3
0	0	0	11	2	2	1	2	2	1	d = 4
0	2	0	0	18	3	0	0	3	1	e = 5
2	0	0	2	3	20	0	0	0	0	f = 6
3	0	2	0	0	1	7	0	2	0	g = 7
1	0	0	2	0	0	1	17	1	1	h = 8
2	4	2	1	3	2	1	2	29	1	i = 9
2	5	1	1	0	0	0	0	0	9	j = 10

## A.6 Appendix F

### A.6.1 Questionnaire aux écrivains ayant évalué *Saga+*

**1) Pensez-vous que vous pourriez composer des romans plus "optimaux" en utilisant un outil comme *Saga+*? Quelle serait pour vous la dimension à optimiser ?**

**A1** Je pense qu'il faudrait que je me familiarise davantage avec l'outil *Saga+* pour que la lecture des représentations graphiques de mes écrits puisse réellement en orienter le développement. Pour l'instant, à supposé que je puisse vraiment définir ce que signifie "optimal" en création littéraire (optimal par rapport à quels buts? Meilleure composition générale, équilibre, harmonie, confort cognitif du lecteur-trice, effets, impacts sur le lecteur-trice?... ) je ne me sens pas encore, après une journée d'utilisation, en mesure d'évaluer les qualités d'un texte ou d'une structure de textes à fragments multiples d'après les graphiques de *Saga+*. (disons que pour l'instant ça ne pourrait pas remplacer l'avis d'un lecteur humain, le retour sur son expérience de lecture).

**A2** Je ne sais pas s'il est possible d'« optimiser » la production d'une œuvre d'art, mais il est certain que *Saga+* offre une possibilité d'analyse synoptique qui peut être très utile pour améliorer un texte dont la longueur et la richesse de l'intrigue excéderaient en partie les capacités cognitives du rédacteur. Il permet de prendre conscience, grâce aux visualisations proposées, des « lieux » textuels qui poseraient un problème de cohérence, vis-à-vis de la présence des personnages ou des relations entre ceux-ci, afin de trouver les « chevilles » adéquates et ne pas perdre les lecteurs en route.

Pour ce qui est des lieux géographiques traversés par une intrigue romanesque, l'outil me semble particulièrement pertinent. Pour les personnages, un outil de résolution-interprétation des pronoms me semblerait utile dans le cas d'un roman impliquant des instances qui s'exprimeraient en première personne : lorsque une telle solution informatique sera développée et fonctionnelle, ce qui ne tardera sans doute pas à arriver, son intégration à *Saga+* ne pourra qu'affiner l'outil. De manière générale toutes les innovations qui concernent l'analyse informatique des textes qui pourront être ajoutée au programme

actuel seront une aide précieuse pour autant que l'interface – actuellement simple et intuitive – ne se trouve pas surchargée de fonctions complexes à paramétrer et prendre en main ; cet aspect de sobriété et de convivialité de l'interface me semble essentiel à la mise au point d'un outil qui puisse être largement utilisé par des auteurs qui ne sont pas forcément des champions en matière de technologie(s) et d'informatique.

**A3** Oui

Pouvoir observer le traitement de différents paramètres plus fins (que la distribution des personnages) au fil du texte, comme par exemple les émotions, les variations d'intensité narrative, etc.

**A4** Je pense que l'outil peut aider à composer différemment, à faire des expériences, mais ça ne voudra pas dire que le résultat sera nécessairement meilleur, juste que ça peut être une aide à la manoeuvre / la tentative / l'essai

**2) *Saga+* introduira bientôt une fonction de "Reading analytics" qui permet d'observer en temps réel comment les textes publiés sur la plateforme sont lus. Pensez-vous que voir en détail la manière dont votre texte est lu pourrait avoir une influence sur votre écriture, et pourquoi ?**

**A1** Oui, dans une certaine mesure. ça m'intéresserait sans conteste de savoir comment se comporte le lecteur. Comment il chemine dans mon texte.

Après, je m'interroge...

– Comment interpréter le comportement ? Par exemple, est-ce qu'une accélération de lecture représente forcément un signe positif ? Doit-on y voir le signe que le lecteur est pris par l'histoire, qu'il y a une tension, un suspens qui le pousse à connaître la suite, ou doit-on y voir le signe que le lecteur pense à autre chose et glisse sur les phrases, vire accélère parce qu'il s'ennuie et il veut passer à la suite ? De même, un lecteur qui lit deux fois un passage : a-t-il mal compris (il faudrait alors l'interpréter comme un signe négatif) ou a-t-il tellement aimé la teneur (poétique, stylistique,

philosophique) du passage qu'il veut la savourer, s'en imprégner une nouvelle fois ?

Est-ce qu'avec de tels instruments on n'est pas en train de limiter notre vision de la littérature au degré d'efficacité de la trame, de la structure (genre policier, pour schématiser), sans prendre en compte d'autres composantes importantes, comme le style ou les résonances créées chez le lecteur ? Ça me semble quand même très délicat de tirer des conclusions à partir de signes totalement extérieurs et factuels de comportement, car l'essentiel va se passer dans la sphère émotionnelle et sensitive du lecteur, et à ces sphères, je ne vois pas comment l'outil de "Reading analytics" va accéder ?? (Il faudrait des électrodes capteurs de plaisir de lecture :)

– d'un point de vue purement éthique je ne suis pas complètement à l'aise avec l'idée d'épier le lecteur dans ses comportements. je me dis que mon plaisir quand je lis c'est l'intimité stricte avec l'oeuvre, la pensée de l'auteur, et me savoir observée (si je me mets en tant que lecteur analysée) me serait désagréable.

- A2 Bien sûr que cette fonction est utile, néanmoins – mais je suis certain que cela a déjà été envisagé – il faudrait que ces statistiques de lecture puissent exclure les clics de curiosité sur une page qui ne serait pas vraiment lue et que la plateforme accueille suffisamment de lecteurs pour que ces mesures soient absolument pertinentes. Ceci résolu, pouvoir visualiser les chapitres où « lâcheraient » les lecteurs, les endroits du texte ou leur compréhension serait mise à une telle épreuve qu'ils en abandonneraient leur lecture, permettrait de retravailler cette partie du texte. Cela rendrait possible une forme de « critique statistique ». Il ne faut néanmoins pas perdre de vue que, via *Saga+* seuls les aspects d'organisation et de cohérence, c'est-à-dire la logique du récit peuvent être envisagés ; l'écriture elle-même (sa stylistique, son lexique, ses tournures, sa rhétorique [figuration, choix des images]) compte aussi beaucoup dans la poursuite ou non d'une lecture : cette question, fondamentale, de goût et d'esthétique ne me semble pas pouvoir être résolue par une application informatique ; bien sûr, en résolvant les points qui relèvent de la logique du récit, l'outil permet à l'auteur de mieux se concentrer sur les aspects non pris en charge par le programme. Malheureusement, il me paraît

impossible de déterminer avec ce genre d'outil si le lecteur abandonne pour des raisons relatives au récit structurellement considéré ou justement pour des raisons liées aux thématiques abordées ou encore à la facture stylistique (registre de langue, etc.).

- A3 Plutôt non, les types de lecture et les voies de l'interprétation sont innombrables; et c'est en soi que se trouvent les clés de l'écriture plutôt que dans les avis des lecteurs. Ce qui ne signifie pas que l'on écrit pour soi seul.
- A4 À court terme, je dirais que non, parce que le texte qui est analysé est déjà écrit et que pour ma part en tout cas je n'imagine pas le changer et que pour les suivants, les analyses ne s'appliqueront pas. Mais il n'est pas exclu qu'à long terme, avec l'expérience venant et une meilleure compréhension de ce qu'indiquent ces résultats, ça puisse avoir une influence.

### 3) Quels sont les fonctionnalités les plus pertinentes pour votre façon de travailler qui manquent encore à *Saga+* ?

- A1 J'en ai déjà parlé un peu le dimanche 13 à Montricher : disons que dans ma façon actuelle de travailler, je suis plus à la recherche de fonctionnalités qui m'aident à maîtriser le monde fictif que je suis en train de construire. J'ai certains besoins en particulier dans le roman, ou dans certaines pièces de théâtre, disons chaque fois que le travail implique d'édifier un cosmos fictif complet, qui doit se tenir. Dans un texte tel que celui que je viens de publier (journal poétique voire documentaire), j'ai moins de besoins techniques ou "de bricolage", car je m'appuie sur un réel, qui est solide, qui existe indépendamment de moi. Ce besoin d'édifier un ensemble de repères (temporalité surtout) est pour l'instant plus important que le besoin de "mesurer le résultat", d'analyser ce qui a été fait.

ces fonctionnalités seraient par exemple celles que j'ai citées le 13 :

– assouplissement et augmentation des possibilités de représentation de l'ordre des chapitres. Par exemple, pour l'instant dans *Saga+* c'est un alignement de petits carrés qui ont tous la même taille. Cela m'aiderait qu'il y ait la possibilité de créer deux lignes, une pour la chronologie des événements,

(avec un rapport de proportions de durée de l'événement et de longueur du "carré" ou rectangle qui le représente) et une pour le récit (ce fameux ordre des chapitres). Avec, comme c'est déjà le cas, le corpus de texte qui se modifie en fonction lorsqu'on change un de ces carrés de place. (le fait que le carré montre la teneur en couleur de tel ou tel personnage ou autre élément me semble très bien)

– possibilité d'écrire sur le carré qui représente un chapitre (synopsis, résumé, notes à propos du chapitre)

– les tags, assignés non automatiquement mais manuellement, qui permettraient de regrouper l'ensemble des passages appartenant à un même "fil", et de les passer en revue lorsqu'on veut relire la trajectoire de ce fil. (mais je me souviens que tu disais que cela posait des problèmes de mémoire?) en tous les cas, je ne crois pas, pour cette fonctionnalité, à un outil qui repérerait automatiquement les passages car parfois le lien est d'une telle subtilité (aucune mention nominale, juste une allusion) que l'outil automatique ne pourrait jamais repérer que le passage est en lien.

Voilà, en gros, je ne fais que répéter mes besoins actuels d'outils qui seraient plutôt là pour m'aider à construire quelque chose d'une certaine ampleur plutôt que de l'évaluer une fois que c'est construit.

Mais je dis peut-être ça car actuellement je commence un gros boulot de roman et que tout est à imaginer. Il faudrait un outil qui permette de se sentir moins perdu, de créer des représentations avant même d'avoir du corpus de texte (des plans).

Mais l'un empêche pas l'autre, peut-être que dans des étapes ultérieures je serai très contente de pouvoir revenir à *Saga+* pour analyser l'effet produit par mes premiers longs rassemblements de chapitres.

**A2** Outre les questions liées à la désignation de personnage par des pronoms ou autre, l'impact de l'unité organisationnelle choisie, ici le chapitre, me semble potentiellement améliorable, par exemple en incluant la possibilité de définition d'unités plus brèves, comme la page (env. 3'000 signes) ou encore le paragraphe. Comme la possibilité d'un travail similaire, mais au

sein d'un chapitre, avec un grain plus fin, plus micro. D'autre part, tout ce qui pourrait encore enrichir l'interface – outil de navigation entre les occurrences dans le texte, outil ou assistant de paramétrage de nouvelles balises pour le référencement d'un personnage, d'un lieu ou d'un pronom (i.e. : je est toujours le narrateur dans ce texte), etc. – peuvent optimiser le fonctionnement fin de *Saga+* dont la prise en main me semble pour le moment remarquablement simple (le risque de contre productivité doit être envisagé dans le cas d'ajouts de nombreuses fonctions et réglages).

**A3** Pouvoir travailler avec des unités plus petites que le chapitre. Comme le fragment.

**A4** Pour ma part, ce serait l'écriture, à savoir la détection automatique d'erreurs d'écriture comme des répétitions ou des maladroites dans la construction de la phrase. Une sorte de Word, mais en efficace.

#### **4) Envisagez-vous d'utiliser *Saga+* dans sa version actuelle pour un projet personnel, ou en recommander l'utilisation à des collègues ?**

**A1** J'en ai déjà parlé à des collègues.

**A2** Absolument, je pense ici à un roman collectif mené à son terme, du point de vue du récit, par une quinzaine d'auteurs mais finalement abandonné pour des raisons qui relèvent de la structure de l'intrigue et d'un nombre de personnage difficilement envisageable seulement par une relecture « humaine », même attentive. Tant et si bien que l'ampleur de la tâche fournir est finalement apparue trop importante pour mobiliser suffisamment les auteurs à un très gros re-travail (le programme peut « aider », faire gagner du temps et proposer des arguments objectifs à la redéfinition de l'économie narrative du roman). *Saga+* pourrait sans doute montrer, grâce à ses représentations visuelles du récit et des relations entre les personnages – un point fort – les endroits du texte qu'il faut retravailler en priorité.

