




Tool criticism in practice. On methods, tools and aims of computational literary studies


J. Berenike Herrmann <berenike_dot_herrmann_at_uni-bielefeld_dot_de>, Universität Bielefeld  <https://orcid.org/0000-0002-5256-0566>

Anne-Sophie Bories <a_dot_bories_at_unibas_dot_ch>, University of Basel  <https://orcid.org/0000-0001-6551-1395>
Francesca Frontini <francesca_dot_frontini_at_ilc_dot_cnr_dot_it(link sends e-mail)>, CNR - Istituto di Linguistica

Computazionale A. Zampolli  <https://orcid.org/0000-0002-8126-6294>
Clémence Jacquot

Steffen Pielström <pielstroem_at_biozentrum_dot_uni-wuerzburg_dot_de>, University of Würzburg  <https://orcid.org/0000-0003-1016-2911>

Simone Rebora <simone_dot_rebora_at_uni-mainz_dot_de>, Johannes Gutenberg University Mainz  <https://orcid.org/0000-0002-1501-3774>

Geoffrey Rockwell <geoffrey_dot_rockwell_at_ualberta_dot_ca>, University of Alberta  <https://orcid.org/0000-0001-7430-4742>

Stéfan Sinclair

Abstract

This paper is a case-driven contribution to the discussion on the method-theory relationship in practices within the field of *Computational Literary Studies* (CLS). Progress in this field dedicated to the computational analysis of literary texts has long revolved around the new, digital tools: tools, as computational devices for analysis, have had here a comparatively strong status as research entities of their own, while their ontological status has remained unclear to the day. As a rule, they have widely been imported from the fields of data science and NLP, while less often being hand-tailored to specific tasks within interdisciplinary settings. Although studies within CLS are evolving to both a higher degree of specialization in method (going beyond the limitations of out-of-the-box tools) and a stronger theoretical modeling, the technological dimension remains a defining factor. An unreflective adoption of technology in the shape of tools can compromise the plausibility and the reproducibility of the results produced using these tools.

Our paper presents a multi-faceted intervention to the discussion around tools, methods, and the research questions that are answered with them. It presents research perspectives first conceived at the ADHO SIG-DLS workshop *Anatomy of tools: A closer look at textual DH methodologies* that took place in Utrecht in July 2019. At that event, the authors discussed selected case studies to address tool criticism from several angles. Our goal was to leverage a tool-critical perspective, in order to “take stock, reflect upon and critically comment upon our own practices” within CLS.

We identified *Textométrie*, *Stylometry*, and *Semantic Text Mining* as three central types of hands-on CLS. For each of these sub-fields, we asked: *What are our tools and methods-in-use? What are the implications of using a tool-oriented perspective as opposed to a methodology-oriented one? How do either relate to research questions and theory?* These questions were explored by case-studies on an exemplary basis.

The unifying perspective of this paper is an applied tool criticism – a critical inquiry leveraged towards crucial dimensions of CLS practices. Here we re-compose the original oral papers and add entirely new sections to it, to create a useful overview of the issue through a combination of perspectives. While we elaborated the thematic connections between the individual case studies, we hope the interactive spirit of an exemplary exchange remains palpable: individual research perspectives shape the case studies reported for *Textométrie*, *Stylometry* and *Semantic Text Mining*, are complemented by further studies showcasing CLS-specific perspectives on *replicability* and *domain-specific research*, and a short section discussing a *tool inventory* as a practical, community-based incarnation of *tool criticism*.

The article reflects thus a rich array of perspectives on *tool criticism*, including the complementary perspective of *tool defense* – arguing that we need tools and methods as a basic common ground on how to carry out fundamental operations of analysis and interpretation within a community.

0. Preliminaries

This paper is a case-driven contribution to the discussion on the method-theory relationship in practices within the field of *Computational Literary Studies* (CLS)^[1] Progress in this field dedicated to the computational analysis of literary texts has long revolved around the new, digital tools: tools, as computational devices for analysis, have had here a comparatively strong status as research entities of their own, while their ontological status has remained unclear to the day. As a rule, they have widely been imported from the fields of data science and NLP, while less often being hand-tailored to specific tasks within interdisciplinary settings. Although studies within CLS are evolving to both a higher degree of specialization in method (going beyond the limitations of out-of-the-box tools) and a stronger theoretical modeling (e.g., [Erlin et al. 2021], [Underwood 2019]), the technological dimension remains a defining factor. An unreflective adoption of technology in the shape of tools can compromise the plausibility and the reproducibility of the results produced using these tools.

Our paper presents a multi-faceted intervention to the discussion around tools, methods, and the research questions that are answered with them. It presents research perspectives first conceived at the ADHO SIG-DLS workshop “Anatomy of tools: A closer look at textual DH methodologies” that took place in Utrecht in July 2019. At that event, the authors discussed selected case studies to address tool criticism from several angles. Our goal was to leverage a tool-critical perspective, in order to “take stock, reflect upon and critically comment upon our own practices” within CLS.^[2]

We identified *Textométrie*, *Stylometry*, and *Semantic Text Mining* as three central types of hands-on CLS. For each of these sub-fields, we asked: “What are our tools and methods-in-use? What are the implications of using a tool-oriented perspective as opposed to a methodology-oriented one? How do either relate to research questions and theory?” These questions were explored by case-studies on an exemplary basis.

The unifying perspective of this paper is an applied tool criticism — a critical inquiry leveraged towards crucial dimensions of CLS practices. Here we re-compose the original oral papers and add entirely new sections, to create a useful overview of the issue through a combination of perspectives. While we elaborated the thematic connections between the individual case studies, we hope the interactive spirit of an exemplary exchange remains palpable: individual research perspectives shape the case studies reported for *Textométrie*, *Stylometry*, and *Semantic Text Mining*. They are complemented by further studies showcasing CLS-specific perspectives on *replicability* and *domain-specific research*, and a short section discussing a *tool inventory* as a practical, community-based incarnation of *tool criticism*.

The *practice of tool criticism*, the evaluation of a tool's suitability to a specific task, is a *sine qua non* in the interdisciplinary setup of a Digital Humanities discipline like CLS, where actors have different degrees of types of technological expertise. Here is where tool criticism caters to “the evaluation of the suitability of a given digital tool for a specific task” [Traub and van Ossenbruggen 2015]. Importantly, the goal is to “better understand the impact of any bias of the tool on the specific task, not to improve the tools performance” [Traub and van Ossenbruggen 2015].

Here, tool criticism is mainly methodological and epistemological: it complements the narrower, domain-dependent practice of *methodological evaluation*, where established ways of testing a method's reliability, objectivity and validity, or precision and recall, respectively, are widely standardized. Tool criticism thus operates in a field of practices where such standards are not conventional yet, but goes beyond them also, by combining a methodological with a *critical* enquiry that *can* be epistemological, but can also extend to financial, pedagogical, disciplinary and other implications. It aims to unearth presuppositions ingrained in the tools that often neither “data scientists” nor “humanities scholars” may be

1

2

3

4

5

6

naturally aware of in interdisciplinary settings. Especially when a tool becomes successful and thus popular, the evaluation of its suitability to specific tasks is often backgrounded or even neglected.

With the 2019 workshop title *Anatomy of Tools*, we chose an analogy from the field of biology to convey our joint aim of *dissecting* case studies with respect to the role of “the tool” respectively. In our work, we noticed that in CLS generally highly diverse incarnations of “tool” are subsumed under one general concept, each tool different in type, specificity, and goal. Therefore, we operationally define the term *tool*^[3] as follows:

7

At the most basic level, in the present paper, a tool is a *computational device used for carrying out “analyses”*. This potentially includes aids for diverse sub-processes such as data collection, data pre-processing, annotation and indexing, as well as “analyses proper”, which may or may not involve frequency counts, algorithmic and statistical modeling, or visualization. A tool is here thus understood as a type of methodological vehicle used for contributing to the pursuit of a particular research goal on some aspect of literary discourse treated as data. In the context of CLS, a tool digitally retrieves, and/or represents, and/or operates upon and/or manipulates literary data,^[4] which principally includes annotations and metadata.^[5] This includes for instance text analysis tools such as Voyant, program libraries for R or Python, as well as general-purpose tools such as Excel spreadsheets or visualization tools.

By contrast to a *method*, a tool is further defined by its typically reified and closed character. As a computational implementation of a method, it exists as a distinctive *entity* with a limited set of functions. A tool often includes a graphical user interface (GUI), which makes it also phenomenologically perceived as a particular device rather than a potentially adaptable set of conditions. Another relevant dimension of tools is its transformative power, as emphasized by Weizenbaum and others: “[T]he tool is much more than a mere device: It is an agent for change” [Weizenbaum 1984, p. 18]. If a tool is successful, which means widely and conventionally used in order to carry out a task, this *change* does not only happen at the level of the *method* of doing something, but potentially also extends to the *object* to which method is applied. The object is then co-constructed by the method – which in turn has effects on the scholarly *subject* using the tool, for example, on their epistemological framework.^[6]

How can actors know enough about the “bias of the tool on the specific task,” [Traub and van Ossenbruggen 2015] how can they gauge its validity, or its capacity for producing plausibility? Or its potential of inducing change in the world? In the present paper, we will not offer many decisive answers to the important questions raised by the use of tools in CLS, and neither can we dive deep into systematic epistemological, media-theoretical, or sociological discussions. However, we do describe critically how a few exemplary tools are applied to several typical research questions within representative sub-fields of CLS. Our aim is to give a practical foothold for a principled *tool-critical awareness*.

8

The paper is structured in the following way: In the introduction (Section 1), we raise issues of “tool criticism” of current CLS at the interface between methodology and theory. Here, we discuss the advantages and disadvantages of the discourse on *tools* vs. *methods*, as well as the dimensions of different types of *user groups*, and a shift from *tools* towards a more integrated perspective on *modeling*. We also make a case for “tool defense.” In the main part, we first present three tool-critical case studies of what we have called “method-driven *schools*”: *Textométrie*, *Stylometry*, and *Semantic Text Mining* (Section 2). Each of its sub sections (Sections 2.1 - 2.3) addresses a pertinent approach, giving an overview of its usability and strengths in the actual research, as well as pointing out problems and formulating specific avenues for further development. Taking a slightly different angle, Section 3 will then discuss the important issue of *replicability* and the range of its potential incarnations in CLS – in between its “humanities” and “science” poles. In the fourth section, we single out an example for one *specific research domain* (poetry), addressing the need for domain-specific tool adaptation in a particular case. The fifth section will present one attempt at a practical solution for handling and gauging methods in Computational Literary Studies, the *SIG-DLS Tool Inventory*, which was designed to provide both a perspective of orientation and criticism. Finally, we draw a summary and conclusion.