**A3** Je le recommanderais volontiers à des collègues, qu'ils soient écrivains ou chercheurs en littérature.

**A4** (sans avis)

## A.7 Appendix G

### A.7.1 Frankenstein (version courte)

#### Le Monstre

« J'ai beaucoup de peine à me rappeler les premiers moments de mon existence. Les événements de cette période m'apparaissent confus et indistincts. Une multitude de sensations étranges m'agitait. Je voyais, j'entendais, je sentais, je touchais – tout de façon simultanée, mais il me fallut un certain temps avant d'apprendre à faire la distinction entre mes divers sens. Peu à peu, je m'en souviens, une violente lumière m'excita si bien que je fus obligé de fermer les yeux. Surgit alors l'obscurité et j'en fus troublé mais à peine en avais-je eu conscience qu'en ouvrant les yeux je revis la lumière. Je me mis à marcher et je descendais, je crois, lorsque se produisit un grand changement dans mes sensations. Auparavant, des corps sombres et opaques m'entouraient, impossibles de toucher ou de voir. Mais voilà que je découvrais que je pouvais me mouvoir en toute liberté et que j'étais capable de surmonter et de contourner les obstacles. La lumière m'oppressait de plus en plus et la chaleur me gênait, au fur et à mesure que je marchais, à telle enseigne que je recherchai un endroit où il y avait de l'ombre. Ce fut une forêt près d'Ingolstadt. Là, je me reposai en bordure d'un ruisseau, jusqu'à être tourmenté par la faim et par la soif. Cela m'arracha de ma torpeur. Je mangeai des baies que je dénichai sur des arbres ou que je ramassai par terre. J'étanchai ma soif au ruisseau et je m'étendis sur le sol pour trouver le sommeil.

« Il faisait sombre quand je me réveillai. J'avais froid et je me sentis effrayé, comme si, indistinctement, je me rendais compte de ma désolation. Avant de quitter ton appartement, ayant éprouvé une sensation de froid, je m'étais couvert de quelques vêtements mais ce n'était pas assez pour me prémunir contre la rosée nocturne. Je n'étais qu'un être misérable, pauvre et sans secours. Je ne connaissais rien, je ne pouvais rien distinguer. Autour tout me parut hostile. Je m'assis et pleurai.

« Bientôt, une légère lueur jaillit dans le ciel et j'éprouvai une sensation de plaisir. Je me dressai et aperçus une forme rayonnante parmi les arbres. Je la contemplai avec admiration. Elle bougeait lentement mais elle éclairait mon chemin et je repartis à la recherche de baies. Il faisait encore froid, pourtant je découvris sous un arbre un large manteau dont je me couvris avant de me rasseoir par terre. Aucune pensée précise ne m'occupait l'esprit. Tout était confus. Je sentais la lumière, la

faim, le froid, l'obscurité. D'innombrables bruits me tintaient aux oreilles et, de toutes parts, montaient des parfums multiples. La seule chose que je pouvais distinguer était la lune lumineuse et je la fixai avec ravissement. Il y eut plusieurs jours et plusieurs nuits. La durée de la nuit avait fortement diminué, lorsque je commençai à différencier mes diverses sensations. Progressivement, je vis le feuillage où j'allais boire et les arbres sous les feuillages desquels je m'abritais. Je fus émerveillé quand je découvris pour la première fois qu'un son agréable qui m'avait souvent charmé les oreilles provenait de la gorge des petites créatures ailées qui, de temps à autres, interceptaient la lumière à mes yeux. Je commençai aussi à observer de façon beaucoup plus nette les formes qui m'entouraient et à percevoir les limites de la rayonnante voûte de lumière au-dessus de moi. Parfois, j'essayais d'imiter les sons mélodieux des oiseaux mais sans succès. Et parfois aussi j'éprouvais le besoin d'exprimer mes sensations de ma propre manière mais les sons rudes et inarticulés qui sortaient de mes lèvres m'épouvantaient et je retombais dans le silence.

« La lune avait disparu de la nuit puis elle resurgit sous une forme, plus mince, et j'étais toujours dans la forêt. Dans l'intervalle, mes sensations étaient devenues bien distinctes et mon cerveau enregistrait chaque jour des idées nouvelles. Mes yeux commençaient à s'habituer à la lumière et à percevoir les objets dans leur forme la plus exacte. Je discernais l'insecte au milieu de l'herbe et, peu à peu, une herbe d'une autre. Je découvrais que le moineau n'émettait que des sons saccadés, alors que le chant du merle ou de la grive était doux et harmonieux.

« Un jour que j'étais tiraillé par le froid, je dénichai un feu que des vagabonds avaient abandonné et cette découverte de la chaleur fut pour moi un délice. Dans ma joie, je plongeai ma main parmi les braises brûlantes mais je la retirai à la hâte en poussant un cri de douleur. Comme il est curieux, pensais-je, que la même cause produise des effets opposés ! J'examinai les matériaux du feu et vis avec contentement qu'ils étaient composés de bois. Je réunis rapidement quelques branches mais elles étaient trop humides et elles ne s'enflammèrent pas. J'en fus peiné et je m'assis pour contempler l'évolution du feu. Le bois humide que j'avais placé près du foyer sécha et, de lui-même, se mit à brûler. Je réfléchis à ce phénomène puis, après avoir ramassé un tas de

branches, j'en découvris la cause et m'efforçai de réunir une grande quantité de bois afin de les faire sécher et d'avoir une bonne provision. Quand tomba la nuit et que je voulus me reposer, j'eus grandpeur que mon feu n'en vînt à s'éteindre. Je le recouvris soigneusement de bois sec et de feuilles et plaçai au-dessus des branches humides. Puis, après avoir déployé mon manteau, je me couchai sur le sol et m'endormis.

### Victor Frankenstein

Un des phénomènes qui avaient singulièrement retenu mon attention était la structure du corps humain, et même de tout être doué de vie. D'où vient, me demandais-je souvent, le principe de la vie ? Une question hardie qui de tout temps avait constitué un mystère. Pourtant, que de secrets ne dévoilerions-nous pas, si la lâcheté et la négligence ne venaient perturber nos recherches ? Je ruminai ces circonstances et décidai bientôt de m'appliquer plus particulièrement au domaine des sciences naturelles qui se rapporte à la physiologie. Si je n'avais pas été animé d'un enthousiasme extraordinaire, l'étude de cette branche m'aurait paru ennuyeuse et presque intolérable. Pour examiner les causes de la vie, nous devons d'abord connaître celles de la mort. Je me tournai vers l'anatomie mais ce ne fut pas suffisant. Je devais aussi observer la décomposition naturelle et la corruption du corps humain. Dans mon éducation, mon père avait pris toutes ses précautions pour que mon esprit ne soit pas impressionné par des horreurs surnaturelles. Je ne souviens pas d'avoir tremblé pour une superstition ni d'avoir craint l'apparition d'un spectre. Les ténèbres n'avaient pas d'effet sur mon imagination et un cimetière était seulement pour moi le reposoir des corps privés de vie qui, après avoir connu la beauté et la force, deviennent la proie des vers. Et maintenant, j'étais amené à examiner les causes et l'évolution de la corruption, à passer mes jours et mes nuits dans des caveaux et des charniers. Mon attention se concentrait ainsi sur l'objet le plus insupportable à la délicatesse des sentiments humains. Je voyais l'enlaidissement et la dégradation des formes les plus pures, j'assistais à l'action dévastatrice de la mort ronger et, détruire la vie, je découvrais la vermine se nourrir de l'œil et du cerveau. Je fixais, j'observais, j'analysais en détail les causes et les effets, les passages de la vie à la mort et de la mort à la vie. Et puis des ténèbres une soudaine leur jaillit dans, mon cerveau une lueur si brillante, si merveilleuse et pourtant si simple que j'en fus ébloui. Elle m'ouvrait d'immenses perspectives et je fus étonné que parmi tous les hommes de génie qui avaient mené des expériences et entrepris des travaux dans le même sens

je fusse le premier à qui devait être réservé le privilège de découvrir un aussi formidable trésor.

Souvenez-vous en, je ne vous rapporte pas la vision d'un fou. Aussi vrai que le soleil brille au firmament, je vous affirme que c'est la vérité. Quelque miracle s'est produit sans doute et pourtant les étapes de ma découverte ont été distinctes et probantes. Après des jours et des nuits de labeur incroyable et de fatigue, je découvrais la cause de la génération et de la vie. Davantage : je devenais capable d'animer la matière inerte.

L'étonnement dont je fus saisi avec cette découverte fit bientôt place à l'allégresse. Après un travail long et pénible, la réalisation de mes desirs constituait une juste récompense. Et cette découverte était si considérable, si prodigieuse que j'oubliai que je n'y étais arrivé que petit à petit et que je ne considérai que le résultat. Ce qui avait été étudié et désiré par les savants les plus éminents depuis la création du monde était à présent à ma portée.

Mais ce n'était pas comme par magie que tout m'apparaissait : la certitude que j'avais acquise était plutôt de nature à diriger mes efforts vers un but précis car celui-ci n'était pas encore atteint.

Je constate, mon ami, à votre, impatience, à l'étonnement et à l'expectative que manifestent vos yeux, que vous vous attendez à ce que je vous révèle mon secret. Je ne peux pas le faire. Écoutez patiemment la suite de mon histoire et vous allez comprendre pourquoi je reste sur la réserve. Je ne peux pas vous entraîner, imprudent et ardent comme je l'étais moi-même, vers votre destruction et votre ruine. Apprenez, sinon par mes préceptes, du moins par mon exemple, combien il est dangereux d'acquérir le savoir et combien est plus heureux l'homme qui croit que sa ville natale est le centre de l'univers et qui n'aspire pas à dépasser ses limites naturelles.

Lorsque je m'aperçus que je possédais un pouvoir aussi étonnant, j'hésitai longtemps sur la manière dont je l'utiliserais. J'étais donc capable d'animer la matière mais créer un organisme avec l'entrelacement de ses fibres, de ses muscles et de ses veines, voilà qui représentait un travail d'une incroyable difficulté. Et d'abord je ne savais pas si je tenterais de créer un être qui me ressemblerait ou un organisme plus, simple. Mon premier succès avait à ce point exalté mon imagination que je ne doutais pas de ma capacité d'animer un animal aussi complexe et aussi merveilleux que l'homme. Les matériaux dont je disposais ne semblaient guère convenir à une entreprise aussi délicate et

aussi ardue mais cela ne devait pas handicaper mon succès. J'étais préparé à affronter une multitude de revers, mes essais pouvaient sans cesse être infructueux et, en définitive, mon œuvre pouvait se révéler imparfaite.

Toutefois, je n'avais qu'à considérer les progrès qui s'effectuaient tous les jours dans le domaine de la science et de la mécanique pour espérer que mes tentatives actuelles constitueraient les fondements de mon futur succès. Dans l'ampleur et la complexité de mon plan, rien ne prouvait que ce fût impossible. Ce fut dans cet état d'esprit que j'entrepris la création d'un être humain. Les dimensions réduites de certaines parties du corps de l'homme m'empêchèrent d'avancer rapidement dans mon travail. Aussi je décidai, au rebours de ma première intention, de mettre au point une Créature de stature gigantesque : il aurait plus ou moins huit pieds de haut et sa carrure serait en proportion de sa taille. Cette décision prise, je passai plusieurs mois à rechercher et à se préparer mon matériel et je me mis au travail.

### Le Monstre

« Insensiblement, j'en vins à faire une découverte d'une importance plus grande encore. Je m'aperçus que ces gens-là possédaient un moyen de communiquer leur expérience et, leurs sentiments par des sons articulés. Je découvris que les mots dont ils se servaient produisaient tantôt le plaisir ou la peine, tantôt le sourire ou la tristesse dans les gestes ou sur la physionomie de ceux qui les entendaient. C'était là, sans nul doute, une science divine et je désirai ardemment l'acquérir. Mais toutes mes tentatives en ce sens échouèrent. Leur prononciation était rapide et les mots qu'ils employaient ne semblaient pas avoir de rapport immédiat avec les objets visibles, et j'étais incapable de découvrir le moindre indice qui aurait pu me permettre de comprendre leurs références. Cependant, avec une grande application, après être resté dans ma hutte le temps de plusieurs révolutions lunaires, je découvris les noms qu'ils donnaient dans leurs dialogues à la plupart de leurs objets familiers. J'appris et employai les mots "feu", "lait", "pain" et "bois". J'appris aussi les noms des fermiers eux-mêmes. Le jeune homme et sa compagne en avaient chacun plusieurs, mais le vieillard un seulement qui était "père". La fille était appelée "sœur" ou "Agatha", et le jeune homme "Félix", "frère" ou "fils". Je ne pourrais pas décrire ma joie quand je compris quelles idées étaient appropriées à chacun de ses sons et quand je fus à même de les prononcer moi aussi. Je distinguai d'autres mots encore mais sans pouvoir les comprendre ni les appliquer, tels que "bon", "très cher",

"malheureux".

« Ainsi se passa l'hiver. Les manières affables et la sympathie des fermiers me les rendirent très chers. Quand ils étaient malheureux, je me sentais déprimé. Quand ils se réjouissaient, je partageais leur allégresse. En dehors d'eux, je voyais peu de gens et jamais personne d'autre n'entrait dans la ferme. Mais les autres avaient des allures frustes et grossières et, par comparaison, ma sympathie pour mes amis ne faisait qu'augmenter. Le vieillard, je pouvais le constater, cherchait souvent à encourager ses enfants – ainsi qu'il les appelait quelquefois – et à dissiper leur mélancolie. Il parlait alors avec un accent de gaieté, avec une expression de bonté qui me procurait du plaisir, Agatha l'écoutait avec respect, les yeux parfois remplis de larmes qu'elle essayait de faire disparaître sans qu'il s'en aperçût. Mais je remarquai que son visage et sa voix étaient généralement beaucoup plus radieux, après qu'elle avait écouté les exhortations de son père. Ce n'était pas pareil avec Félix. Il était toujours le plus triste du groupe et, malgré mon manque d'expérience, il me donnait l'impression d'avoir davantage souffert que ses compagnons. Pourtant, s'il avait une physionomie plus affligée, sa voix était caressante, plus douce que celle de sa sœur, surtout quand il s'adressait au vieillard.

« Je pourrais mentionner d'innombrables exemples qui illustreraient clairement les bonnes dispositions de ces aimables fermiers. Au milieu de la pauvreté et de la gêne, Félix offrait spontanément à sa sœur la première petite fleur blanche qui avait percé sous le tapis de neige. Très tôt le matin, avant qu'elle ne fût levée, il balayait la neige qui obstruait le chemin de l'étable, tirait de l'eau du puits et ramenait chez lui une provision de bois, qu'une main inconnue, à son grand étonnement, continuait de lui fournir. Pendant la journée, il travaillait, je crois, dans une ferme du voisinage car il partait souvent tôt le matin et ne rentrait que le soir, sans rapporter du bois. À d'autres moments, il travaillait au jardin mais, comme il y avait peu de besogne en cette saison froide, il faisait la lecture au vieillard et à Agatha.

« Ces lectures, au début, m'avaient extrêmement intrigué. Mais, peu à peu, je me rendis compte que les sons qu'il émettait lorsqu'il parlait étaient les mêmes que ceux qu'il émettait lorsqu'il lisait. Je supposai donc qu'il trouvait sur le papier des signes qui lui permettaient de parler et qu'il comprenait et je voulus moi aussi les connaître. Mais était-ce possible puisque je ne pouvais pas saisir les sons correspondant à ces signes ? Néanmoins, je fis de notables progrès en ce domaine mais ils n'étaient pas suffisants pour me permettre de

suivre une conversation quelconque, quelle que fût l'application avec laquelle je m'attelais à cette tâche. J'avais une grande envie de révéler ma présence aux fermiers mais je m'apercevais bien que je ne devais rien tenter avant d'avoir réussi à maîtriser leur langage – et peut-être, en étant capable de parler, pouvais-je aussi faire oublier la difformité de ma figure, car sur ce point-là aussi j'avais appris à mesurer les différences existant entre nous.

« J'avais admiré la perfection des corps des fermiers leur grâce, leur beauté, la délicatesse de leur allure. Comme j'étais terrifié lorsque je voyais mon reflet dans l'eau ! La première fois, je m'étais jeté en arrière, ne pouvant pas croire que c'était moi que le miroir réfléchissait. Mais lorsque je fus pleinement convaincu que j'étais un authentique monstre, je ressentis une profonde, une humiliante amertume. Hélas ! Je ne connaissais pas tout à fait encore les conséquences fatales de ma misérable difformité !

### Victor Frankenstein

Ce fut par une sinistre nuit de novembre que je parvins à mettre un terme à mes travaux. Avec une anxiété qui me rapprochait de l'agonie, je rassemblai autour de moi les instruments qui devaient donner la vie et introduire une étincelle d'existence dans cette matière inerte qui gisait à mes pieds. Il était une heure du matin et la pluie frappait lugubrement contre les vitres. Ma bougie allait s'éteindre lorsque tout à coup, au milieu de cette lumière vacillante, je vis s'ouvrir l'œil jaune stupide de la Créature. Elle se mit à respirer et des mouvements convulsifs lui agitèrent les membres.

Comment pourrais-je décrire mon émoi devant un tel prodige ? Comment pourrais-je dépeindre cet être horrible dont la création m'avait coûté tant de peines et tant de soins ? Ses membres étaient proportionnés et les traits que je lui avais choisis avaient quelque beauté. Quelque beauté ! Grand Dieu ! Sa peau jaunâtre, tendue à l'extrême, dissimulait à peine ses muscles et ses artères. Sa longue chevelure était d'un noir brillant et ses dents d'une blancheur de nacre. Mais ces avantages ne formaient qu'un contraste plus monstrueux avec ses yeux stupides dont la couleur semblait presque la même que celle, blême, des orbites. Il avait la peau ridée et les lèvres noires et minces.

Les avatars multiples de l'existence ne sont pas aussi variables que les sentiments humains. J'avais, pendant deux ans, travaillé sans répit pour donner la vie à un corps inanimé. Et, pour cela,

j'avais négligé mon repos et ma santé. Ce but, j'avais cherché à l'atteindre avec une ardeur immodérée – mais maintenant que j'y étais parvenu, la beauté de mon rêve s'évanouissait et j'avais le cœur rempli d'épouvante et de dégoût. Incapable de supporter la vue de l'être que j'avais créé, je sortis de mon laboratoire et longtemps je tournai en rond dans ma chambre à coucher, sans trouver le sommeil. Enfin la fatigue l'emporta et je me jetai tout habillé sur mon lit pour chercher, quelque temps, l'oubli de ma situation. En vain. Je dormis sans doute mais ce fut pour être assailli par les rêves les plus terribles. Je crus voir Élisabeth, débordante de santé, se promener dans les rues d'Ingolstadt. Charmé et surpris, je l'enlaçai mais, alors que je posais mes lèvres sur les siennes, elle devint livide comme la mort. Ses traits se décomposèrent et j'eus l'impression que je tenais entre mes bras le cadavre de ma mère. Un linceul l'enveloppait et, à travers les plis, je vis grouiller les vers de la tombe. Je me réveillai avec horreur.

Une sueur glacée me couvrait le front, mes dents claquaient, j'étais saisi de convulsions. Puis, la lumière jaunâtre de la lune se glissa à travers les croisées de la fenêtre et j'aperçus le malheureux – le misérable Monstre que j'avais créé. Il soulevait le rideau de mon lit et ses yeux, si je puis les appeler ainsi, étaient fixés sur moi. Ses mâchoires s'ouvrirent et il fit entendre des sons inarticulés, tout en grimaçant. Peut-être parlait-il mais je ne l'entendis pas. Une de ses mains était tendue, comme pour me retenir. Je pris la fuite et me précipitai vers les escaliers. Je cherchai refuge dans la cour de la maison où je passai le reste de la nuit, marchant fébrilement de long en large, aux aguets, attentif au moindre bruit, à croire qu'il annonçait chaque fois l'approche du démon à qui j'avais si piteusement donné la vie.

Oh ! Quel mortel pourrait supporter l'horreur d'une telle situation ! Une momie à qui l'on rendrait l'âme ne pourrait pas être aussi hideuse que ce misérable. Je l'avais observé avant qu'il ne fût achevé : il était laid à ce moment-là, mais quand ses muscles et ses articulations furent à même de se mouvoir, il devint si repoussant que Dante lui-même n'aurait pas pu l'imaginer.

Je passai une nuit épouvantable. Quelquefois, mon pouls battait si vite et si fort que je sentais la palpitation de chacune de mes artères. Il m'arrivait aussi de chanceler, tant ma fatigue était grande, tant ma faiblesse était profonde. Et mêlée à cette horreur, l'amertume née de mon dépit me tirait. Les rêves dont je m'étais nourri et qui avaient soutenu mon exaltation devenaient à présent un enfer. Le changement avait été si



brutal, la désillusion si complète !

Le jour, enfin, commença à paraître – un jour sombre et pluvieux. Mes yeux découvrirent le clocher blanc de l'église d'Ingolstadt et l'horloge qui marquait six heures.

Le portier ouvrait les portes de la cour qui, cette nuit, avait été mon asile. Je sortis, allai précipitamment par les rues, un peu comme si je voulais fuir le misérable, craignant de le rencontrer à chaque carrefour. Je n'osais pas retourner dans mon appartement, je me sentais le besoin irrésistible de marcher, bien que trempé par la pluie qui tombait à verse d'un ciel sombre et bas.

J'errai longtemps de la sorte, cherchant par la fatigue physique de me soulager du poids qui m'accablait l'esprit.

Je parcourus les rues sans savoir où j'étais ni ce que je faisais. Mon cœur battait au rythme de la peur et j'allais en titubant, sans un seul regard en arrière.

Tel celui qui, sur la route solitaire,

Chemine dans la peur et l'épouvante,

Et qui, après s'être retourné, va de l'avant

Sans ne plus regarder derrière lui ;

Parce qu'il sait qu'un affreux démon

Marche, menaçant, dans son dos.

### Le Monstre

« Maudit, maudit créateur ! Pourquoi est-ce que je vis ? Pourquoi, à cet instant, n'ai-je pas éteint l'étincelle de vie que tu as si étourdiment allumée en moi ? Je ne sais pas. Le désespoir ne s'était pas encore emparé de mon être ; je n'étais animé que par la rage et que par la vengeance. C'était avec délectation que j'aurais détruit le chalet et ses occupants, que je me serais réjoui de leurs cris d'épouvante et de leur malheur.

« Quand la nuit tomba, je quittai ma cabane et allai me promener dans le bois. À présent, je n'éprouvais plus la crainte d'être découvert. Je libérai mon angoisse en poussant des hurlements effroyables. Ainsi qu'une bête sauvage qui vient briser ses chaînes, je détruisais les objets qui se dressaient devant moi, fonçant parmi les, taillis à la vitesse d'un cerf. Oh ! Quelle affreuse nuit j'ai passée ! Les froides étoiles se moquaient de moi, les arbres dépouillés étendaient leurs branches

au-dessus de ma tête, de loin en loin la douce voix d'un oiseau venait déchirer l'universel silence. Tout, sauf moi, se reposait ou s'amusait. Et moi, démon parmi les démons, je portais l'enfer en mon sein. Ne trouvant personne avec qui sympathiser, je voulais arracher les arbres, semer autour de moi la ruine et la destruction avant de m'asseoir pour admirer mon œuvre.

« Mais c'était là, un paroxysme insupportable. Ces excès physiques m'avaient fatigué et je m'étendis sur l'herbe humide, frappé d'impuissance et de désespoir. Parmi les myriades d'hommes existait-il un seul qui pourrait avoir pitié de moi ou qui pourrait me secourir ? Devais-je éprouver de la bonté envers mes ennemis ? Non ! À partir de ce moment-là, je déclarai la guerre au genre humain et, par-dessus tout, à celui qui m'avait façonné et qui avait provoqué chez-moi cette détresse intolérable. » Le soleil se leva. J'entendis des voix d'homme et me rendis compte qu'il n'était pas possible de regagner mon abri pendant la journée. Je me cachai dans d'épais taillis, déterminé à passer les heures suivantes à réfléchir sur ma situation.

« À ce moment, un léger sommeil dissipait déjà ma rancœur quand je fus réveillé par l'arrivée d'un beau garçon qui, plein d'agilité, venait en courant vers l'abri que je m'étais choisi. Et soudain, en le voyant, j'eus l'idée qu'une petite Créature ne pouvait pas avoir, elle, de préjugés et qu'elle n'avait assez vécu pour connaître l'épouvante et la laideur. Aussi, si je parvenais à m'emparer de lui, si je réussissais à en faire un ami et un compagnon, je ne serais plus seul dans ce monde peuplé d'hommes' » Obéissant à mon impulsion, je saisis le garçon au passage et l'attirai vers moi. Dès que ma physionomie lui fut révélée, il plaça ses mains devant les yeux et poussa un cri formidable. Je lui tirai énergiquement les mains du visage et lui dis :

« – Pourquoi fais-tu cela, mon enfant ? Je n'ai pas l'intention de te nuire. Écoute-moi.

« Il se débattit violemment.

« – Lâchez-moi, hurla-t-il. Monstre ! Abominable Créature ! Vous voulez me manger et me mettre en pièces. Vous êtes un ogre. Laissez-moi partir ou je le dirai à mon papa.

« – Tu ne reverras plus jamais ton père, mon garçon. Tu dois venir avec moi !