9

Introduction

Tool criticism [van Es et al. 2018] [Koolen et al. 2018] has made us aware that the digital tools widely distributed in DH have the power to reify theoretical *a priori*s [Flanders and Jannidis 2019] [McCarthy 2005]. Therefore, the community needs a handle for gauging their validity, or capacity for producing plausibility [Winko 2015], possibly applying a sense of craft [Piper 2017]. But tools are not “just tools” – the current panorama in CLS presents a plethora of computational devices used for carrying out “literary analyses: instruments, protocols and practices for processing, analyzing and visualizing data” (see our operational definition of *tool* above). On a general level, all of these are used to examine aspects of “literature”, but a closer look reveals that the methodological “heirdom” of CLS is a precarious patchwork. Just loose connections link tools from stylometry to those from NLP and computational linguistics, to those from corpus linguistics, and to ones more genuinely developed within literary studies. At the moment we are dealing with a rich, but also atomized situation. Some tools, but by far not all, involve aggregation and statistical models, while others center on visualization, while yet others have a different epistemic approach, centering on implementing a hermeneutic or deconstructivist mode of enquiry. Some tools combine multiple methods.

10

What is more, CLS practices are diverse also in the degree of reduction and generalization. They vary in the way they explicitly address the fit of (digital) data and method to literary modeling [Flanders and Jannidis 2019] [Piper 2018] [Underwood 2019]. The practices run the gamut extending from a “computational”, or “quantitative,” paradigm [Herrmann 2017] with computational linguistics, text mining, quantitative linguistics, and corpus linguistics and an “analog”, “qualitative” paradigm with structuralist, hermeneutic, or deconstructivist approaches. It is evident how these have substantially distinct requirements, intentions and ways of defining the limitations imposed by the digital.

11

As the role of digital tools in these [sic!] type of studies grows, it is important that scholars are aware of the limitations of these tools, especially when these limitations might bias the outcome of the answers to their specific research questions. While this potential bias is sometimes acknowledged as an issue, it is rarely discussed in detail, quantified or otherwise made explicit. On the other hand, computer scientists (CS) and most tool developers tend to aim for generic methods that are highly generalisable, with a preference for tools that are applicable to a wide range of research questions. As such, they are typically not able to predict the performance of their tools and methods in a very specific context. This is often the point where the discussion stops. [Traub and van Ossenbruggen 2015]

In the current paper, we address selected tools in concrete scenarios of application within CLS, as well as the preconditions of *replicability/recapitulation* and *domain specificity* as well as practical questions of how to conduct tool criticism more systematically and openly.

12

There is a fertile ground for this in CLS, which have clearly developed a “sense of tool criticism” that includes an awareness of built-in bias in generic tools “inherited” from Computer Science and NLP [Noble 2018] [O’Neil 2016], and appears to evolve to an explicit sense of digital methodology [Hayles 2012]. Tool criticism, posited by Karin van Es and colleagues as “a rigorous inquiry into the tools used for research” [van Es et al. 2018, 24] is indeed already becoming an “essential element of the overall research process” [van Es et al. 2018, 24]. Accordingly, more emphasis has been put on the adaptation and scaling of methods to specific research questions, combining reflected and critical approaches with the affordances of digitization, constructively managing constraints. Most recently, the replication-debate that originated from psychology has demonstrated that common and explicit criteria, or at least a differentiated conversation about underlying axioms, are desirable (see replies to “the Da-paper” [Da 2019]^[7]).

13

Interpreted as a contribution to serious scholarship, the – also clearly polemic – paper by Nan Z. Da^[8] highlights that constructive criticism is needed for methodological and theoretical progress, which examines whether a certain method is actually applied well to some type of research question. We believe that the time is ripe for a constructive “literary” method criticism, which can foster a uniquely humanities perspective of digital enquiry. One of the most interesting positions emerging from our workshop is that of tools as – ideally – well-calibrated instruments for “getting things done in CLS.” From here, we have started thinking along the lines of a *tool defense*, which concerns the balance between methodological and content-related research. Driven by the affordances of the digital transformation, CLS has so far understandably put an emphasis on the “how” – the development of methodology, as well as digital resources and

14

standards. Meanwhile, the “what,” that is, historically contextualized and systematic enquiries about literary texts, structures, and discourse have not received equal attention. Looking out at the future of CLS, we realize that the discussion about tools directly relates to the definition of digital literary studies as a discipline: is it predominantly a methodological experiment, or is it a serious data-driven, digitally enhanced, enquiry into “things literary?” Starting to answer these questions means to address strategies of tool evaluation, but also standards of methodological training.

In light of these issues, when juxtaposing our section 2 on *Textométrie*, *Stylometry*, and *Semantic Text Mining* we also asked *How do practices relate to different types of actors in the field?* In CLS, scholars originally trained in literary studies typically approach methodology differently from scholars originally trained in data science. For example, our discussion of textométrie (2.1) raises very different questions from that of semantic text mining (2.3). At the same time, there are clear differences between the statistical models applied in the different “computational fields” – NLP, corpus linguistics, and computational linguistics, which only in part coincide with the division between explanatory / exploratory / predictive perspectives.

Scholars, by discipline and training, may vary in their research focus, some preferring to solve methodological questions, while others concentrate on questions about periods, stylistic features, or the history of ideas, to name a few. While it is indispensable to understand the way a tool works at a fundamental level, not every scholar is interested in the development of methodology for its own sake. Along these lines, an important point made in the discussion was *that we need a well-calibrated arsenal of instruments* in the hands of a *scholarly majority*: transparent tools/methods as a reliable basis for the emerging mainstream, who rather than advance a method, wishes to pursue literary research questions. Such methods fulfill multiple functions, one of which is providing a basic common ground on how to carry out fundamental operations of analysis and interpretation.^[9] This is precisely where we have identified the perspective of *tool defense*. Section 2 aims at striking a balance in between criticism and defense.

At the moment, one of the most pertinent questions in the broader field of “science” is that of *replication*, addressed by section 3, which emphasizes that digital humanistic enquiry may, however, entail a qualitative, sometimes even emphatic, “understanding” of cultural phenomena embedded in rich historical contexts. Not only where interpretation involves the appreciation of texts or parts of texts through aesthetic and otherwise hermeneutic processes, we have to deal with an irreducible subjectivity. But there is a clash with “replicability” approached as one of the criteria of the scientific method.^[10] In CLS, we thus need to discuss whether, when and where we want to strive for “replicability” based on accounts of “objectivity” – and whether, when and where the softer criterion of “intersubjectivity” may be more adequate. Section 3 prompts questions such as: *What is the status of replicability in CLS, and are there possibly specific “literary” types of replicability? What culture of replication do we want to develop in the humanities? What traditions can we draw on?*

Closely related to the need for standards and ready-made tools in CLS is that of *domain specificity* addressed by section 4: While CLS should at one level advance towards a “standard” inventory of methods and tools, in ultimate instance valid and meaningful results can be achieved only if sources and methodologies are tailored to the specific research questions, which in turn are nested within domains of expert knowledge. Thus, a practical CLS tool criticism can hardly avoid factoring in the affordances of “target domains” such as prose, drama, and poetry: *How can the gap between domain generality and specificity be bridged in practice?* Section 4 singles out one specific domain, poetry, as an exemplary use case. Section 5 on the *DLS-Tool Inventory* follows by illustrating an ongoing initiative to address documentation, comparability and community-driven evaluation: *How can we systematically record and compare methods and tools, gauge their usability and the fundamental assumptions incarnated?* It showcases one attempt at making available tools as embedded in concrete case studies, offering a foothold for judging their applicability, as well as practical replication and/or recapitulation of particular studies.

2. Three Methodological Schools within CLS

Based on what has emerged so far from the *DLS-Tool Inventory* and our observations of research practices, we have identified three “schools” that coincide with different types of (handling) digital tools – or indeed methods: *Textométrie*, *Stylometry*, and *Semantic Text Mining*. In the following sub-sections, three short tool-critical case studies will address

each of these by practical example, addressing its advantages and limitations.

2.1 Textométrie: Applying a general tool to a specific research question

“Textometric tools” are widely used to explore literary corpora by a mix of quantitative and qualitative methods. Textométrie is an approach to statistical text analysis, developed in France during the 1970s, following lexicometry, a traditional statistical approach to study lexical particularities of literary texts, such as common words, hapax legomena, and specific keywords [Guiraud 1953] [Lafon and Muller 1984]. In addition, textometry applies methods of data analysis, for instance factor analysis or clustering, that enable mapping of words and texts as they are similar or opposed to each other within a corpus.^[11] Its objective is to produce interpretable statistics of textual data, in a contrasting perspective. Bolstered by the development of accessible software platforms incorporating textual statistics, textometry has produced a number of relatively generic tools, alongside a productive body of research in the domain of stylistics and corpus linguistics [Heiden et al. 2010]. As textometry involves the principled interaction between quantitative (reductive) and qualitative (contextualizing) enquiry [Pincemin 2011], it is an attractive approach for scholars from the field of stylistics, which is traditionally qualitative and interpretative. In mixed methods, they are able to capitalize on pattern detection of quantitative operations for larger-scaled studies and re-contextualization that often leads to original observations.

20

Currently, the most pertinent textometric tools are probably *Hyperbase*^[12] (Etienne Brunet, University of Nice) and *TXM* (Serge Heiden, Ecole Normale Supérieure of Lyon). In the following, we will report on a study carried out with *TXM* (version 0.5). We chose *TXM* as it may be referred to as a “canonical tool.” It is a text analysis environment which is well documented, available in open-source, and compatible with texts encoded in XML. It also comprises a graphical client based on CQP and R and is available for Microsoft Windows, Linux, Mac OS X and as a J2EE web portal.^[13]

21

In the following, we will reflect on using *TXM* as a comprehensive, out-of-the-box tool for a diachronic stylistic analysis by applying it to one exemplary study of French poetry. *How can the gap between domain generality of the tool and the specificity of the research question be bridged in practice? What advantages and disadvantages need to be faced?*

22

In our *TXM* case study, we explore the poetic style of the poems written by the French poet Guillaume Apollinaire (1880-1918) from a diachronic perspective. The literary analysis of the poetry of Apollinaire is meant as an illustration only: this subsection primarily aims to highlight and to scrutinize the strengths and the limitations of the *TXM* tool. Apollinaire is one of the most eminent French poets of the early twentieth century. He is particularly known as an important figure in the renewal of poetic forms alongside contemporary avant-garde movements in art (cubism, dadaism). His poetry is marked by *discontinuity*, *heterogeneity* and *fragmentation*, similar to other authors of this period (e.g. Romain Rolland, Salmon, Jacob). One important dimension for interpretation and analysis of his poetic writing is “perpetual reorganization”: texts being rewritten, reused in other contexts, transformed from prose into poetry, see [Debon 2008] [Décaudin 1969] [Follet 1987] [Jacquot 2012] [Moore 1995]. These sign-marks of Apollinaire's writing appear at several levels, including the composition of the collections, different types of “tone”, lexical resources, influences, and thematic issues.