« – Hideux Monstre ! Laissez-moi partir. Mon papa est un syndic' C'est M. Frankenstein.. Il vous punira. Vous n'oserez pas me garder !

« – Frankenstein ! Tu es donc de la famille de mon

ennemi, de celui envers lequel je nourris une éternelle vengeance. Tu seras ma première victime !

« L'enfant se débattait toujours et m'accablait d'injures qui me déchiraient le cœur. Je le pris à la gorge pour le faire taire mais, en un rien de temps, il tomba mort à mes pieds.

« Je contemplai ma victime et mon cœur se gonfla d'exultation et d'un triomphe infernal. En battant des mains, je m'écriai :

« – Moi aussi, je peux créer la désolation. Mon ennemi n'est pas invulnérable. Cette mort le remplira de désespoir et mille autres misères le tourmenteront et l'annihileront !

« Comme j'avais les yeux sur l'enfant, je vis quelque chose briller sur son cou. Je m'en emparai. C'était le portrait d'une très belle femme. En dépit de ma hargne, il me séduisit et me fascina. Pour un court moment, je fus sous le charme de ses yeux sombres frangés de longs cils et de ses lèvres exquises. Mais très vite ma rage reprit le dessus. Je me rappelai que j'étais à jamais privé des joies qu'une créature aussi belle aurait pu m'octroyer et je me dis que si celle dont je contemplais le visage me voyait elle n'aurait plus cet aspect délicieux mais une expression de dégoût et d'horreur.

« Peux-tu t'étonner que de telles pensées aient at-tisé ma fureur ? Je me demande pourquoi sur le moment même, au lieu de donner libre cours à mes sentiments de douleur par des exclamations, je ne me suis pas précipité parmi les hommes en cherchant, au risque de perdre la vie, de les tuer.

« Mais ces pensées m'avaient épuisé et je quit-tai l'endroit où j'avais commis le meurtre afin de dénicher un abri plus sûr. J'entrai dans une grange qui m'avait paru vide. Sur la paille, une femme y dormait. Elle était jeune, pas aussi belle que celle qui figurait sur le portrait, avenante pourtant, pleine de charme et de santé. Je me dis qu'une telle créature était de celles dont les radieux sourires ne me seraient jamais destinés. Je me penchai sur elle et lui murmurai :

« – Réveille-toi, ma douce, ton amant est à tes côtés – il est prêt à te donner sa vie pour un seul de tes regards affectueux. Réveille-toi, mon amour !

« La femme qui dormait remua et un frisson de terreur me parcourut. Et si elle se réveillait ef-fectivement, si elle me voyait, si elle me maud-issait, si elle dénonçait mon meurtre ? Elle le ferait sans nul doute dès qu'elle ouvrirait les yeux et m'apercevrait. Cette idée attisa ma folie, ranima

ma hargne. Non, ce ne serait pas moi qui souf-frirais, mais elle ! Le crime que j'avais commis parce jamais je n'aurais pu obtenir tout ce qu'elle aurait dû me donner, ce serait le sien. Elle en était la cause, c'est elle qu'on punirait. Grâce aux leçons de Félix sur les lois sanguinaires des hommes, j'avais appris à présent comment faire le mal. Je me penchai de nouveau et glissai soigneusement le portrait dans un des plis de sa robe. Elle bougea encore et je pris la fuite.

« Durant quelques jours, je hantai l'endroit où s'étaient produits ces événements, tantôt dans l'espoir de te voir, tantôt de quitter à jamais le monde et ses misères. Finalement, j'allai errer dans les montagnes et j'en ai exploré tous les recoins, animé par une passion brûlante que toi seul tu peux satisfaire. Nous ne nous séparerons pas avant que tu n'aies accédé à ma demande. Je suis seul et misérable. L'homme ne veut pas de moi. Seule une femme, aussi laide et aussi horrible que moi, souffrirait ma compagnie. Elle devrait être de la même engeance et avoir tous mes défauts. Cet être-là, c'est à toi de le créer ! »

### Victor Frankenstein

La Créature se tut et me regarda fixement, dans l'attente d'une réponse. Mais j'étais décontenancé, perplexe, incapable d'ordonner suffisamment mes idées pour comprendre toute l'étendue de cette proposition. Il reprit la parole :

- Tu dois créer une femme avec laquelle je peux vivre et partager toutes les affections qui sont nécessaires à mon existence. Toi seul, tu le peux. Je l'exige et c'est un droit que tu ne peux pas me refuser.

La dernière partie de son récit avait réveillée en moi la colère qui s'était estompée, alors qu'il me racontait sa vie paisible au chalet. Mais, avec ce qu'il avait dit, il ne m'était plus possible de contenir ma rage.

- Je refuse, lui répondis-je. Et aucune torture ne réussirait à m'arracher mon accord. Tu peux faire de moi l'homme le plus misérable mais tu ne pourras jamais m'abaisser à ce point ! Créer une autre Créature pareille à toi pour que vous jetiez ensemble la désolation sur le monde ? Va-t'en ! Je t'ai répondu. Tu peux me torturer mais je n'accepterai jamais !

- Tu te trompes, reprit le Monstre. Au lieu de te menacer, je suis disposé à discuter avec toi. Si je suis mauvais, c'est parce que je suis malheureux. Ne suis-je pas banni et repoussé par tout le genre humain ? Toi, mon créateur, tu

veux m'anéantir et triompher. Réfléchis donc et demande- toi pourquoi je devrais avoir de la pitié envers ceux qui n'en manifestent pas à mon égard? Tu n'appellerais pas cela un meurtre si tu pouvais me précipiter dans une de ces crevasses et détruire mon corps, ton oeuvre, de tes propres mains? Dois-je avoir du respect pour l'homme qui me méprise? Qu'il ait de l'affection pour moi et, au lieu de lui faire du mal, je le servirai s'il l'accepte avec des larmes de gratitude. Mais ce n'est pas possible: les sentiments humains forment une barrière infranchissable pour notre union. Jamais pourtant je ne me soumettrai à un esclavage aussi abject. Je veux venger les injustices que j'ai subies. Si je ne peux pas inspirer l'amour, je répandrai la peur, et principalement sur toi, mon plus grand ennemi, parce que tu m'as créé et que je nourris envers toi une haine inextinguible. Je serai l'instrument de ta destruction jusqu'à te retourner le coeur et te faire maudire le jour où tu es né!

Tout en parlant, il était tirillé par une rage féroce. Ses traits étaient parcourus de contorsions tellement épouvantables qu'aucun regard humain n'aurait pu les supporter. Puis, il se calma et poursuivit:

- J'avais l'intention d'être raisonnable. Mon emportement m'est nuisible car tu dois te dire que c'est toi qui es la cause de mes excès. Si quelqu'un m'accordait des sentiments de bienveillance, je les lui rendrais au moins au centuple. Pour plaire à une seule créature, je ferais la paix avec l'humanité tout entière. Mais je ne veux pas non plus me laisser aller à des rêves de bonheur qui ne peuvent pas s'accomplir. Ce que je te demande est raisonnable et commode - une créature du sexe opposé aussi affreuse que moi. C'est là une maigre consolation mais c'est aussi tout ce que je peux recevoir et je m'en contenterai. Il est vrai que nous serons des monstres à l'écart du monde mais, pour cette même raison, nous serons davantage attachés l'un à l'autre. Nos vies ne seront pas heureuses mais elles seront sans tache et je serai libéré de la détresse que j'éprouve. Oh, mon créateur, rends-moi heureux! Fais en sorte que je te sois reconnaissant! Laisse-moi me rendre compte que je suis à même de susciter la sympathie de quelqu'un. Ne rejette pas ma requête!

J'étais secoué. Je tremblais en pensant aux conséquences possibles d'une telle solution mais je sentais aussi qu'il y avait du vrai dans ses arguments. Son récit et les sentiments qu'il exprimait prouvaient qu'il était une créature qui avait du bon sens. Est-ce que moi qui l'avais fait je ne devais pas lui offrir des bribes de bonheur, pour autant que cela était dans mon pouvoir? Il remarqua que

mes sentiments s'étaient modifiés et dit:

- Si tu consens, plus jamais aucune créature humaine ne me reverra. Je partirai pour les vastes contrées sauvages de l'Amérique du Sud. Ma nourriture n'est pas celle des hommes, je ne tue ni l'agneau ni le chevreuil pour apaiser ma faim. Les racines et les baies me suffisent largement. Ma compagne aura la même complexion que la mienne et se contentera de la même chose. Nous ferons notre couche parmi les feuilles. Le soleil brillera pour nous comme pour les hommes et fera naître notre nourriture. Cette description que je te donne est paisible et humaine et tu dois sentir que ce serait faire preuve de méchanceté et de cruauté que de me refuser cela. Tu as été impitoyable envers moi mais maintenant je lis la compassion dans ton regard. Permets-moi de profiter de cet instant favorable et laisse-moi te persuader d'obtenir ce que je désire avec tant d'ardeur.

- Tu me proposes, dis-je, de fuir la proximité des hommes et de gagner des contrées sauvages où les animaux seront tes seuls compagnons. Toi qui cherches l'amour et la sympathie des êtres humains, comment pourrais-tu persévérer dans cet exil? Tu reviendras, tu redemanderas leur affection et tu rencontreras de nouveau leur haine. Tes passions diaboliques renaîtront et tu auras alors une compagne pour t'aider dans ton oeuvre de destruction. Ce n'est pas possible. Cesse de discuter là-dessus car je ne suis pas d'accord.

- Comme tes sentiments sont inconstants! Il y a quelques instants encore, tu étais touché par mes paroles. Pourquoi mes doléances te rendent-elles de nouveau hostile? Je le jure sur cette terre où je suis, sur toi qui m'as fabriqué, si tu me donnes une compagne, je quitterai le voisinage des hommes et j'irai me réfugier, s'il le faut, dans les lieux les plus sauvages! Mes passions diaboliques n'existeront plus puisque je connaîtrai l'affection. Ma vie se passera paisiblement et, à l'heure de ma mort, je ne maudirai pas mon créateur.

Ces mots eurent un étrange effet sur moi. J'avais pitié de lui et, en même temps, je voulais le consoler. Mais, lorsque je le regardais, quand je voyais sa masse difforme balloter au moment où il prenait la parole, mon coeur se soulevait et je me sentais horrifié et dégoûté. J'essayai de chasser ces sensations. Je pensais que si je pouvais éprouver de la sympathie pour lui, je n'avais pas le droit de lui refuser non plus ce maigre bonheur qu'il était en mon pouvoir de lui accorder.

- Tu jures, dis-je, que tu seras bon. Tu t'es déjà montré si malicieux que j'ai naturellement toutes

## Appendix A. Appendices

---

les raisons de me méfier de toi! Et si tout cela était une feinte destinée à accroître ton triomphe et à précipiter ta soif de vengeance?

- Comment cela ? Je ne veux pas qu'on se moque de moi et j'exige une réponse. Si je n'ai ni attache, ni affection, la haine et le vice seront mon lot. L'amour annihilerait la cause de mes crimes et je deviendrais une créature dont l'existence serait ignorée de tous. Mes vices sont les fruits de cette solitude forcée que j'abhorre. Les vertus grandiront nécessairement en moi lorsque je vivrai en communion avec une de mes semblables. J'éprouverai les sentiments d'un être sensible et je ferai alors partie, au lieu d'en être exclu, du processus ordinaire de l'existence. Je pris le temps de réfléchir à tout ce qu'il venait de développer et aux divers arguments auxquels il avait recouru. Je songeai qu'il avait eu, au début de son existence, quelques qualités et que par la suite celles-ci avaient subi un choc, à cause du mépris que lui avaient manifesté ses protecteurs. Dans mes calculs, je ne pouvais pas ne pas tenir compte de sa force et de ses menaces. Une créature qui était capable de vivre parmi les glaciers et de fuir le long des précipices inaccessibles possédait un pouvoir contre lequel il était vain de lutter. Après avoir longuement médité, je conclus qu'en toute justice je devais, aussi bien pour lui que pour tous mes semblables, répondre favorablement à sa requête. Aussi je me tournai vers lui pour lui dire:

- J'accepte ce que tu me demandes, à condition que tu me jures formellement de quitter l'Europe pour toujours ainsi que tout lieu où il y aurait des hommes, une fois que je t'aurai donné cette femme qui t'accompagnera dans ton exil.

- Je le jure, cria-t-il, par le soleil, par le ciel, par le feu de l'amour qui me consume le coeur, que, si tu exauces ma prière, jamais plus tu ne me rever-

ras. Rentre donc dans ta maison commence ton travail. J'en attendrai le résultat avec une angoisse immense. Mais n'aie pas peur, quand tout sera prêt, je ferai mon apparition!

Sur ces mots, il me quitta précipitamment, craignant sans doute que je ne change d'avis. Je le vis descendre la montagne à toute vitesse, tel un aigle qui volait, et disparaître rapidement parmi les ondulations de la mer de glace.

Son récit avait occupé une journée entière et le soleil touchait déjà l'horizon quand il partit. Je savais que je devais me dépêcher de rejoindre la vallée si je ne voulais pas être surpris par les ténèbres. Mais j'avais le coeur lourd et ma démarche était lente. Je peinaï sur les petits sentiers montagneux, mes pas manquaient de fermeté, tant j'étais indécis, remué par tous les événements de la journée. La nuit était déjà fort avancée lorsque je parvins au refuge situé à mi-route et je m'assis auprès d'une fontaine. De loin en loin, au milieu des nuages qui passaient, brillaient les étoiles. Des sapins sombres se dressaient devant moi et, par places, des arbres déracinés jonchaient le sol. C'était un spectacle d'une solennité extraordinaire qui m'arracha des pensées étranges. Je pleurai amèrement. Angoissé, je joignis les mains et m'écriai:

- Oh! Étoiles, nuages, vents! Vous vous moquez tous de moi! Si vous me prenez en pitié, débarrassez-moi de toute sensation, de toute mémoire! Réduisez-moi à néant. Sinon, partez, partez et laissez-moi parmi les ténèbres! C'étaient des pensées misérables et ridicules mais j'ai du mal à vous dire combien j'étais accablé par la vue de ces étoiles qui scintillaient sans relâche, alors que soufflaient les rafales de vent, comme si c'était un violent sirocco qui allait me consumer.

## Bibliography

- [mcc(2011)] Linguistic Inquiry and Word Count (LIWC): Pronounced “Luke,” . . . and Other Useful Facts. In Philip M. McCarthy and Chutima Boonthum-Denecke, editors, *Applied Natural Language Processing: Identification, Investigation and Resolution*, pages 206–229. IGI Global, October 2011. ISBN 978-1-60960-741-8 978-1-60960-742-5. URL <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60960-741-8>.
- [Abrams et al.(2004)] Chadia Abrams, Diane Maloney-Krichmar, and Jenny Preece. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.
- [Agarwal et al.(2012)] Apoorv Agarwal, Augusto Corvalan, Jacob Jensen, and Owen Rambow. Social network analysis of alice in wonderland. In *Workshop on Computational Linguistics for Literature*, pages 88–96, 2012.
- [Agarwal et al.(2014)] Apoorv Agarwal, Sarthak Dash, Sriramkumar Balasubramanian, and Jiehan Zheng. Using Determinantal Point Processes for Clustering with Application to Automatically Generating and Drawing xkcd Movie Narrative Charts. May 2014. ISBN 978-1-62561-000-3. URL <http://www.ase360.org/handle/123456789/47>.
- [Aggarwal(2012)] Charu C. Aggarwal, editor. *Mining text data*. Springer, New York, NY, 2012. OCLC: 775736101.
- [Akyigit et al.(2014)] Erk Ediz Akyigit, Tugkan Cengiz, Onur Burak Yildirim, and Selim Balcisoy. Visual exploratory tool for storyline generation. pages 215–216. IEEE, October 2014. ISBN 978-1-4799-6227-3. doi: 10.1109/VAST.2014.7042497. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7042497>.
- [Albanese(2011)] Andrew Albanese. Small Demons Makes Big Splash at Frankfurt Book Fair. *Publishers Weekly*, (258.43):6–8, October 2011.
- [Albuquerque et al.(2011)] Alexandre C. Albuquerque, Cesar Tadeu Pozzer, and Angelo E. M. Ciarlini. The Usage of the Structural-Affect Theory of Stories for Narrative Generation. pages 250–259. IEEE, November 2011. ISBN 978-0-7695-4648-3. doi: 10.1109/SBGAMES.2011.38. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6363240>.
- [Alter(2012)] Alexandra Alter. Your E-Book Is Reading You. *The Wall Street Journal*, July 2012.
- [Arac(2002)] Jonathan Arac. Anglo-globalism? *New Left Review*, 16:35, 2002.
- [Arakji and Lang(2007)] Reina Arakji and Karl Lang. Digital Consumer Networks and Producer-Consumer Collaboration: Innovation and Product Development in the Video

## Bibliography

---

- Game Industry. *Journal of Management Information Systems*, 24(2):195–219, October 2007.
- [Ayckbourn(1985)] Alan Ayckbourn. *Intimate exchanges: a play*. S. French, London ; New York, 1985.
- [Azpeitia et al.(2014)] Andoni Azpeitia, Montse Cuadros, Seán Gaines, and German Rigau. NERC-fr: Supervised Named Entity Recognition for French. In Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, editors, *Text, Speech and Dialogue*, volume 8655, pages 158–165. Springer International Publishing, Cham, 2014. ISBN 978-3-319-10815-5. URL [http://link.springer.com/10.1007/978-3-319-10816-2\\_20](http://link.springer.com/10.1007/978-3-319-10816-2_20).
- [Babauta(2015)] Leo Babauta. 10 Free Minimalist Word Processors for Greater Productivity, June 2015. URL <https://gigaom.com/2007/06/22/10-free-minimalist-word-processors/>.
- [Barraud(2014)] Emmanuel Barraud. Premiers pas dans la «littérature potentielle 2.0», April 2014. URL <http://actu.epfl.ch/news/premiers-pas-dans-la-litterature-potentielle-20/>.
- [Baum(2006)] F. Baum. Participatory action research. *Journal of Epidemiology & Community Health*, 60(10):854–857, October 2006.
- [Bernstein et al.(2015)] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. *Communications of the ACM*, 58(8):85–94, July 2015.
- [Bienvault(2014)] Hervé Bienvault. Daniel de Roulet : La Simulation humaine, oeuvre numérique, 2014. URL [http://aldus2006.typepad.fr/mon\\_weblog/2014/05/daniel-de-roulet-la-simulation-humaine-oeuvre-num%C3%A9rique.html](http://aldus2006.typepad.fr/mon_weblog/2014/05/daniel-de-roulet-la-simulation-humaine-oeuvre-num%C3%A9rique.html).
- [Black(2005)] Rebecca W Black. Online fanfiction: What technology and popular culture can teach us about writing and literacy instruction. *New Horizons for Learning Online Journal*, 11(2), 2005.
- [Boinon et al.(2011)] D Boinon, C Charles, S Dauchy, and S Sultan. Les effets de l’expression émotionnelle sur l’ajustement au cancer: une revue systématique de la littérature. *Psycho-oncologie*, 5(3):173–190, 2011.
- [Booth(1967)] Taylor L Booth. *Sequential machines and automata theory*. Wiley, New York, 1967. OCLC: 523861.
- [Bornet and Kaplan(2015)] Cyril Bornet and Frédéric Kaplan. Anatomy of a Drop-Off Reading Curve. In *DH2015*, Sydney, Australia, 2015.
- [Bornet and Kaplan(2017)] Cyril Bornet and Frédéric Kaplan. A simple set of rules for characters and place detection in French novels. *Frontiers in Digital Humanities*, 5, March 2017.
- [Bornet et al.(2014)] Cyril Bornet, Daniel de Roulet, and Frédéric Kaplan. La simulation humaine : le roman-fleuve comme terrain d’expérimentation narrative. *Cahiers de Narratologie*, (27), December 2014.
- [Bransford(2013)] Nathan Bransford. *How to write a novel: 47 rules for writing a stupendously awesome novel that you will love forever*. Nathan Bransford, United States, 2013. OCLC: 889902581.

- [Brewer(1984)] William F Brewer. The story schema : universal and culture-specific properties. *Center for the Study of Reading Technical Report*, (322), August 1984.
- [Brewer and Lichtenstein(1982)] William F. Brewer and Edward H. Lichtenstein. Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, 6(5-6):473–486, December 1982.
- [Brewer and Ohtsuka(1988)] William F. Brewer and Keisuke Ohtsuka. Story structure, characterization, just world organization, and reader affect in American and Hungarian short stories. *Poetics*, 17(4-5):395–415, October 1988.
- [Bridgwood and Radford(1986)] Christine Bridgwood and Jean Radford. Family Romances: The Contemporary Popular Family Saga. In *The Progress of Romance. The Politics of Popular Fiction*, pages 167–193. Routledge & Kegan Paul, London ; New York, 1986. ISBN 0-7102-0963-0.
- [Buckingham and Willett(2013)] David Buckingham and Rebekah Willett. *Digital Generations Children, Young People, and the New Media*. Taylor and Francis, Hoboken, 2013.
- [Buntinx et al.(2017)] Vincent Buntinx, Cyril Bornet, and Frédéric Kaplan. Studying Linguistic Changes over 200 Years of Newspapers through Resilient Words Analysis. *Frontiers in Digital Humanities*, 4, February 2017.
- [Büchner and Mulvenna(1998)] Alex G. Büchner and Maurice D. Mulvenna. Discovering Internet marketing intelligence through online analytical web usage mining. *ACM SIGMOD Record*, 27(4):54–61, December 1998.
- [Calle-Gruber and Simon(2011)] Mireille Calle-Gruber and Claude Simon. *Claude Simon: l'inlassable réancrage du vécu*. Number 75 in Les essais. Différence, Paris, 2011.
- [Cardie and Wilkerson(2008)] Claire Cardie and John Wilkerson. Text annotation for political science research. 2008.
- [Carter and Carter(1982)] Robert C. Carter and Ellen C. Carter. High-contrast sets of colors. *Applied Optics*, 21(16):2936, August 1982.
- [Catinchi(2014)] Philippe-Jean Catinchi. Désintégrations. *Le Monde*, July 2014.
- [Catmull and Rom(1974)] Edwin Catmull and Raphael Rom. A class of local interpolating splines. *Computer aided geometric design*, 74:317–326, 1974.
- [Cesário et al.(2016)] Vanessa Cesário, Paulo Freitas, Diana Pimentel, and Valentina Nisi. Children's Books: Paper VS Digital, What Do They Prefer? pages 625–630. ACM Press, 2016. ISBN 978-1-4503-4313-8. doi: 10.1145/2930674.2936004. URL <http://dl.acm.org/citation.cfm?doid=2930674.2936004>.
- [Chaplin(2010)] Heather Chaplin. The Future of Reading and Writing is Collaborative, November 2010. URL [http://spotlight.macfound.org/featured-stories/entry/the-future-of-reading-and-writing-is-collaborative/?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed:+macfound/iQaL+Spotlight+on+Digital+Media+and+Learning](http://spotlight.macfound.org/featured-stories/entry/the-future-of-reading-and-writing-is-collaborative/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+macfound/iQaL+Spotlight+on+Digital+Media+and+Learning).
- [Checkland and Holwell(1998)] Peter Checkland and Sue Holwell. Action Research: Its Nature and Validity. *Systemic Practice and Action Research*, 11(1):9–21, 1998.

## Bibliography

---

- [Chung and Pennebaker(2012)] Cindy K Chung and James W Pennebaker. Linguistic Inquiry and Word Count (LIWC). 2012.
- [Clement et al.(2009)] Tanya Clement, Catherine Plaisant, and Romain Vuillemot. The Story of One: Humanity scholarship with visualization and text analysis. *Relation*, 10(1.43): 8485, 2009.
- [Cortázar(2006)] Julio Cortázar. *Rayuela*. Punto de Lectura, Madrid, 2006. OCLC: 76966303.
- [Craig et al.(2006)] James Craig, Irene Korol Scala, and William Bevington. *Designing with type: the essential guide to typography*. Watson-Guption Publications, New York, 5th ed., rev. ed edition, 2006. OCLC: ocm62090824.
- [Culy and Lyding(2011)] Chris Culy and Verena Lyding. Corpus Clouds - Facilitating Text Analysis by Means of Visualizations. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, and Zygmunt Vetulani, editors, *Human Language Technology. Challenges for Computer Science and Linguistics*, volume 6562, pages 351–360. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-20094-6 978-3-642-20095-3. URL [http://link.springer.com/10.1007/978-3-642-20095-3\\_32](http://link.springer.com/10.1007/978-3-642-20095-3_32).
- [de Geest and Goris(2010)] D. de Geest and A. Goris. Constrained Writing, Creative Writing: The Case of Handbooks for Writing Romances. *Poetics Today*, 31(1):81–106, March 2010.
- [de Roulet(1993)] Daniel de Roulet. *Virtuellement vôtre!: roman*. Canevas éditeur, Dole [France], 1993.
- [de Roulet(1995)] Daniel de Roulet. *La ligne bleue: roman*. Seuil, Paris, 1995.
- [de Roulet(1996)] Daniel de Roulet. *Bleu siècle: roman*. Seuil, Paris, 1996.
- [de Roulet(1999)] Daniel de Roulet. *Gris-bleu: roman*. Seuil, Paris, 1999.
- [de Roulet(2001)] Daniel de Roulet. Davos Terminus, December 2001. URL <http://www.largeur.com/?p=930>.
- [de Roulet(2005)] Daniel de Roulet. *L'homme qui tombe*. Buchet Chastel, Paris, 2005.
- [de Roulet(2007)] Daniel de Roulet. *Kamikaze Mozart: roman*. Buchet-Chastel, Paris, 2007.
- [de Roulet(2009)] Daniel de Roulet. *Le silence des abeilles: roman*. Buchet-Chastel, Paris, 2009.
- [de Roulet(2012)] Daniel de Roulet. *Fusions: roman*. Buchet/Chastel, Paris, 2012.
- [de Roulet(2013)] Daniel de Roulet. *Ecrire la mondialité: essais*. Baconnière, Genève, 2013.
- [de Roulet(2014a)] Daniel de Roulet. *Le démantèlement du coeur*. Buchet-Chastel, Paris, 2014a.
- [de Roulet(2014b)] Daniel de Roulet. Labo-Cité : rencontre avec Daniel de Roulet - la Simulation humaine, October 2014b. URL <https://www.youtube.com/watch?v=B2OSkahSu1U>.
- [de Roulet(2016)] Daniel de Roulet. Eco-fiction. Comment la fiction permet-elle d'interroger notre rapport aux technologies et à la nature?, January 2016. URL <http://www.espacestemp.net/articles/eco-fiction/>.