23

Usually, analyses are based on thematic differences within Apollinaire's work, especially between his two main collections: *Alcools* (1913) and *Calligrammes* (1918). Approaching style as a dynamic factor [Herschberg-Pierrot 2006] [Jenny 2011], our case study deliberately adopts a “fresh” look, with the intention to be unbiased with regard to a thematic approach, centering in on the diachronic evolution of writing.

24

- *If style is conceived of as a dynamic factor, how may Apollinaire's heterogeneous poetry be analyzed textometrically as a homogeneous unit?*
- *How may the evolution of Apollinaire's style be textometrically traced over the years?*
- *Can we textometrically re-assess value judgments predominantly based on thematic aspects?*

We depart from the hypothesis of a temporal evolution of stylistic, specifically, syntactic, structures in Apollinaire's poetic oeuvre. As there are issues with the clear temporal reference of Apollinaire's texts, we resort to the criterion of “relative dating,” enriching the corpus by temporal metadata. Going by publication dates, Apollinaire's anthomous poetic oeuvre appears in several collections of varying scope over a period of about eight years: *Le Bestiaire ou Cortège d'Orphée*

25

(1911), *Alcools* (1913), *Vitam impendere amori* (1917) and *Calligrammes* (1918).

Metadata covering just these dates would present a distortion of the genealogy of writing, however. *Alcools*, for instance, a striking example of the phenomenon of perpetual rewriting, was published in 1913, but actually collects poems originally composed over more than a decade (1898 — 1913).^[14] This temporal heterogeneity poses a substantial problem for designing a metadata schema: dates need to be provided for each text in order to facilitate a diachronic contrastive analysis of the corpus.

26

In order to get a handle on a potential evolution of Apollinaire's style, it appears fruitful to consider syntactic configurations. Striking a balance between descriptive accounts of style, for instance, *discontinuity* and *fragmentation*, and the affordances of a “formal” tool such as TXM, we focused on the distribution and the use of grammatical words and especially what we will call here the “syntactic link markers”, i.e. the lexical links between sentences, for example conjunctions and relative pronouns. Interestingly, the use of these grammatical words is explicitly mentioned in Apollinaire's poetics, for example, in 1915, where he posits a “simplification of the syntax of poetry” [Apollinaire 2005, 77]. Our method is thus to compare several sub-divisions of Apollinaire's poetic corpus for the distribution of grammatical words, but paying special attention to the date of writing, which needs to be manually re-attributed to each poem. In TXM, we thus use basic database functions to shape the representation of the data to our needs.

27

The first step in using TXM is corpus construction and pre-processing. This includes tokenization, lemmatization, and automatic annotation of textual features, such as part of speech, as well as the enrichment by metadata such as authorship, data of publication, genre, text collection. It is thus pivotal that TXM “takes care” of important analytical steps, the researcher relying on the sub-modules and parameter settings of the “general tool” provided for the community. In our annotation of the Apollinaire corpus, we aim to represent its “original structural organization” from the largest unit, the subset constituted by Apollinaire's poetic collections (*Alcools*, *Bestiaire*, *Calligrammes* ...) down to that of the “verse”: title of the poetry collection > section > subdivision (if needed) > title of the poem > stanza > verse. These structural metadata are completed by information about textual particularities, especially at the “poem”-unit level. We use six units (optional marked with *): title, subtitle*, epigraph*, date of writing*, date of edition, and date of publication (see [Jacquot 2017]).

28

The “research logic” of TXM requires systematic temporal data. Our relative dating – defined as the marking up of each text according to its writing date, either supposed or authorized [Jacquot 2017, 91] – has the advantage of transcending the potentially problematic “unity” of texts created by the poetry collections.^[15] The collection *Alcools*, for example, is based on a set of themes and alternates cycles and inspirations. Using the TXM database detaches the poems from their previously perceived unit (“the collection”) and thus liberates them from efficacious prior ascription, facilitating new perspectives. The textometric analysis works on sub-sets indexed by the newly added temporal attributes, potentially identifying, in the various collections, new organizations of Apollinaire's poems.

29

For the diachronic analysis, we apply TXM's “specificity calculation” [Lafon 1980], which highlights the positive as well as negative specificities in terms of a defined set of linguistic markers, comparing all temporal subsets with each other.

30

Building upon the part-of-speech-tagging in TXM (launched at ingesting a corpus), and our division of temporal sub-sets (by year), we examine the diachronic distribution of syntactic link markers (for example: *lorsque*, *quand*, *qui*, *que*, *parce que*). Figure 1 shows positive and negative specificities in the distribution of “syntactic link markers” (significant over- vs. underuse in comparison with the other temporal subsets). It highlights clear trends in the way Apollinaire uses the syntactic link markers.

31

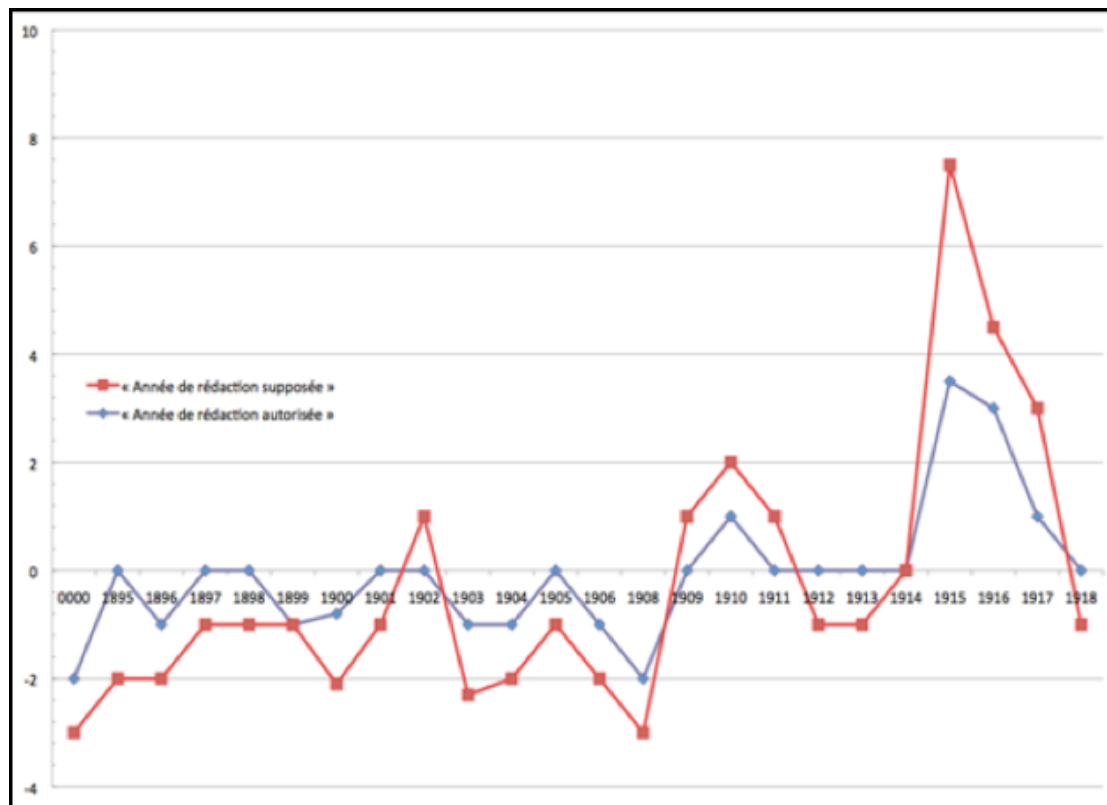


Figure 1. Diachronic distribution of subordinating conjunctions in Apollinaire's poetic corpus "specificity calculation." Comparison between "Supposed Writing Years" (in red) and "Authorized Writing Years" (in blue)

The figure also depicts a comparison between the distribution specificities for the metadata "Authorized Writing Years" (in blue), and "Supposed Writing Years" (in red).

For the "assumed writing years", the graph reveals a striking over-representation of the style markers in 1915, 1916, and 1917, and even the "Authorized Writing Years" (in blue) depict a significant overuse in 1915 and 1916. The year 1917 shows a difference between the assumed and the authorized writing date — the latter does not deviate much from a statistical chance result, falling into the "banality" zone (in textometry, the significant specificity scores are usually below -2 and above +2). It is on the basis of this corpus's analyses that we choose, through several pilot studies, to favor the "assumed writing years" for our stylistic studies. Adding relative temporal data indeed seems to provide us with genealogical-philologically more appropriate and often more meaningful results than the authorized year of writing (which is after all, most often more artificial).

32

The data (including further features, see [Jacquot 2017]) show a trend that is in line with the author's poetics: Over time, stylistic links become more common in Apollinaire's poems. This finding actually defies the idea of a deep heterogeneity, as it points to the increase of devices used for creating cohesion and coherence, at least at the stylistic level. The data also show that the attributed dates render more pronounced amplitudes than the authorized dates.

33

Our example of textometric study shows that a generic tool can be put to "specific use" when this use is based on intricate expert knowledge (here about how to handle the metadata). Using a tool like TXM allows respecting the complexity of the literary subject, even if it — as in our case, the dating of Apollinarian texts — is fragile. The scholar is not relieved of their critical responsibility. As a tool capable of processing structured data, TXM allows for the modeling of data based on criteria that are not always explicit. Entering the estimated dates into the database has the potential (and risk) of their reification, enhanced through visualization and ensuing analysis, especially where no automatic reflection is built into the analytical process regarding the basis for temporal attribution. At the same time, the textometric analysis offers several opportunities for a heuristic change of perspectives, specifically through

34

simultaneous representation of estimated and authorized dates, which allows for systematic comparison of different types of ascription. For users from stylistics, TXM has the advantage of flexibly partitioning the corpus into several different sub-corpora, comparable to each other. It should be mentioned that the whole analysis hinges on pivotal pre-processing steps, such as tokenization, lemmatization, and part-of-speech tagging. The accuracy of these automatic steps needs to be rigorously gauged (are the words correctly separated, including compositional lexical units that cross white space boundaries? Are parts of speech correctly identified?) The same holds at a different level for the research logic tacitly presented (e.g., analytic relevance of basic unit “word,” of “part of speech”). These present possible “blind spots” that users need to be aware of and are required to thoroughly check in each analysis.

2.2 Word Frequencies in Stylometry

Stylometry uses a series of tools and methods for the statistical analysis of style, based on advanced calculations on stylistic properties of texts, whereby mostly focusing on word frequencies. Its main applications have been both authorship attribution and distant reading. Initially developed through the use of spreadsheets, stylometric methods have been fully implemented into programming languages such as R and Python [Eder et al. 2016] [Evert et al. 2017], and enhanced by a wide variety of visualizations, derived from research fields such as phylogenetics and network theory [Eder 2017]. However, from a very pragmatic point of view, it can be stated that the most widely and frequently used “tool” in stylometry are simple algorithms that count words, on which methods like “Delta distance” [Burrows 2002] are applied.