- [Dick(1993)] Bob Dick. You want to do an action research thesis?, 1993. URL <http://www.aral.com.au/resources/arthesis.html>.
- [Dumais(2005)] Susan T. Dumais. Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1):188–230, September 2005.
- [Duval and Pétry(2016)] Dominic Duval and François Pétry. L'analyse automatisée du ton médiatique: construction et utilisation de la version française du Lexicoder Sentiment Dictionary. *Canadian Journal of Political Science*, 49(02):197–220, 2016.
- [Dällenbach(1988)] Lucien Dällenbach. *Claude Simon*. Number 1 in Les Contemporains. Seuil, Paris, 1988.
- [Délétréz(2014)] Stéphane Délétréz. Les humanités digitales vont-elles assassiner la littérature ?, April 2014. URL <http://www.rts.ch/la-1ere/programmes/cqfd/5755935-lire-autrement-la-simulation-humaine-22-04-2014.html>.
- [Eden and Huxham(1996)] Colin Eden and Chris Huxham. Action Research for Management Research. *British Journal of Management*, 7(1):75–86, March 1996.
- [Eisner(1998)] Elliot W Eisner. *The enlightened eye: qualitative inquiry and the enhancement of educational practice*. Merrill, Upper Saddle River, N.J., 1998.
- [Ellson et al.(2002)] John Ellson, Emden Gansner, Lefteris Koutsofios, Stephen C. North, and Gordon Woodhull. Graphviz- Open Source Graph Drawing Tools. In Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, Petra Mutzel, Michael Jünger, and Sebastian Leipert, editors, *Graph Drawing*, volume 2265, pages 483–484. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002. ISBN 978-3-540-43309-5 978-3-540-45848-7. URL [http://link.springer.com/10.1007/3-540-45848-4\\_57](http://link.springer.com/10.1007/3-540-45848-4_57).
- [Elson et al.(2010)] David K Elson, Nicholas Dames, and Kathleen R McKeown. Extracting social networks from literary fiction. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 138–147. Association for Computational Linguistics, 2010.
- [Erten et al.(2004)] Cesim Erten, Philip J Harding, Stephen G Kobourov, Kevin Wampler, and Gary Yee. Exploring the computing literature using temporal graph visualization. In *Electronic Imaging 2004*, pages 45–56. International Society for Optics and Photonics, 2004.
- [Eskicioglu et al.(2003)] Ahmet M Eskicioglu, John Town, and Edward J Delp. Security of digital entertainment content from creation to consumption. *Signal Processing: Image Communication*, 18(4):237–262, April 2003.
- [Fan Bao and Jia Chen(2014)] Fan Bao and Jia Chen. Visual framework for big data in d3.js. pages 47–50. IEEE, May 2014. ISBN 978-1-4799-4565-8 978-1-4799-4566-5. doi: 10.1109/IWECA.2014.6845553. URL <http://ieeexplore.ieee.org/document/6845553/>.
- [Fayyad et al.(2002)] Usama M. Fayyad, Georges G. Grinstein, and Andreas Wierse, editors. *Information visualization in data mining and knowledge discovery*. The Morgan Kaufmann series in data management systems. MK/Morgan Kaufmann Publishers, San Francisco, 2002.
- [Fleiss and Cohen(1973)] J. L. Fleiss and J. Cohen. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, 33(3):613–619, October 1973.

## Bibliography

---

- [Frøslev Christensen and Maskell(2003)] Jens Frøslev Christensen and Peter Maskell. *The industrial dynamics of the new digital economy*. E. Elgar, Cheltenham, UK; Northampton, MA, 2003.
- [Geno(1992)] Marie Gontier Geno. *A la française: correct French for English speakers*. University Press of America, London ;New York, 1992.
- [Gilleo and Lind(2016)] Wayne M. Gilleo and Mary Lind. Adoption of Cloud Computing in Firms to Enable Software as a Service.. *International Journal of Operations Research and Information Systems*, 7(4):1–37, October 2016.
- [Glass and Bangay(2007)] Kevin Glass and Shaun Bangay. A naïve, salience-based method for speaker identification in fiction books. pages 1–6, Durban, South Africa, November 2007. PRASA. ISBN 978-1-86840-656-2.
- [Goh et al.(2012)] Hui-Ngo Goh, Lay-Ki Soon, and Su-Cheng Haw. Automatic Identification of Protagonist in Fairy Tales Using Verb. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Pang-Ning Tan, Sanjay Chawla, Chin Kuan Ho, and James Bailey, editors, *Advances in Knowledge Discovery and Data Mining*, volume 7302, pages 395–406. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-30219-0. URL [http://link.springer.com/10.1007/978-3-642-30220-6\\_33](http://link.springer.com/10.1007/978-3-642-30220-6_33).
- [González et al.(2013)] J.C. González, J.L. Guzmán, S. Dormido, and M. Berenguel. Development of interactive books for Control Education. *IFAC Proceedings Volumes*, 46(17): 150–155, 2013.
- [Gottron(2009)] Thomas Gottron. Document Word Clouds: Visualising Web Documents as Tag Clouds to Aid Users in Relevance Decisions. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Maristella Agosti, José Borbinha, Sarantos Kapidakis, Christos Papatheodorou, and Giannis Tsakonas, editors, *Research and Advanced Technology for Digital Libraries*, volume 5714, pages 94–105. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-642-04345-1 978-3-642-04346-8. URL [http://link.springer.com/10.1007/978-3-642-04346-8\\_11](http://link.springer.com/10.1007/978-3-642-04346-8_11).
- [Green-Armytage(2010)] Paul Green-Armytage. A Colour Alphabet and the Limits of Colour Coding. *Colour: Design & Creativity*, 5(10):1–23, 2010.
- [Grevisse and Lenoble-Pinson(2009)] Maurice Grevisse and Michèle Lenoble-Pinson. Majuscules. In *Le français correct : Guide pratique des difficultés*, pages 88–92. Duculot, 6e édition revue et corrigée (12 août 2009) edition, 2009. ISBN 978-2-8011-0051-6. URL <https://books.google.ch/books?id=Cva6FIQ-5nYC&pg=PA88#v=onepage&q&f=false>.
- [Grosjean(2015)] Marianne Grosjean. Daniel de Roulet a roulé sa bosse au bout du monde. *Tribune de Genève*, January 2015.
- [Hall(1989)] Oakley M. Hall. *The art & craft of novel writing*. Writer’s Digest Books, Cincinnati, Ohio, 1989.
- [Hirschberg and Manning(2015)] J. Hirschberg and C. D. Manning. Advances in natural language processing. *Science*, 349(6245):261–266, July 2015.

- [Hoeken and van Vliet(2000)] Hans Hoeken and Mario van Vliet. Suspense, curiosity, and surprise: How discourse structure influences the affective and cognitive processing of a story. *Poetics*, 27(4):277–286, May 2000.
- [Holmes(1998)] D. I. Holmes. The Evolution of Stylometry in Humanities Scholarship. *Literary and Linguistic Computing*, 13(3):111–117, September 1998.
- [Horner et al.(2005)] Robert H Horner, Edward G Carr, James Halle, Gail McGee, Samuel Odom, and Mark Wolery. The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2):165–179, 2005.
- [Hubbs and Hubbs(1953)] Carl L. Hubbs and Clark Hubbs. An Improved Graphical Analysis and Comparison of Series of Samples. *Systematic Zoology*, 2(2):49, June 1953.
- [Ingermanson and Economy(2010)] Randy Ingermanson and Peter Economy. *Writing fiction for dummies*. Wiley, Hoboken, N.J., 2010. OCLC: 806268515.
- [Jansen et al.(2009)] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, November 2009.
- [Jessop(2008)] M. Jessop. Digital visualization as a scholarly activity. *Literary and Linguistic Computing*, 23(3):281–293, September 2008.
- [Jick(1979)] Todd D. Jick. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4):602, December 1979.
- [Jockers(2011)] M Jockers. Detecting and characterizing national style in the 19th century novel. *Digital Humanities 2011*, 2011.
- [John et al.(2016)] Markus John, Steffen Lohmann, Steffen Koch, Michael Wörner, and Thomas Ertl. Visual Analytics for Narrative Text - Visualizing Characters and their Relationships as Extracted from Novels:. pages 27–38. SCITEPRESS - Science and and Technology Publications, 2016. ISBN 978-989-758-175-5. doi: 10.5220/0005669800270038. URL <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005669800270038>.
- [Jovanovic et al.(2014)] Jelena Jovanovic, Ebrahim Bagheri, John Cuzzola, Dragan Gasevic, Zoran Jeremic, and Reza Bashash. Automated Semantic Tagging of Textual Content. *IT Professional*, 16(6):38–46, November 2014.
- [Juola(2006)] Patrick Juola. Authorship attribution. *Foundations and Trends in information Retrieval*, 1(3):233–334, 2006.
- [Karsdorp et al.(2012)] Folgert Karsdorp, Peter van Kranenburg, Theo Meder, and Antal van den Bosch. Casting a spell: Identification and ranking of actors in folktales. pages 39–50, Lisbon, Portugal, 2012. Edições Colibri.
- [Kemeny(1976)] John G. Kemeny. *Finite Markov chains*. Undergraduate texts in mathematics. Springer-Verlag, New York, 1976.
- [Kergosien et al.(2015)] Eric Kergosien, Pierre Maurel, Mathieu Roche, and Maguelonne Teisseire. Senterritoire pour la détection d’opinions liées à l’aménagement d’un territoire. *Revue Internationale de Géomatique*, 25(1):11–34, 2015.

## Bibliography

---

- [Kershner(1964)] Alan M. Kershner. Speed of reading in an adult population under differential conditions. *Journal of Applied Psychology*, 48(1):25–28, 1964.
- [Khondker(2011)] Habibul Haque Khondker. Role of the New Media in the Arab Spring. *Globalizations*, 8(5):675–679, October 2011.
- [Kim et al.(2010)] Nam Wook Kim, Stuart K Card, and Jeffrey Heer. Tracing genealogical data with timenets. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 241–248. ACM, 2010.
- [Kirk and Miller(1986)] Jerome Kirk and Marc L. Miller. *Reliability and Validity in Qualitative Research*. SAGE Publications, Inc., 2455 Teller Road, Newbury Park California 91320 United States of America, 1986.
- [Kirschenbaum(2007)] Matthew G Kirschenbaum. The remaking of reading: Data mining and the digital humanities. In *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD*. Citeseer, 2007.
- [Kline(2003)] Stephen Kline. *Digital play: the interaction of technology, culture, and marketing*. McGill-Queen's University Press, Montréal ; London, 2003.
- [Kluytmans(1998)] Joseph Kluytmans. Pierrick remonte le temps. *Mickey Jeux*, (86), October 1998.
- [Koch et al.(2014)] Steffen Koch, Markus John, Michael Worner, Andreas Muller, and Thomas Ertl. VarifocalReader - In-Depth Visual Analysis of Large Text Documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732, December 2014.
- [Krishnaswamy(2004)] Ajit Krishnaswamy. Participatory research: strategies and tools. *Practitioner: Newsletter of the National Network of Forest Practitioners*, 22:17–22, 2004.
- [Kurkjian and Livingston(2007)] Catherine Kurkjian and Nancy Livingston. Story Characters, Problems, and Settings Are Elemental. *The Reading Teacher*, 60(5):494–500, February 2007.
- [Kuteeva(2011)] Maria Kuteeva. Wikis and academic writing: Changing the writer-reader relationship. *English for Specific Purposes*, 30(1):44–57, January 2011.
- [Kydos and Anastasiadis(2015)] Dimitrios Kydos and Anastasios Anastasiadis. Social network analysis in literature. The case of The Great Eastern by A. Embirikos. In *Proceedings of the 5th European Congress of Modern Greek Studies of the European Society of Modern Greek Studies*, volume 4, pages 681–702, 2015.
- [Lambert(2012)] Joe Lambert. *Digital storytelling: capturing lives, creating community*. Routledge, New York ; London, 4th ed edition, 2012.
- [Landow(2006)] George P. Landow. *Hypertext 3.0: critical theory and new media in an Era of Globalization*. Parallax. Johns Hopkins University Press, Baltimore, 3rd ed. edition, 2006.
- [Leblond(2010)] Aude Leblond. *Poétique du roman-fleuve, de Jean-Christophe à Maumort*. PhD thesis, Université de la Sorbonne nouvelle-Paris III, 2010.
- [Lelièvre(1999)] Jacques Lelièvre. Yoël-Krahn le magnifique. *Mickey jeux*, (80):47–51, June 1999.

- [Lumbroso and Pagès(2016)] Olivier Lumbroso and Alain Pagès. *L'écriture du cycle*. PUPS, Paris, 2016. OCLC: 951102728.
- [Lustig et al.(2006)] Cindy Lustig, Lynn Hasher, and Simon T. Tonev. Distraction as a determinant of processing speed. *Psychonomic Bulletin & Review*, 13(4):619–625, August 2006.
- [Madigan(1984)] Chris Madigan. The Tools that Shape Us: Composing by Hand vs. Composing by Machine. *English Education*, 16(3):143–150, October 1984.
- [Malrieu and Rastier(2001)] Denise Malrieu and Francois Rastier. Genres et variations morphosyntaxiques. *Traitement Automatique des langues*, 42(2):548–577, 2001.
- [Marshall(1997)] Catherine C. Marshall. Annotation : from paper books to the digital library. pages 131–140. ACM Press, 1997. ISBN 0-89791-868-1. doi: 10.1145/263690.263806. URL <http://portal.acm.org/citation.cfm?doid=263690.263806>.
- [Martin DuGard and Camus(1993)] Roger Martin DuGard and Albert Camus. *OEuvres complètes*. Number 113 in Bibliothèque de la Pléiade. Gallimard, Paris, réimpr edition, 1993. OCLC: 831241431.
- [Mason(2010)] Mark Mason. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(3), 2010.
- [McCarty(2011)] Willard McCarty, editor. *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*. Open Book Publishers, Cambridge, 2011. OCLC: 837949146.
- [McNaught and Lam(2010)] Carmel McNaught and Paul Lam. Using Wordle as a supplementary research tool. *The qualitative report*, 15(3):630, 2010.
- [Meadows(2003)] Daniel Meadows. Digital Storytelling: Research-Based Practice in New Media. *Visual Communication*, 2(2):189–193, June 2003.
- [Meisel and Sullivan(2002)] John B. Meisel and Timothy S. Sullivan. The impact of the Internet on the law and economics of the music industry. *info*, 4(2):16–22, 2002.
- [Meister(2014)] Anna-Maria Meister. Multimedia Review: Scrivener 2.0 (MacOS and Windows). *Journal of the Society of Architectural Historians*, 73(1):158–159, March 2014.
- [Merga and others(2014)] Margaret K Merga and others. Are teenagers really keen digital readers?: Adolescent engagement in ebook reading and the relevance of paper books today. *English in Australia*, 49(1):27, 2014.
- [Merriam-Webster, Inc(1998)] Merriam-Webster, Inc, editor. *Merriam-Webster's manual for writers and editors*. Merriam-Webster, Springfield, Mass, completely rev. [ed.] edition, 1998.
- [Michel et al.(2011)] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014):176–182, January 2011.
- [Miller(2004)] Carolyn Handler Miller. *Digital storytelling: a creator's guide to interactive entertainment*. Focal Press, Amsterdam ; Boston, 2004.

## Bibliography

---

- [Mordock and Krasny(2001)] Kalay Mordock and Marianne E. Krasny. Participatory Action Research: A Theoretical and Practical Framework for EE. *The Journal of Environmental Education*, 32(3):15–20, January 2001.
- [Moreno(1934)] Jacob Levy Moreno. *Who shall survive?*, volume 58. JSTOR, 1934.
- [Moretti(2000)] Franco Moretti. Conjectures on world literature. *New left review*, 1:54, 2000.
- [Moretti(2005)] Franco Moretti. *Graphs, maps, trees: abstract models for a literary history*. Verso, London ; New York, 2005.
- [Moretti(2011)] Franco Moretti. Network theory, plot analysis. *New Left Review*, 2011.
- [Moretti(2013)] Franco Moretti. *Distant reading*. Verso, London ; New York, 2013.
- [Morón Arroyo(2002)] Ciriaco Morón Arroyo. *The humanities in the age of technology*. Catholic University of America Press, Washington, D.C, 2002.
- [Moser(1977)] Heinz Moser. *Praxis der Aktionsforschung: e. Arbeitsbuch*. Kösel Ausbildung. Kösel, München, 1977.
- [Mueller(2003)] Erik T. Mueller. Story understanding through multi-representation model construction. volume 9, pages 46–53. Association for Computational Linguistics, 2003. doi: 10.3115/1119239.1119246. URL <http://portal.acm.org/citation.cfm?doid=1119239.1119246>.
- [Mueller(2006)] Erik T. Mueller. Story Understanding. In Lynn Nadel, editor, *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd, Chichester, January 2006. ISBN 0-470-01619-1 978-0-470-01619-0 0-470-01886-0 978-0-470-01886-6. URL <http://doi.wiley.com/10.1002/0470018860.s00082>.
- [Muralidharan and Hearst(2013)] A. Muralidharan and M. A. Hearst. Supporting exploratory text analysis in literature study. *Literary and Linguistic Computing*, 28(2):283–295, June 2013.
- [Muralidharan and Hearst(2011)] Aditi Muralidharan and Marti Hearst. Wordseer: Exploring language use in literary text. In *Fifth Workshop on Human-Computer Interaction and Information Retrieval*, 2011.
- [Mwalongo et al.(2016)] F. Mwalongo, M. Krone, G. Reina, and T. Ertl. State-of-the-Art Report in Web-based Visualization. *Computer Graphics Forum*, 35(3):553–575, June 2016.
- [Neefs and Barel-Moisan(2015)] Jacques Neefs and Claire Barel-Moisan, editors. *Balzac, l'éternelle genèse*. Manuscrits modernes. Presses universitaires de Vincennes, Saint-Denis, 2015.
- [Nield(2015)] David Nield. 9 Minimal Word Processor Apps for Distraction-Free Writing, September 2015. URL <http://fieldguide.gizmodo.com/9-minimal-word-processor-apps-for-distraction-free-writ-1730016034>.
- [Nieva(2016)] Richard Nieva. Instagram Stories, cloned from Snapchat, has 100M daily viewers, October 2016. URL <https://www.cnet.com/news/instagram-stories-snapchat-100m-daily-viewers/>.
- [Noble(2011)] William Noble. *Three rules for writing a novel: a guide to story development*. Write Thought.com, Sanger, CA, 2011. OCLC: 824148428.

- [Ogawa and Ma(2010)] Michael Ogawa and Kwan-Liu Ma. Software evolution storylines. In *Proceedings of the 5th international symposium on Software visualization*, pages 35–42. ACM, 2010.
- [Omid(2011)] Mahmoud Omid. Design of an expert system for sorting pistachio nuts through decision tree and fuzzy logic classifier. *Expert Systems with Applications*, 38(4):4339–4347, April 2011.
- [Ousby(2009)] Louise Ousby. Whatever it takes: an exploration of writing tools and strategies for completing a novel. 2009.
- [Ouvroir de Littérature Potentielle(1990)] Ouvroir de Littérature Potentielle, editor. *La littérature potentielle: (créations, re-créations, récréations)*. Number 95 in Collection Folio @Essais. Gallimard, Paris, 1990. OCLC: 28361056.
- [Ozanne and Saatcioglu(2008)] Julie L. Ozanne and Bige Saatcioglu. Participatory Action Research. *Journal of Consumer Research*, 35(3):423–439, October 2008.
- [Parv(2014)] Valerie Parv. The tortoise and the hare. *Writing Queensland*, (239):10, 2014.
- [Passer(2014)] Christophe Passer. La saga numérique de Daniel de Roulet. *L'Hebdo*, pages 63–64, April 2014.
- [Patton(2002)] Michael Quinn Patton. *Qualitative research and evaluation methods*. Sage Publications, Thousand Oaks, Calif, 3 ed edition, 2002.
- [Pedersen(2001)] Ted Pedersen. A decision tree of bigrams is an accurate predictor of word sense. pages 1–8. Association for Computational Linguistics, 2001. doi: 10.3115/1073336.1073347. URL <http://portal.acm.org/citation.cfm?doid=1073336.1073347>.
- [Pence(2016)] Charles H. Pence. RLetters: A Web-Based Application for Text Analysis of Journal Articles. *PLOS ONE*, 11(1):e0146004, January 2016.
- [Pennebaker et al.(2001)] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71: 2001, 2001.
- [Perec et al.(1995)] Georges Perec, Hans Hartje, Bernard Magné, and Jacques Neefs. *Cahier des charges de "La vie mode d'emploi"*. CNRS éd. ; Zulma, Paris; Cadeilhan, 1995. OCLC: 489634802.
- [Peyret(2016)] Jocelyn Peyret. "La simulation humaine", une ambitieuse saga de Hiroshima à Fukushima, 2016. URL <http://www.sortirdunucleaire.org/La-simulation-humaine>.
- [Plaisant et al.(2006)] Catherine Plaisant, James Rose, Bei Yu, Loretta Auvil, Matthew G Kirschenbaum, Martha Nell Smith, Tanya Clement, and Greg Lord. Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pages 141–150. ACM, 2006.
- [Pury(2011)] Cynthia LS Pury. Automation can lead to confounds in text analysis Back, Küfner, and Egloff (2010) and the Not-So-Angry Americans. *Psychological science*, 2011.
- [Qi Zhu(2013)] Nick Qi Zhu. *Data Visualization with D3.js Cookbook*. Packt Publishing, 2013. OCLC: 862363135.
- [Queneau(2004)] Raymond Queneau. *Exercices de style*. Gallimard, s.l, 2004.

## Bibliography

---

- [Queneau et al.(1982)] Raymond Queneau, Sheila Bourne, and John Crombie. *Un conte à votre façon*. KickShaws, 1982.
- [Reagan et al.(2016)] Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. The emotional arcs of stories are dominated by six basic shapes. *arXiv preprint arXiv:1606.07772*, 2016.
- [Reda et al.(2011)] Khairi Reda, Chayant Tantipathananandh, Andrew Johnson, Jason Leigh, and Tanya Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. In *Computer Graphics Forum*, volume 30, pages 1061–1070. Wiley Online Library, 2011.
- [Robinson(2010)] Peter Robinson. Electronic editions for everyone. In *Text and Genre in Reconstruction: Effects of Digitization on Ideas, Behaviours, Products and Institutions*. Cambridge: Open Book Publishers, pages 145–163. OpenBook Publishers, 2010. ISBN 978-1-906924-24-9. URL <http://www.openbookpublishers.com/reader/64>.
- [Rochat(2014)] Yannick Rochat. Character networks and centrality. Technical report, University of Lausanne, 2014.
- [Rochat et al.(2013)] Yannick Rochat, Frédéric Kaplan, and Cyril Bornet. A social network analysis of Rousseau’s autobiography “Les Confessions”. In *DH 2013 book of abstracts*, 2013.
- [Rommel(2004)] Thomas Rommel. Literary Studies. In Siemens Schreibman, John Unsworth, and Raymond George Siemens, editors, *A companion to digital humanities*, number 26 in Blackwell companions to literature and culture, pages 88–95. Blackwell Pub, Malden, MA, 2004. ISBN 978-1-4051-0321-3.
- [Roser et al.(2007)] Nancy Roser, Miriam Martinez, Charles Fuhrken, and Kathleen McDonnold. Characters as Guides to Meaning. *The Reading Teacher*, 60(6):548–559, March 2007.
- [Rumelhart(1975)] David E Rumelhart. Notes on a schema for stories. In *Representation and understanding: Studies in cognitive science*, volume 211, pages 211–236. 1975. ISBN 978-0-12-108550-6.
- [Rushkoff(2014)] Douglas Rushkoff. *Present shock: when everything happens now*. 2014.
- [Ryan(2004)] Marie-Laure Ryan. Will new media produce new narratives? In *Narrative across media: The languages of storytelling*, pages 337–359. U of Nebraska Press, 2004, 2004. ISBN 978-0-8032-8993-2. URL [http://books.google.ch/books?hl=fr&lr=&id=b7Y-1c0DC6IC&oi=fnd&pg=PA337&dq=Will+new+media+produce+new+narratives&ots=ebrexifkRT&sig=qIMzo8omi2B0EoTYtanrdG3k\\_eQ#v=onepage&q=Will%20new%20media%20produce%20new%20narratives&f=false](http://books.google.ch/books?hl=fr&lr=&id=b7Y-1c0DC6IC&oi=fnd&pg=PA337&dq=Will+new+media+produce+new+narratives&ots=ebrexifkRT&sig=qIMzo8omi2B0EoTYtanrdG3k_eQ#v=onepage&q=Will%20new%20media%20produce%20new%20narratives&f=false).
- [Sack(2011)] Graham Sack. *Quantifying imaginary social forms: Character networks in the 19th century British novel*. PhD thesis, Master’s thesis, Columbia University, 2011.
- [Sack(2014)] Graham Alexander Sack. Character networks for narrative generation: Structural balance theory and the emergence of proto-narratives. In *Complexity and the human experience: Modeling complexity in the humanities and social sciences*, page 81. Pan Stanford, 2014. ISBN 978-981-4463-26-3.