In the abovementioned, fairly dramatic, critique of computational literary studies, [Da 2019] made a controversial case against the application of quantitative methods to literary texts.^[16] She argues that much work in this field essentially boils down to “counting words.” We agree that this view is somewhat reductive, but also find it not without merit: it certainly applies to much of the present-day approaches that are dominant in stylometry and, consequently, to many of the tools that are available. While this methodological focus is to some extent justified by empirical work, the under-explored options for stylometry to move beyond naive word counting need specific attention. Stylometrists, for instance, often take pride in the fact that their tools typically work on raw texts that require little preprocessing. In this, stylometry ignores much of the achievements of literary theory in the twentieth century, such as the importance of focalization or the (actual or implied) reader [Bal 2017] [Genette 1972] [Miall 2018].^[17] Richer (pre)processing pipelines, that also tap into syntax and discourse, might allow stylometry to revitalize its connection with literary theory, but come with significant barriers for non-Anglo-Saxon literatures.

In their thought-provoking article, [Hirst and Feiguina 2007] already proposed a method that took into consideration not only word count, but also syntactic tags as features for stylometric analysis. By working on the results of a partial parser, which performs “a structural analysis that is more than mere chunking but less than the parse tree of a fully recursive grammar” [Hirst and Feiguina 2007, 408], it became possible to “measure” the syntactic structure of sentences. Hirst and Feiguina tested this method on the attribution of the works of the Brontë sisters, reaching some very promising accuracy scores. Curiously enough, however, the method did not find extensive application in the following years. While [Eder 2015] cited it as one of the most promising methods, approaches like *Delta* (a similarity measure based just on the most frequent words) have dominated the field of stylometry, with just a few attempts [van Cranenburgh 2012] [Frontini et al. 2017] to include part-of-speech tags and phrase-structures as well.

The main reason for such a forgetfulness is the fact that similar approaches have proved quite inefficient when compared to others in competition-like setups. In the computational linguistics community, the most relevant of these setups is PAN (*Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection*),^[18] a competition held each year in the context of the Conference and Labs of the Evaluation Forum (CLEF) conference, where teams of programmers are invited to write scripts to solve a series of shared tasks, generally focused on authorship attribution. This “shared task” can be considered in general as a good research practice, as it implies a criticism of tools that are constantly rethought and improved through competition [Gius et al. 2019].

Even if the final goals of stylometry clearly move away from simple authorship attribution, aiming at a quantitative characterization of authorial style, it should be noted how attribution is still the most frequently adopted task when

35

36

37

38

39

verifying the efficiency of stylometric methods (cf. [Burrows 2002] [Eder 2015] [Evert et al. 2017]). In fact, if it cannot be proven that a computer is actually able to tell apart the style of two authors, any approach that adopts it to study authorial style can be easily dismissed as frail or inconsistent.

An overview of the methods presented for the authorship identification task at PAN 2014 notes how the features preferred by participants are “low-level measures”, such as “character measures (i.e., punctuation mark counts, prefix/suffix counts, character n-grams, etc.) or lexical measures (i.e., vocabulary richness measures, sentence/word length counts, stopword frequency, n-grams of words/stopwords, word skip-grams, etc.)” [Stamatatos et al. 2014, 16]. Only one approach was based on “high-level” features, obtaining some of the lowest scores. Hirst himself participated in the 2012 PAN competition, getting the worst result in the “authorship attribution” task.^[19] Finally, in one of the most recent PAN competitions, focused on multi-authored fanfiction texts, just a few teams used POS tags and they were clearly outperformed by those who chose the simplest methods, based on character and word n-grams [Kestemont et al. 2018]. In conclusion, notwithstanding the originality and the theoretical consistency of the idea, the approach seems to prove inefficient when applied to an effective measuring of style. Reasons may be many, starting from the necessary automated (pre-)processing of texts, which might generate errors in itself. In addition, overfitting may be a source of error in machine learning approaches (that are more and more frequently used in stylometry): a combination of too many high-level features can make a model strong for a specific case study (see the Brontë sisters in [Hirst and Feiguina 2007]), but can also hinder its generalizability. From this point of view, simple wordcount might prove less insightful, but much more stable. Finally, it is undeniable that, while NLP tools have been developed extensively for English, efficient applications to other languages are still missing, where research is a few – if not many – steps behind. Therefore, while wordcount proves reliable whatever the language of application [Eder 2015], the same cannot be said with approaches that rely on “high-level” features.

40

However, during the past few years, the interest towards more “linguistically informed” approaches has never fully extinguished. In fact, there is the impression that state-of-the-art stylometric methods work efficiently not because they are able to catch the very nature of style, but just because they scrape the surface of a phenomenon that has much more profound implications. In fact, by modeling stylistic distance as a statistical difference in the frequency of use of some words (or characters), stylistic choices such as syntactic construction and discourse structuring might be implicitly modeled (as they determine these very frequencies). However, such implicit modeling catches them only indirectly, and thus incompletely (see also [Bubenhofer and Dreesen 2018]).

41

In recent years, positive experimental results have finally started to support this theoretical need. In the PAN 2019 Cross-Domain Authorship Attribution Task (once again, focused on fanfiction texts in four different languages), one team proposed a fruitful integration between word- (and punctuation-) count, stemming, text distortion, and POS-tagging. The tagging was performed using Spacy,^[20] which reaches “an accuracy that varies from 95.29% to 97.23%” [Bacciu et al. 2019]. Notwithstanding the errors generated by the tagger, the approach ranked second (and even first for Spanish texts, cf. [Daelemans et al. 2019]), confirming how, with an improvement of NLP tools, the dream of a stylometry finally able to move beyond wordcount and reach some more theoretically-pregnant dimensions might not be a dream anymore.

42

2.3 Semantic Text Mining

Semantic Text Mining methods, such as sentiment analysis, topic modelling, and word embeddings, apply tools for text analysis and visualization based on external semantic information and co-occurrence methodologies. They offer the potential of addressing key questions in literary theory and narratology, from the identification of genre to the visualization of plot. These emerging approaches are now beginning to broaden the scope of computational literary studies and to open up new, still unexplored, potentialities. But it is not only these potentialities that still await exploration, but also the limitations and caveats of the methods' application in humanities research that in many cases require more systematic investigations.

43

Among these methods, topic models based on Latent Dirichlet Allocation (LDA) and Gibbs Sampling [Blei 2012] [Steyvers and Griffiths 2007] have become particularly popular in digital humanities research in recent years (see e.g.

44

[Binder and Jennings 2014] [Mitrofanova 2015] [Schöch 2017]). They allow for explorations and analyses of the content and the semantic structure of digital text corpora. They permit researchers to model a corpus' content in terms of so-called “topics,” groups of words which are apparently semantically related, and show the distribution of these topics within the corpus. By observing single, meaningful topics, scholars can scan large text collections for documents relevant to a specific discourse, or estimate how the prominence of such a discourse developed along the time axis (e.g. [Pavlova and Fischer 2018]). Texts can be sorted according to their content, and it is also possible to derive content-related features for text classification from topic models (e.g. [Henny et al. 2018]). Thanks to an increasing number of available tools and libraries, the method is, by this day, accessible to a wide range of users^[21] (Figure 2).

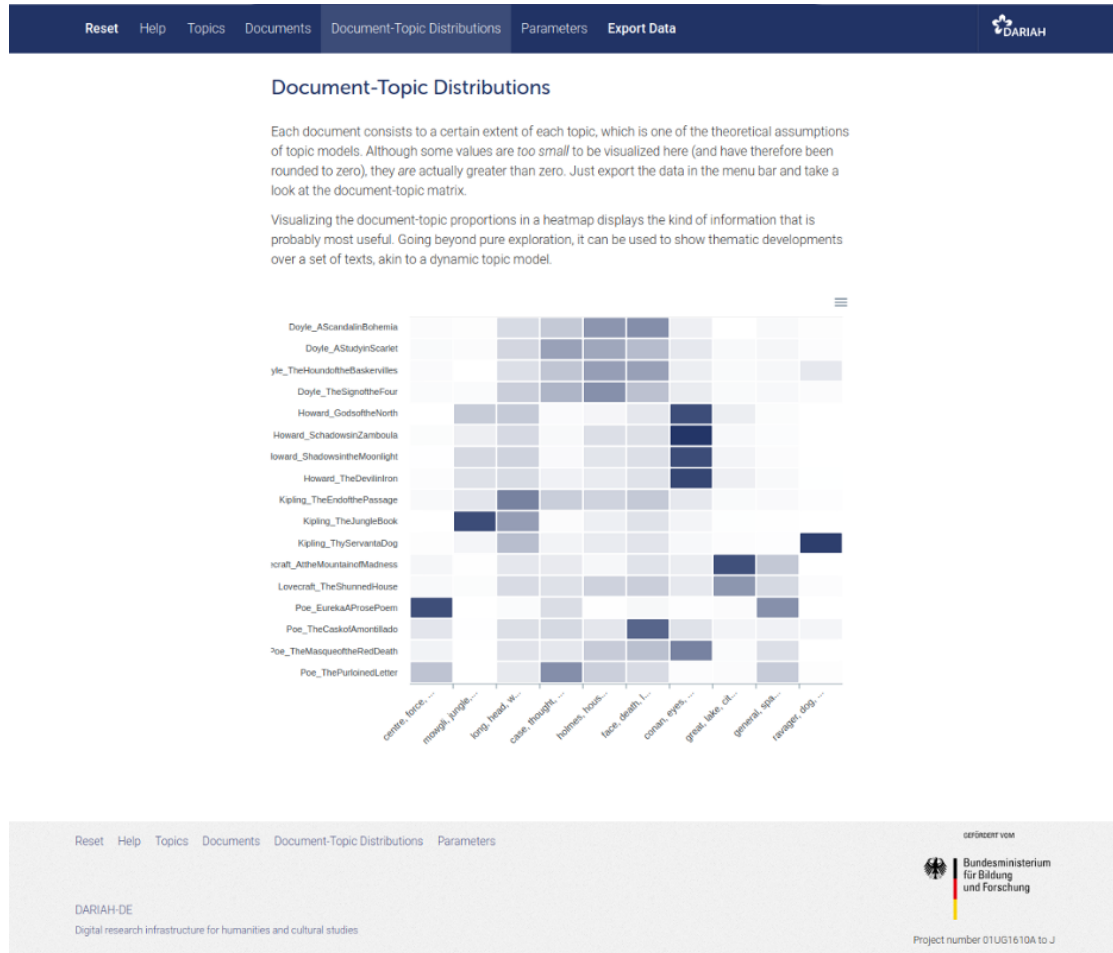


Figure 2. Interface of the DARIAHTopicsExplorer as an example for how accessible topic modeling nowadays is to users. Among other results, the interface produces a heat map overview of the distribution of topics in the corpus (here, a small collection of English short stories). This heat map shows us, among other things, that a topic in which the words “mowgli” and “jungle” weigh most heavily is very present in Kipling’s *The Jungle Book*.

The popularity of topic modeling in digital humanities combined with its technical accessibility may suggest that it is a well understood method – safe to use without a second thought – but this impression is misleading. Basically, there are three issues to consider when using topic modeling: (1) in order to use it, one has to make decisions that require a basic understanding of the algorithm, (2) for many of these decisions there are still no best practices and recommendations rooted in systematic, empirical, methodological research, and (3) attempts to “validate” a computational method for semantic analysis is generally not unproblematic.

(1) The methodology of topic modeling is rather intricate: LDA topic models are based on relatively advanced probabilistic procedures, which are probably understood only vaguely by many users. The method requires a number of informed decisions that have to be taken by these users – about the number of topics, iterations, chunk sizes, even hyperparameter settings and the innumerable possible ways to preprocess the texts.

(2) Moreover, even after years of application in the field, it is still hard to find clear recommendations and best practices concerning the choice of model parameters or preprocessing steps in topic modeling. As [Du 2019] points out in a survey on how topic modeling has been used in digital humanities in the past, many studies do not even bother to report the parameter settings of their experiments. If experiments are reported in more detail, studies often copy the decisions that seemingly worked well in other studies. The field would certainly benefit from more thorough methodological research providing empirical grounds for clear and systematic recommendations and best practices tailored towards the specific demands of the field.

47

(3) But there is a more fundamental problem that affects the search for parameters that produce good topic models: the impossibility to clearly define a good topic model. In semantic text mining we attempt to map a mathematical concept onto the much more elusive phenomenon of semantic meaning.^[22] Human readers are often stunned by the combinations of words found in topic models. One of the most popular German tutorials on topic modeling introduces the concept with an example topic composed of the keywords “theater, actor, play, role, applause...”.^[23] It is obvious to the reader that grouping these words together seems to be meaningful, and that is why scholars are attracted to topic modeling. But to evaluate the method, it is necessary to quantify how meaningful this combination is compared to others.

48

Other areas of text mining can deal with that kind of problem much more straightforwardly. To evaluate a method for authorship attribution or POS-tagging, we can create test sets containing instances the classes of which we already know and observe how many texts are attributed correctly to their authors, or how many words are tagged correctly with their respective POS tags. But what is the “correct” semantic description of a fictional text? Some approximations have been tried. Probably the closest to the intuition of meaningfulness of a topic is the word intrusion method [Bhatia et al. 2018] [Lau and Baldwin 2016], where a random word is added to a topic and human annotators are tasked to identify it. If a random intruder is easy to spot, the original words of the topic are considered semantically close. Another popular idea is that a meaningful interpretable topic can be identified by measuring the semantic coherence of its keywords-based co-occurrence in an external source (usually wikipedia), e.g. in terms of Pointwise Mutual Information (PMI). A third approach is based on previous knowledge about the internal structure of an evaluation corpus: if the “content topics” of the texts in a collection are known in advance, thanks for example to key words provided by the authors, a good topic model could be defined as one that allows a classifier to attribute texts to keywords using the topic distributions as features [Schöch 2017]. These are all but approximations to our intuition of “meaningfulness,” and future research may even show that they can produce contradicting recommendations in some cases.^[24]

49

Despite the current challenges, topic modeling is a useful tool for exploring the content – literally speaking, the *topoi* – of a corpus. It thereby can, without a doubt, considerably assist us in hypothesis generation. However, the fact that this method has both been factually established in the digital humanities community for some years and is relatively accessible should not lead us to believe that our understanding of how to apply it in our field is exhaustive and complete, and that we can reliably base matter-of-fact statements with ease on the results of topic modeling. To get to this point, further research dedicated specifically to the methodology of topic modeling and its application on research questions in digital humanities will be required.

50

3. Recapitulation, Replication, Reanalysis, Repetition, or Revivication

In Section 2, we examined textometry, stylometry, and semantic text mining as three subfields of CLS that are driven by methodology. The different sub sections each showed how very different tools can indeed be well-suited to the intended tasks, requiring, however, both methodical skill and expert knowledge of the content under scrutiny. The required knowledge of the inner workings of the tool at hand was increasingly complex from textometry over stylometry to topic modeling, while the required domain knowledge was almost inversely graded from Apollinaire over general style markers to recurrent topics in a small collection of English short stories. A main difference was that the Apollinaire study is a fully fledged case study, while stylometry and semantic text mining each discuss more generically the implications and limitations of their tools.

51

For all types of studies, however, pre-processing steps, such as tokenization, lemmatization, and part-of-speech tagging emerged as being of high importance. They need to be rigorously gauged and even epistemologically reflected, as for example the analytic relevance of basic unit “word” and “part of speech.” It was demonstrated that methodological and analytical experimentation is fruitful beyond the single “word” level across approaches. However, this is where we enter into the realm of methodological evaluation, testing whether method is tailored to its task. In many fields, this includes the fundamental question whether replication of a study's results is possible. In the present section on *Recapitulation, Replication, Reanalysis, Repetition, or Revivication*, we will discuss an example of how replication can in fact be addressed in a digital humanities setting such as CLS.

In 1973 John B. Smith [Smith 1973] published an article on “Image and Imagery in Joyce's Portrait: A Computer-Assisted Analysis” with at its heart a visualization of the intensity of verbal images in the text. As Smith points out, commenting on the visualization, the richness of the imagery peaks at the end of Chapter 1, as he expected, at the “pandybat” episode. [25]

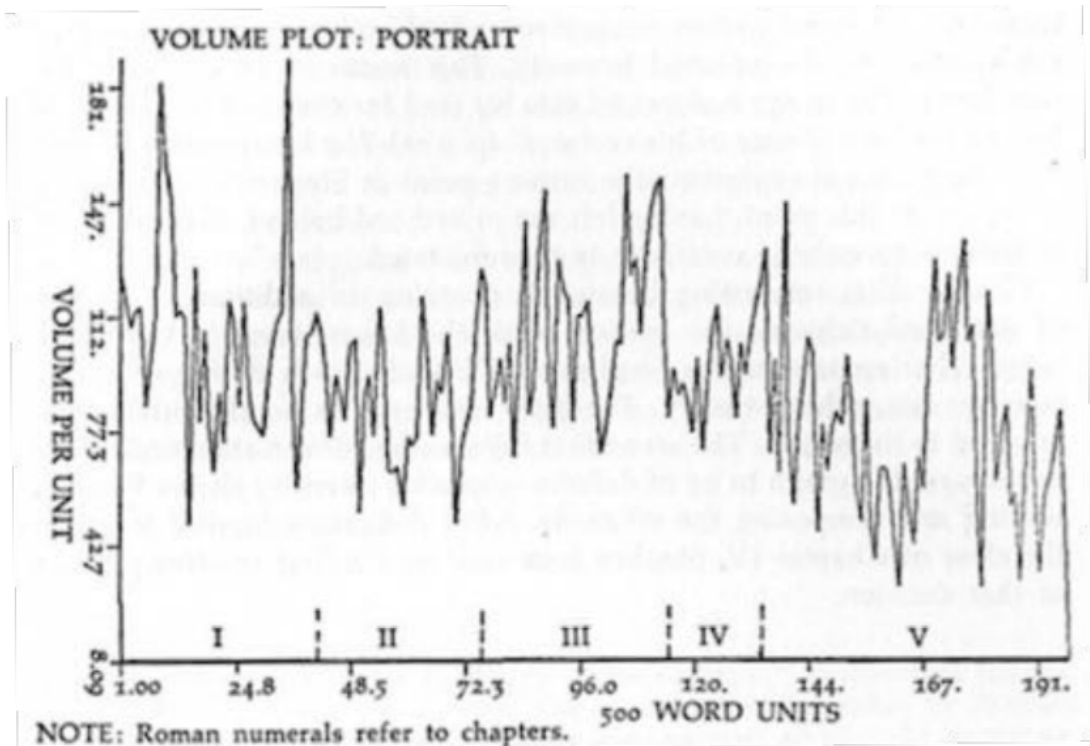


Figure 3. Smith's Visualization

How did Smith get his visualization (Figure 3) and does it do what he says it does? We set out to try to replicate this visualization as a way of understanding Smith and early experiments in textual visualization. This is part of a larger project which, like *Hermeneutica*, is a hybrid combination of book [Rockwell and Sinclair 2016] and tool (<https://voyant-tools.org>). Drawing on this research the following section will:

- Talk about Smith's visualization and how it was brought back to life,
- Show some examples of experiments in revivication of techniques, and
- Reflect on what the practice might be doing and what we might call it.

How did Smith generate the visualization of Joycean imagery? What Smith did, according to the article, was to develop a custom dictionary of some 1,300 terms with emotional valence to track and visualize image intensity or volume through the novel. He divided the text into 500-word chunks, approximately the number of words on a page, as he puts it, and he then counted the number of imagery words from his dictionary in each chunk, weighting some, and graphed the results.

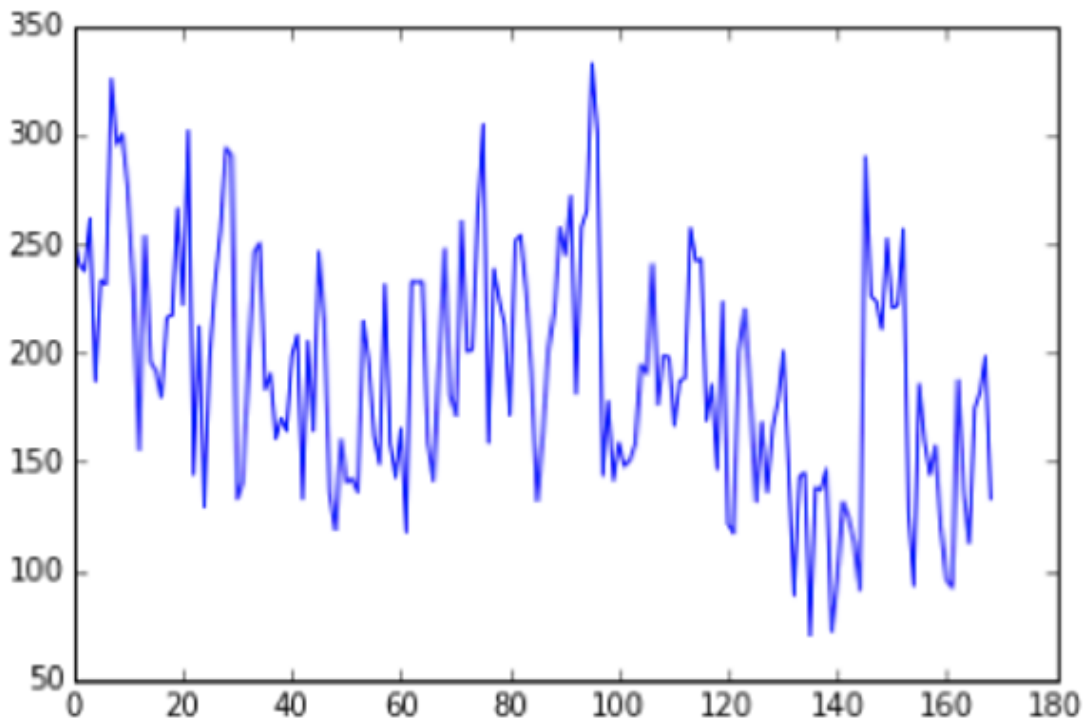


Figure 4. Replication of Smith's Visualization

When we tried to reproduce his results, Figure 4 is what we got. Not quite the graph that Smith got, but similar. We developed our revived interpretation of Smith's technique, or "Zombie tool," in a Jupyter Python notebook using a version of the text from Gutenberg. The first difficulty we ran into recapitulating Smith was reconstituting his dictionary of words with emotional valence. Fortunately, Smith is still alive and he pointed us to an appendix to his book on Joyce titled *Imagery and the Mind of Stephen Dedalus* [Smith 1980] which we OCR'd and corrected in order to recreate his method.