- [Scheirer et al.(2011)] W J Scheirer, A Rocha, R J Micheals, and T E Boulton. Meta-Recognition: The Theory and Practice of Recognition Score Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689–1695, August 2011.
- [Schreibman et al.(2004)] Susan Schreibman, Raymond George Siemens, and John Unsworth. *A companion to digital humanities*. Blackwell Pub., Malden, MA, 2004. OCLC: 184983791.
- [Schöch(2016)] Christof Schöch. Does Shorter Sell Better? Belgian author George Simenon’s use of sentence length, August 2016. URL <http://dragonfly.hypotheses.org/922>.
- [Shejuti et al.(2015)] Silvia Shejuti, June Abbas, Sam Huskey, and Chris Weaver. Storyline Visualization with Force Directed Layout. Chicago, Illinois, USA, October 2015. URL <http://ieevis.org/year/2015/info/overview-amp-topics/posters>.
- [Shixia Liu et al.(2013)] Shixia Liu, Yingcai Wu, Enxun Wei, Mengchen Liu, and Yang Liu. StoryFlow: Tracking the Evolution of Stories. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2436–2445, December 2013.
- [Si et al.(2005)] Luo Si, Tapas Kanungo, and Xiangji Huang. Boosting performance of bio-entity recognition by combining results from multiple systems. page 76, Chicago, Illinois, USA, 2005. ACM Press. ISBN 1-59593-213-5. doi: 10.1145/1134030.1134044. URL <http://portal.acm.org/citation.cfm?doid=1134030.1134044>.
- [Siemens and Schreibman(2013)] Ray Siemens and Susan Schreibman. *A Companion to Digital Literary Studies*. Wiley, New York, 2013. OCLC: 833765852.
- [Smith and Kelly(2002)] Joseph A Smith and Colleen Kelly. Stylistic constancy and change across literary corpora: Using measures of lexical richness to date works. *Computers and the Humanities*, 36(4):411–430, 2002.
- [Snickars and Vonderau(2012)] Pelle Snickars and Patrick Vonderau, editors. *Moving data: the iphone and the future of media*. Columbia University Press, New York, 2012.
- [Stamatatos(2009)] Efstathios Stamatatos. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3): 538–556, March 2009.
- [Stross(2013)] Charlie Stross. Why Microsoft Word must Die, October 2013. URL <http://www.antipope.org/charlie/blog-static/2013/10/why-microsoft-word-must-die.html>.
- [Svensson(2010)] Patrik Svensson. The landscape of digital humanities. *Digital Humanities Quarterly*, 4(1), 2010.
- [Tanahashi and Kwan-Liu Ma(2012)] Y. Tanahashi and Kwan-Liu Ma. Design Considerations for Optimizing Storyline Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, December 2012.
- [Tangherlini(2015)] Timothy R Tangherlini. Facebook for Vikings: Social Network Analysis and Egil’s Saga. *Egil, the Viking Poet: New Approaches to Egil’s Saga*, pages 149–172, 2015.
- [Tapaswi et al.(2014)] Makarand Tapaswi, Martin Bauml, and Rainer Stiefelhagen. StoryGraphs: Visualizing Character Interactions as a Timeline. pages 827–834, June 2014. URL [http://www.cv-foundation.org/openaccess/content\\_cvpr\\_2014/html/Tapaswi\\_StoryGraphs\\_Visualizing\\_Character\\_2014\\_CVPR\\_paper.html](http://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Tapaswi_StoryGraphs_Visualizing_Character_2014_CVPR_paper.html).

## Bibliography

---

- [Tapuria et al.(2007)] Archana Tapuria, Bill Wheeldin, and Multiprofessional Education. The CLEF corpus: semantic annotation of clinical text. 2007.
- [Thompson(2013)] John B Thompson. *Books in the Digital Age the Transformation of Academic and Higher Education Publishing in Britain and the United States*. Wiley, Hoboken, 2013.
- [Turner et al.(2003)] Mark Turner, David Budgen, and Pearl Brereton. Turning software into a service. *Computer*, 36(10):38–44, 2003.
- [Van Rijsbergen(1979)] C. J. Van Rijsbergen. *Information retrieval*. Butterworths, London ; Boston, 2d ed edition, 1979.
- [Viégas and Wattenberg(2008)] Fernanda B. Viégas and Martin Wattenberg. TIMELINES: Tag clouds and the case for vernacular visualization. *interactions*, 15(4):49, July 2008.
- [Vonnegut and Simon(2007)] Kurt Vonnegut and Daniel Simon. *A man without a country*. Random House Trade Paperbacks, New York, random house trade pbk. ed edition, 2007. OCLC: ocm71223500.
- [Warwick(2012)] Claire Warwick. Studying Users in Digital Humanities. In Melissa M. Terras and Julianne Nyhan, editors, *Digital humanities in practice*, pages 1–21. Facet Publishing in association with UCL Centre for Digital Humanities, London, 2012. ISBN 978-1-85604-766-1. OCLC: ocn707962666.
- [Weber et al.(2001)] Marc Weber, Marc Alexa, and Wolfgang Müller. Visualizing Time-Series on Spirals. In *Infovis*, volume 1, pages 7–14, 2001.
- [Weiland(2011)] K. M Weiland. *Outlining your novel: map your way to success*. 2011. OCLC: 781744823.
- [Weiland(2013)] K. M Weiland. *Structuring your novel: essential keys for writing an outstanding story*. 2013. OCLC: 858592420.
- [Whitelaw(2012)] Mitchell Whitelaw. Towards generous interfaces for archival collections. *Comma*, 2012(2):123–132, January 2012.
- [Whyte(1989)] William F. Whyte. Advancing Scientific Knowledge Through Participatory Action Research. In *Sociological Forum*, volume 4. September 1989.
- [Wischenbart et al.(2012)] Ruediger Wischenbart, Carlo Carrenho, Veronika Licher, and Jennifer Krenn. *The global ebook market current conditions & future projections*. O'Reilly, [S.I.], 2012.
- [Wu et al.(2010)] Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and Huamin Qu. OpinionSeer: interactive visualization of hotel customer feedback. *IEEE transactions on visualization and computer graphics*, 16(6):1109–1118, 2010.
- [Xu et al.(2014)] Anbang Xu, Shih-Wen Huang, and Brian Bailey. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1433–1444. ACM, February 2014. ISBN 978-1-4503-2540-0.

- [Youn and McLeod(2007)] Seongwook Youn and Dennis McLeod. A Comparative Study for Email Classification. In Khaled Elleithy, editor, *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pages 387–391. Springer Netherlands, Dordrecht, 2007. ISBN 978-1-4020-6263-6. URL [http://link.springer.com/10.1007/978-1-4020-6264-3\\_67](http://link.springer.com/10.1007/978-1-4020-6264-3_67).
- [Zhu(2001)] Kevin Zhu. Internet-based Distribution of Digital Videos: The Economic Impacts of Digitization on the Motion Picture Industry. *Electronic Markets*, 11(4):273–280, December 2001.



Cyril Bornet  
Chemin des Lilas 4, 1950 Sion, Switzerland  
+41 79 579 52 43  
info@cybor.ch  
[www.cyrilbornet.com](http://www.cyrilbornet.com)

Born on 10-04-1986  
Single, no children  
Swiss and French citizen  
Driver's License  
Linked  profile



*Entrepreneurship, perseverance and efficiency in teamwork are for me natural skills that I refined by founding and administrating my own company. This enriching experience also taught me fundamentals of project management and compliance with economic imperatives. At ease both in development and research fields, I know to bring the necessary attention to details while keeping long-term goals in mind. Having a strong interest and ease in languages and linguistics, I am always passionate about new challenges and currently looking for a long-term position in which my contribution will make a difference.*

## Education

- 2016    **PhD**, Management of Technology, EPFL (DHLAB). (To be defended in January 2017)
- Subject: Changes in digital literature from a writer's perspective. This research aimed at qualifying in which respects digital reading technologies and modern writing tools are reshaping novels writing, and was conducted with the active participation of the Swiss novelist Daniel de Roulet. It lead to the release of innovative text analysis and visualisation tools that were used for the publication of an unique of its kind 300 chapters digital saga ([simulationhumaine.com](http://simulationhumaine.com)).
- Publications list available on the [EPFL Infoscience Portal](http://EPFL-Infoscience-Portal).
- 2010    **MSc** in computer science, specialization in Internet Computing, EPFL.
- 2008    **BSc** in computer science, EPFL.
- 2005    **Maturité cantonale** (bilingual), opt. maths/physics, Lycée-Collège des Creusets, Sion.

## Current positions

- EPFL    **PhD candidate** and **IT manager** at Digital Humanities Laboratory, EPFL (DHLAB). Actively implicated in decision making, planning and deploying network and server network solutions used for most research projects.
- Teaching assistant** for the Master courses Digital Humanities (prof. F. Kaplan) and Business Information Systems (prof. S. Riboni). Improvement and rationalization of course materials preparation, students projects supervision and setup of reusable remote teaching structures.
- Cybor    **Founding and administration** since 2008. The company currently employs 3 people, providing various Internet services (generic as well as dedicated web hosting), custom web-oriented software and mobile applications development (iOS and Android).
- As a software engineer (until 2014), consultancy (analyzing the customers' needs, finding efficient solutions and drawing specifications sheets), undertaking and implementation of innovative solutions meeting the customers' needs.
- As a CEO and project manager (from 2014 on), establishing quotes and contracts, planification and supervision, delivering final products to the clients and ensuring perfect adoption by end users.

## Past experience and academic projects

- 2008 *EPFL* (semester project): Artificial intelligence system generating music playlists for the robotic computer QB1 by OZWE, publicly presented at the «Cité des sciences» exhibition in Paris (Nov. 14th-16th 2008, Grand Palais). Software in C++ characterized by high efficiency, stability and low memory consumption in order to run as an "*Internet of things*" component.
- 2007 *Daily-Soft Ltd*, Lausanne: development and release of the remote teaching website *didactis.ch*, used for several years by private schools Lemania, IUKB and Buissonnets.
- 2006 *Boutique "Au Fil des Mois"*, Sion: development and release of custom ERP software in RealBasic (billing, inventory and basic accounting), used on a daily basis since then.
- 2005 Freelance webmaster, more than 50 websites released, most still active in 2016 and highly ranked (first results page on local Google search in their domains). Several projects in partnership with professional designers and other third parties.

## Technical Skills

**Desktop programming:** C++, Swift / Objective-C, Python, OpenGL, Java, Xojo (RealBasic), Scala, C Ansi, assembler and compilation basics.

**Web programming:** PHP, MySQL, PostgreSQL, JavaScript, XML/XSLT/XHTML, J2EE, LDML (Lasso), networking protocols (TCP/UDP, HTTP, SSH, DNS,...) and testing tools.

**Frameworks and environments:** Laravel, Node.js, Gulp, Semantic UI, Wordpress, Facebook and Twitter APIs.

**Scripting languages:** Shell, Perl, AppleScript/Automator.

**Systems administration:** Web servers (OSX, Linux Ubuntu, OpenBSD, VMWare ESXi). Apache technologies (HTTP Server, Solr/Lucene, Hadoop), BIND, R Shiny.

**Miscellaneous:** LaTeX, Infographics, typography, hardware prototyping (Arduino), movies editing, architectural modeling of buildings, relevant experience in accounting and Swiss fiscal law.

## Extra-curricular activities

Real estate management (budgeting and accounting)

Active contributor to Wikipedia and OpenStreetMap (more than 70'000 edits in 2013-2015)

## Languages

French: Mother tongue, perfect grammar and writing

English: C2; Master/PhD lectures and work environment in English at EPFL

German: C1; Maturité cantonale bilingue

Swiss German: Very good understanding (second family language)

Japanese: A2; Japanese Language Proficiency Test level 3 (2010)

Italian: A2; 1 year in high school (basic knowledge)

Mandarin: A1; Started in 2016