#	IMAGE	FREQ: TOTAL	CH1	CH2	CH3	CH4	CH5
1	ABBEY	3	3	0	0	0	0
2	ABBOTS	2	2	0	0	0	0
3	ABLAZE	2	0	0	0	0	2
4	ABORO	1	0	0	0	1	0
5	ABYSS	5	0	1	4	0	0
6	ACCENT	10	0	3	1	2	4
7	ACHE	8	0	0	6	1	1
8	ACOLYTE	1	0	0	0	1	0
9	ADAM	6	0	0	6	0	0
10	AFLANE	2	0	0	0	2	0
11	AFRAID	17	7	6	1	0	3
12	AFTERGLOW	2	0	0	0	0	2
13	AGLOW	1	0	0	0	1	0
14	AGONY	11	1	1	9	0	0
15	AIR	87	25	10	11	18	23
16	ALICE	2	0	0	2	0	0

Figure 5. Part of Table of Imagery Words from [Smith 1980]

Smith also explained in email correspondence that for his graph he didn't use raw counts per segment but instead weighted each count by the logarithm of the frequency of the word in the entire text – this serves to add more importance to higher frequency terms.

Alas, our updated graph,^[26] despite having access to the original data beyond the published paper, still doesn't really match the original results, which raises questions about the significance and purpose of such replications. Does it call into question Smith's interpretation of Joyce, or his model, or his method of implementing the model, or is our replication at fault? Did we miss something?

57

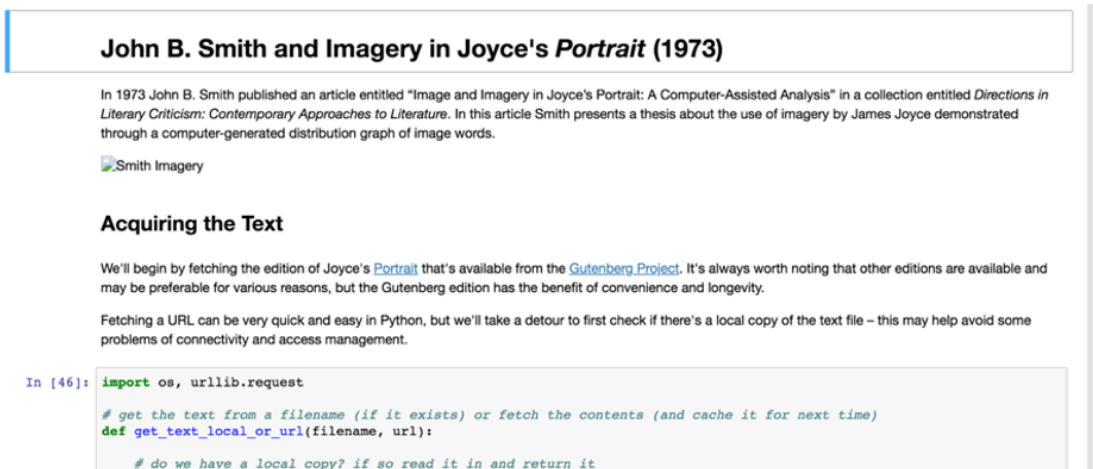


Figure 6. The Jupyter Notebook

One of the issues that arises when you try bring “historical” research and associated tools back to life is the question of how accurately to redevelop the tools used in the replication. Does one need to use exactly the same tools as in the original? In this case, should we be working on a mainframe using something close to the analytical environment used by Smith? In Figure 6 you can see a screenshot of the Jupyter Python notebook, which is the environment we decided to use for this revivification. We chose the notebook model as it encourages a programming style where you explain what you are doing in the spirit of replication rather than emulating the original environment. Ironically, that means our code is by definition different from Smith's original because it is implemented in a deliberately reflective literary programming environment.^[27]

58

One might wonder why we wanted to revivify Smith's work? Why bring back work that is largely forgotten in the digital humanities community?

59

One reason to recover Smith was that he was one of the first to reflect on visualization and computer criticism. Despite being forgotten, his visualization is interesting because it is one of the first text visualizations published not as part of a technical document, but in a paper addressed to literary critics. Smith was a pioneer in visualization and criticism, as a later 1978 paper in the journal *Style* titled “Computer Criticism” [Smith 1978] showed. As such you could say that our project was one of media archaeology – recovering a mediating technology (visualization of emotional imagery) that has recently seen a revival in various forms, including sentiment analysis.

60

A second reason was to revisit some of the methods that Smith pioneered and the ideas they bore. The problem with tools is that like all Zombies they don't last that long without human reuse and no one pays them much attention when they are working. Who remembers ARRAS [Smith 1984], arguably one of the most influential early tools in the history of humanities computing?^[28]

61

Part of the reason instruments have largely escaped the notice of scholars and others interested in our modern techno-scientific culture is language, or rather its lack. Instruments are developed and used in a context where mathematical, scientific, and ordinary language is neither the exclusive vehicle of communication nor, in many cases, the primary vehicle of communication. Instruments are crafted artifacts, and visual and tactile thinking and communication are central to their development and use. [Baird 2004, XV]

Following [Baird 2004] we believe that tools can and do bear knowledge, including theories of interpretation, just as

discourses do, but they bear them differently. This raises the question of how one interprets tools if they can bear knowledge. Obviously one approach is to try to use them as they were used at the time, but how does one use a dated tool when you don't have the code or the platform to run it on? In other words, *how does one bring something back to life so one can see how it worked in bearing knowledge?* 62

Our answer is: interpretative replications, or what we playfully call “zombie tools.” The idea is to replicate the interpretative ideas and methods rather than the particular materiality of the tool. The advantage of the notebook style of programming for replications is that they bring the interpretation of the replication to the fore, allowing one to create a type of Frankenstein's monster, a hybrid of explanation and code that mutually preserve knowledge. In other words, a zombie or revenant created through reanimation of the ideas. 63

What was interesting to us about Smith's work was how he uses text analysis and visualization to model a theory of interpretation drawn from the novel he is interpreting. This allowed him to show how he believes Joyce applied to his own writing the aesthetic theory of artistic imagery presented explicitly by Daedalus, the lead character in Chapter V of the *Portrait* [Joyce 1968]. In other words, Smith translated his interpretative thesis into a software tool or model that could then be evaluated by a computer. His thesis was an interpretation of a widely discussed aesthetic theory drawn from the novel itself. He applied the tool back onto the novel as a way of evaluating his model and ultimately testing whether Joyce actually followed his own aesthetic theory. 64

For our purposes, the computational model of Joyce's imagery and whether it translates the hermeneutical thesis is not that important. What we propose is interpretative replication as a way of bringing ideas and techniques back to life and that this could be important to computer criticism and stylistics. To that end we want to make a number of points about all the “re-”words that we can use: Revivification, Recapitulation, Recapture, Recovery, Replication, Reflection, Reproduction, Relive, Respond, Reinterpretation, Revenant ... 65

First of all, what this project is not trying to do is scientific replication. As [Collins 1975] notes in *The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics*, there is a lot of enculturation going on in what gets called “scientific reproduction.” Replication, in Collins' account, is not straightforward, as almost no complex experiment can be described in sufficient detail to be replicated precisely without having to make assumptions, as we did in the Smith case. Collins goes further to suggest that what is often needed is the *tacit knowledge* that only shared training, close communication or shared researchers moving between labs can provide. Thus, enculturation is the combination of training and negotiation that leads to experiments being considered for replication in the first place and then ensures that the experiments are replicated in a fashion that extends the field. In short, sociologists of science point out how replication is a practice that in each field is constructed by the field. The question for us in the digital humanities then is how do we want to construct replication? *What culture of replication do we want to develop in the humanities? What traditions can we draw on?* 66

One possible tradition is what gets called “experimental archaeology” [Ascher 1961] [Millson 2010]. While this set of practices has its problems, we can learn from the way they are defining the replication in their field. For example, in [Outram 2008]'s “Introduction to experimental archaeology” he talks about some of the types of problems with reproductions that are not sufficiently scientific for what they want to define as truly experimental, in the sense of replicable experiments: 67

1. Lack of clear aims
2. Insufficient detail on materials and methods
3. Compromises over authentic materials
4. Inappropriate parameters
5. Lack of academic context

We assume that generally for (digital) humanities scholars, the ultimate goal of research is not proof, or prediction, but understanding.^[29] When applied to interpretative replication, our aim is therefore not, for example, to use only authentic materials, but to learn about the potential today of past methods. We agree that aims should be clear, but they do not have to be to prove that Smith was right or wrong, especially given that we do not have all the materials and the exact

weights for words. Instead, in interpretative replication and its failures, we learn about ourselves: what we know and do not know. We unpack our assumptions and imagine what we could do rather than learn only about Smith.

An alternative practice that does not aim for scientific credibility is what the artist Judith Buchanan [Buchanan n.d.] calls “revivification”. What she and her performers do is to bring to present life and understanding silent films of Shakespeare plays. She does this through a combination of lecture, commentary and live performance while the film plays. She calls it “revivification” because what they do is collaborate with the dead performers (of the silent films) to create renewed knowledge. When you attend a performance you find that she lectures, live actors perform certain lines, and the original silent films play out on the screen. This is replication enriched by her interpretation and that of the actors with the goal of understanding the phenomena.

68

Her theorization is appealing as it acknowledges that any replication is an interpretation of the replicated. We in the humanities are in a collaboration with the dead. We engage with them rather than just stand on their shoulders. The humanities have a different relationship with our histories than the sciences. The hermeneutic circularity of Smith, who draws theory from a novel in order to interpret the novel, is appropriate to the humanities in a way it would not be to most sciences. There is a similar hermeneutic circularity to interpretative replication. It is both about understanding ourselves and the past. It is a thinking-through of the past.

69

For all these reasons we see replication in the digital humanities as a practice of replying, and re-doing in a larger dialogue, but also re-plying in the sense of folding back onto our past. This is appropriate in the humanities where we eat our past to stay current.

70

4. Domain Specificity. The Example of Plotting Poetry

In this section, we shift our focus from the perspective of tools, methods and replication towards a systematic area of application, poetry. The choice of that area is arbitrary. However, the situation of poetry is somewhat specific. Versification is intrinsically related to various formal measures, and was a subject of quantitative analyses long before any computational methods were available for counting meters, feet, syllables, rhymes and such. However, readily available tools developed for the computational study of literary texts – though they can sometimes be applied to poetry – are often not a perfect fit for versified texts, due to the intrinsic specificities of the highly regulated material that is verse. As for the poetry-specific devices being produced, their transfer from one text to another is made more difficult by language and time-period differences as well as by individual variations in the use of rules between poets. Furthermore, the general challenge – not restricted to poetry – posed by the utter diversity of hermeneutic goals, is made somewhat more acute by the relatively smaller number of scholars working on poetry.

71

Because poetry is so fundamentally linked to a number of forms, and because the variation and creativity in the form are so essential to the production of meaning in poems, the use of quantitative methods for the study of poetry is first geared towards versification. Meters, rhymes, caesuras, stresses can all be very aptly studied with the use of computational methods. Progressively, the fastidious manual collection of data is being replaced as much as possible with automatic data collections. Computational tools are being produced to detect or predict metric syllables, rhyming phonemes, etc., but this effort is somewhat disseminated and has yet to produce universally performant and flexibly adaptable tools. Indeed, the rules of versification form a vast and variable series of systems.

72

The very principles of meter vary from one language to the other, with tonic, syllabic, syllabo-tonic and other systems requiring very different rules to be detected. Furthermore, even within one language area, different eras – and different poets – have produced a range of practices within versification systems. Such systems are often quite strict, and have a high interpretative value, yet remain singular to a language, time period, sub-genre, or to a specific form (the sonnet, for instance, has its own set – or sets – of strict rules). These telling systems being so worthy of a close inspection, and of a systematic one, researchers are actively producing their own tools to automatically collect precise data to quantify verse features. There is, obviously, no obligation to focus solely on genre-specific features of poetic texts, and tools that are useful in analyzing prose texts may be simply transposed to the study of poetry. Yet in doing so, one has two obstacles to overcome: the resistance of the material itself, and the hermeneutic relevance of the endeavor.

73

There is, obviously, no obligation to focus solely on genre-specific features of poetic texts, and tools that are useful in analyzing prose texts may be simply transposed to the study of poetry. Yet in doing so, one has two obstacles to overcome: the resistance of the material itself, and the hermeneutic relevance of the endeavor. 74

First, the use of language in poetic texts is particularly prone to anomalies, constrained as it is by both the corset of verse and the effort towards an economy of words and depth of meaning. Poetry often relies on the ambiguity, the polysemy, the shortcomings of language, to produce a richer layering of less obvious meanings. Automatic taggers of any kind struggle with this linguistic prolificity, such as the lack of a clear and linear syntax, higher degree of polysemy, redundancy, or words being used with unexpected or isolated meanings. POS-taggers, for instance, wrestle with the relatively frequent misuse or category change of words in poetry, as well as with the unusual word order, which is unfortunate as POS-categories are very useful to the study of poetry and to the understanding of versification. Similarly, topic modeling and other clustering methods suffer from the lack of linearity of a versified text, where the verse unit as a structural element plays an important role in the construction of meaning, alongside the syntactic units and the physical proximity, with which it may agree or disagree. 75

Second, stylistic analyses designed for prose might not be as relevant, on their own, once applied to poetry. In focusing on relatively less foregrounded structures, they risk missing the point of a text by failing to address its genre-specific and language-specific form: its versification, its controlled polysemy, the significations of its rhymes, enjambments, choice of meter, degree of obedience to the rule. The latter issue, although not specific to digital humanities, is at the core of how digital humanities and the study of poetry interact, for a focus on versification is bound to produce large amounts of data. Versification descriptions typically include fine information not just about each poem, stanza, rhyme, line, hemistich and syllable group, but also numerous information about each syllable and about its parts. This need for a high granularity has long prompted versification scholars to try and mechanize the handling, if not the collection of such data. 76

There are, thus, two broad ways to explore poetic corpora with the help of computational methods. One is to focus on features shared with prose, and for this, one can borrow any of the more generic tools of digital literary studies, such as TXM in France (see above), or CATMA (<https://catma.de>). Researchers may use parts of speech distribution, word frequencies, topic modeling, word vectors, even sentiment analysis, to gather data of ever-improving quality, although not necessarily as precise, and thus as meaningful and interpretable as data collected by learned humans. Machine Learning, deep learning, neural networks are also used, producing convincing results in authorship or genre attribution for instance, whilst so far, the black box effect of such endeavors bars the traditional untangling of how poetic devices function. In using these tools, not designed with poetry in mind, one important challenge for the researcher is to make good use of them, to gather enough hermeneutical benefit, to get insight into new or crucial issues, and to avoid banalities. 77

The other approach, focusing on poetry-specific features, is seldom addressed by readily available tools. The rules of versification are so language-specific that a tool developed for one language might need considerable adaptations to be reused in another. As in other literary genres, it is difficult to predict the many features that poetry scholars might want to model, and stylistic questions geared towards interpretation tend to focus on features so exclusively characteristic that a tool developed by one team to describe one phenomenon, although it may form part of a further exploration taking said phenomenon into account, might not fit the precise needs of another team. 78

Although the very object's lack of universality is an obstacle to the interoperability of devices, there is a genuine need for tools tailored for poetry and usable for a range of research questions. Such tools are progressively being developed, mostly by researchers trying to address their own needs, in particular for the exploration of versification features [Bories et al. 2021] [Plecháč et al. 2021] [Bories et al. 2023]. Some teams procure fully integrated applications, many share Python packages for all to use, and many more researchers – the group *Plotting Poetry* now has 65 members^[30] – develop precise methods to fit their own research goals, without having a neat tool to share with the wider community. To mention but a few, Valérie Beaudouin's *Métromètre* [Beaudouin 2002], Eliane Delente and Richard Renault's *Malherbe* [Delente and Renault 2015] or Benoît Brard and Stéphane Ferrari's work for French [Brard and Ferrari 2015], Klemens Bobenhausen's *Metricalizer* for German [Bobenhausen and Hammerich 2015], Daniele Fusi's *Chiron* for Latin 79

and Ancient Greek [Fusi 2015], more recently the Postdata project's *Anja*, *Skas* and *Disco* for Spanish [Martínez Cantón et al. 2017], Arto Anttila and Ryan Heuser [Anttila and Heuser 2016] for English and Finnish, Petr Plecháč and Robert Kolár's *versologie* tools [Plecháč and Kolár 2017], Igor Pilshchikov and Anatoli Starostinor's [Pilshchikov and Starostin 2015] or David Birnbaum and Elise Thorsen's works [Birnbaum and Thorsen 2015] for Russian, all automatically analyze various components of verse. Further efforts towards interoperability are made, and one must salute the teams who provide well-thought web ontologies focused on versification phenomena, such as the Postdata project.^[31]

Besides the purer automatic detection efforts, where the quest for data is so time and energy consuming that it risks delaying the hermeneutical goals, many computer-aided poetry researchers mix manually collected and automatically collected data, to achieve a compromise, get smaller, high quality datasets, and allow text interpretation based on a mix of distant and close reading, one informing, testing and guiding the other [Bories 2020]. And some scholars simply use data management tools, such as Excel or Matlab, to handle data collected manually, typically versification data to explore the stylistic relevance, the poetics of versification routines, or the evolution of practices, sometimes other types of data, for instance to reflect on the diachronic reception of a sub-genre, or on more contingent reader response.

Another aspect of digital poetry studies to be mentioned is the production of various poetry generators, which evolved from relatively simple literary devices such as Raymond Queneau's *Cent mille milliards de poèmes* [Queneau 1961] to elaborate computer programs such as Pablo Gervas' WASP.^[32] Their development and examination through initiatives such as Thierry Poibeau and Valérie Beaudouin's OUPOCO project ^[33] provides remarkable insights into what constitutes the uniqueness of a poet's voice.

Poetry studies, thus, are still in the process of developing both interoperable tools, and more uniquely tailored methodologies, with a majority of researchers so far piecing together their own methods, mixing a variety of programming, data analysis and visualization techniques with manual data collections, with or without a view to interpretation. In this observation, we wish to stress that the journey towards relevant quantitative stylistics research is not always linear, nor should it be; the devices developed should continually feed a methodological reflection and reuse, and their applications should be tested and renewed, so as to improve hermeneutical benefit. The group Plotting Poetry, founded by Anne-Sophie Bories and now part of the SIG-DLS, gathers scholars of very diverse geographical, linguistic and indeed methodological realms around a yearly conference.^[34] This platform for the sharing of practices, challenges, and results, is fostering collaborations, and through those, the emergence of well-needed interoperable tools.

5. The Digital Literary Stylistics-Tool Inventory (DLS-TI)

The above has shown that CLS is defining itself in many facets, including chief methodological areas such as textometry, stylometry, and semantic text mining, but also in very specific applications, for example to the genre of poetry, and with regard to its own range of types of replication. We have shown that this definition draws to a large extent from the tools and methods applied in concrete research. Systematically assessing the usage of tools is not only a way of taking stock and boosting information, but also to foster replication/re-analysis and comparability. Moreover, and crucially for our analysis, it provides a way to analyze how different tools embody different hermeneutical approaches and address different methodological needs, how one and the same tool can be used by different communities with different goals, how one community may resort to various tools to target different stylistic phenomena.

The idea of gathering information on tool usage from the community is not new, and has been inspired by pre-existing initiatives. An example that seems particularly interesting is the LRE-map.^[35] Launched at LREC 2010 and soon extended to other conferences, it has been collecting data on the use of Language Resources in submitted papers [Calzolari et al. 2012]. Opposite to traditional catalogues, the LRE-map typically has several entries for the same resource, corresponding to the different papers in which their use is described. Opposite to other catalogues of language resources, where entries are compiled by creators and thus mostly represent their point of view on resource and its intended uses, the LRE map has developed over time to represent rather the user's point of view i.e. the way in which it was applied in actual research. A lexical resource such as WordNet, for instance, is often cited as an ontology in the LRE map, for this is the way it is often applied in NLP today. Also, the diachronic span of the LRE map allows for

the detection of trends in the use of resources over time. In part, the success of the LRE-map is due to the decision to work with a very limited set of metadata, and to the bottom-up collection of data in conjunction with large events.

Within the field of DH various notable comparable initiatives can be named. We cite among others the DIRT directory,^[36] a comprehensive catalogue of digital tools, albeit without links to use cases; the Catalogue of Digital Editions,^[37] an interesting example of community driven collection of metadata for a particular type of resource; the review of Tools and Environments for Digital Scholarly Editing^[38] is currently ongoing by the German Institute for Documentology and Scholarly Editing (IDE); the ACDH Tool Gallery, linked to a series of training workshops, as well as the recently launched TAPoR database of tools,^[39] drawn from the ADHO-proceedings [Barbot et al. 2019].^[40]

The idea of a Tool Inventory curated by the SIG-DLS (DLS-TI)^[41] has developed out of such initiatives, as well as from the epistemological reflections outlined in the introduction to this paper, as a first attempt to gather information on tool usage in CLS, in a way that would do justice to all different approaches and perspectives currently existing in the field, and the following paragraphs offer an overview of this. Accordingly, the definition of “tool” adopted sees them as incarnations of methods. Tools, just like other types of resources such as corpora, lexicons, grammars, etc. are not neutral, but often the reification of theoretical a priori.

From the practical point of view our initiative draws from the aforementioned previous initiatives in that:

- it is based on bottom up contributions of descriptions by researchers;
- it is intended to be use-case-oriented; the idea is not just to collect lists of tools but concrete real-world usage; crucially we aim to collect contributions which do not come from the developers of the tool;
- it relies on a simple method of collection and a limited set of descriptors, aimed at identifying the name and type of tool, including with reference to the TaDiRAH ontology, and the type of use in real DH scenarios such as published research, such as papers, books, blog posts, but also projects, or academic courses.

One line in the inventory represents a use case contribution. The same tool may thus be entered several times, by different contributors or even by one and the same contributor providing different use cases. As the aim is to review existing practices in the community, the focus is on off the shelf tools which are applied to diverse aims.

We decided to accept a broad but text-analysis oriented definition of “tool”, defined as any method or technique that provides ways of data manipulation and interpretation, including desktop GUIs, online Virtual Research environment and libraries for R or Python (see above). In the inventory, we also include general-purpose tools such as Excel spreadsheets when they are used for central CLS tasks such as data manipulation, but exclude, for instance, data acquisition libraries such as lxml for Python used to scrape xml. Visualization tools are also taken into account, as this task is a central one for the interpretation of data.

In 2018 the first call was launched to populate the inventory. To date, 35 use cases have been collected. The results are still limited, but some first interesting trends can be already identified, such as the importance of R packages and Python libraries for the community. In turn, this is linked to the fact that many of the tools that have been entered are not specifically designed for literary or stylistic analysis, but are implementations of Natural Language Processing algorithms or statistics measures. On the other hand of the spectrum, a number of entries refer to desktop applications (TXM, WCopyFind, Docuscope, ...) which are used for very specific purposes and seem to be more specifically geared towards CLS.

Such preliminary results mirror our intuition about the current practices in the community. On the one hand, there is a tendency towards the development of highly specialized methods based on generic building blocks. On the other hand, there is simultaneously the need for dedicated tools, with a well-identified and -tested set of functions, which address both usability and repeatability needs.

As we hope to gather an increasing number of entries from future campaigns, we believe that the DLS-TI may become an important instrument to monitor the needs of the community, which in turn may benefit those institutions and infrastructures – such as CLARIN^[42] and DARIAH ERIC and their national consortia - that are currently devoted to

creating resources for the DH community. Through its perspective on use cases, it also allows scrutinizing methods epistemologically, in relation to research questions and theoretical frameworks. Relatedly, as a systematic inventory not only of tools, but also of case studies, the DLS-TI can function as a basis for replication and reanalysis - and thus help progress in the field. Finally, initiatives such as the DLS-TI could be beneficial in identifying the needs of the community, and in deciding where best to invest time and money be it in terms of development of user-friendly interfaces, or in the organization of training and events that promote the exchange of best practices.

6. Conclusions. The many facets of tool criticism

This intervention began as a series of “anatomies” of tools – a playful concept applied to different approaches, which, we hope, documents relevant vantage points within CLS. In addition to the predictable “quantitative” ones, *validity* and *replicability*, we have offered *hermeneutic plausibility* [Winko 2015] and *open-ended tinkering* as among the ultimate causes or mindsets of computational literary studies. Having said this, *validity*, *reliability*, and *objectivity* of explanatory hypothesis testing do resonate with much of what we believe computational literary studies are and should be about. This apparent conflict in fact appears indicative of not only the diversity of approaches within the emerging field, but also of that of research stages. What seems crucial though, is to maintain an awareness of one's aims and axioms within this rich conversation – which necessarily involves a sense of (computational) method.

93

Rather than “just applying” tools prepared by some pioneering expert, CLS scholars have become more interested and knowledgeable about methods and their fit to research questions and data. Future cohorts of scholars may take tested and tried methods for granted, and most of them may concentrate on testing and exploring research questions, rather than on method development (as today is still so much needed in the case of Topic Modeling, for instance). When computational methods have become reliable and transparent vehicles for most, this means that fewer scholars need to concentrate on honing them. A matter of division of tasks is only functional, however, where a critical level of validation and practical experience has been reached and is embedded as continuous practice. As we have shown, humanistic interactions, not only with artefacts and phenomena of the past, but also with historical tools and methods, form an important source of insight.

94

We may or may not have already entered a phase of the digital transformation of literary studies in which digital methodology may be reintegrated with the field and its various sub-fields. In any case, further maturation is likely to produce a new type of division of labor, where continuous development, calibration and reflection of tools is but one, albeit focal, activity.

95

Taking the argument of “tool defense” voiced in the introduction one step further, “the instrumentalist” dimension of algorithmic data science [Jones 2018] is worthy of some positive attention: using “an instrument” to further sophisticated literary modeling, for example machine learning for predictive modeling, is as warranted as a “truth-oriented,” or explanatory, type of modeling, whose goal is to develop a mechanistic model of the process that originally produced the observed data and which strives to estimate the parameters of this process. Here, of course, critical enquiry into algorithmic logic is a precondition, possibly at a larger scale than “just CLS.”

96

A few other topics: Annotation is a scholarly universal [Unsworth 2000], and one of the externalizations of modeling in DH practices [Herrmann forthcoming]. Enrichment of raw texts by annotation [Rapp 2017] is one of the key practices where scholarly knowledge and theory are made explicit. This includes hermeneutic, progressively altered annotations that depict a rather inductive approach [Gius and Jacke 2017] [Percillier 2017] and rule-based annotation systems such as MIPVU [Herrmann et al. 2019] [Steen et al. 2010], as well as the pertinent issue of inter-coder reliability [Kuhn 2019].

97

Another important aspect is modeling through metadata, as highlighted for example in the TXM study. Metadata reside at the interface of data-scientific and philological dimensions of literature studies, as they allow adding multiple variables to the data model. Together with text sampling and markup, metadata constitute the philological object, ideally in an explicitly theory-driven way. In research practice, this also involves important decisions about schemata, but also about missing or incorrect entries.

98

A crucial dimension of CLS that was addressed in our discussion is *source criticism*. Building corpora is no trivial task,

99

as they are where the research question, but also general and specific assumptions about literary discourse are modeled both conceptually and as data [Bode 2018] [Herrmann and Lauer 2018] [Herrmann forthcoming]. We need source criticism and reflection about what it means that literature is represented either digitally or in an analog medium, we need to understand what we sample, and how we sample. We need to be aware of bias, but also know that “doing CLS,” defining research questions and operationalizations, necessarily involves “highlighting” some aspects and therefore “hiding” others, in an interpretation of Popper. On that note, the discussion about hermeneutic vs. algorithmic criticism has become less irreconcilable since attention has shifted to the role of modeling, which has highlighted the logical relation between theory, data, and method [Flanders and Jannidis 2019]. The attention to modeling emphasizes the need for scholarly control and critique of all pertinent levels. With [Popper 2002, p. 67], we think it is a constructive vision of computational literary studies to say that it “passes on its theories; but it also passes on a critical attitude towards them. The theories are passed on, not as dogmas, but rather with the challenge to discuss them and improve upon them”. This approach at the level of modeling, we think, does not preclude an “emphatic” way of approaching our objects at certain stages of research.

We have shown how the question of digital methodologies, epistemic practices, and research motives cuts directly to the identity of “computational (or digital) literary studies”: In fact, CLS's putative emphasis on tools / methodology may actually indicate that digital literary enquiry has not been as fundamentally detached from “traditional” studies – literary studies, stylistics, cultural studies etc. – as the administrative and institutional perspective may suggest. For ontological and other definitions of digital humanities, see for example [Svensson 2015] and contributions in [Terras et al. 2013].^[43] Rather, CLS emerges as one of the incarnations of literary studies, capitalizing on the affordances of the digital – and simultaneously being shaped by them. Admittedly, whether this may allow for eventually more freedom, on a grander scale, depends on the willingness of established Humanities disciplines and gatekeepers to venture for an update. Provided some more time, a fundamental change may just be at our doorstep, with literary data-driven modeling pursued not at the sidelines of the field, but complementing the practices at its center.

100

Notes

[1] In the past few years, the term *Computational Literary Studies* (CLS) has become more prominent than its alternatives, including *Digital Literary Stylistics* (DLS). While DLS focuses more on aspects of style (stylometry, corpus stylistics), we decided to use the term CLS in the current paper, except when referring to resources that are explicitly named “DLS,” such as the ADHO-Special Interest Group “Digital Literary Stylistics” (SIG-DLS), and the DLS tool inventory.

[2] SIG DLS organized a pre-conference workshop at DH2019 in Utrecht on 9 July 2019, see <https://dls.hypotheses.org/activities/anatomy>

[3] The term *tool* is heavily underspecified in digital humanities. For example, the glossary of ForText, a popular DH-training resource in the German-speaking context, does not incorporate a lemma for “tool” <https://fortext.net/ressourcen/glossar> (retrieved 13 October 2022). In the following, in absence of a comprehensive definition to draw from, we provide a largely operational one.

[4] We use the term *literary data* to include a wide array of data. CLS presently centers mostly on textual data, and our case studies do the same. However, a growing body of CLS research is dedicated to studying multi-modal, especially audio and visual data. For a recent overview see [Herrmann et al. 2021].

[5] Our predicate “operates upon” includes for example visualization, analysis, annotation, curation, comparison cf. [Unsworth 2000]. In his “Scholarly Primitives” paper, John Unsworth gives an indirect definition in the table at the bottom of the document in which he transfers the human genome project into a “human genre project:” “2.Tools: What protocols and tools for data submission, viewing, analysis, annotation, curation, comparison, and manipulation will you need to make maximal use of the data? What sorts of links among datasets will be useful?”.

[6] Computational methods and resources are *change agents* in Weizenbaum's sense: they alter literary scholars' practices and the modeling of objects and theories. For example, Herrmann (forthcoming) shows how corpora are externalizations of the philological object in practical and

theoretical senses, and thus are key arenas of this change. A digital literary corpus is different both from a collection of books and a 'mere digital database', because it has a more deliberate modeling function: It is built to construct "narratives" [Hayles 2012, 176].

[7] <https://culturalanalytics.org/2019/09/special-forum-on-responses-to-nan-z-da/>

[8] Post-hoc analyses of Nan Z. Da's polemic paper and the ensuing debate have revealed that instead of a methodological paper that it appeared to be at the surface, it rather worked as an disciplinary in- and outgroup profiling.

[9] A good candidate for the most fundamental academic practice is comparison [Descartes 1959], which, externalized by computational methods, presently is evolving in new ways (see also the CRC1288 "Practices of Comparing. Ordering and Changing the World" <https://www.uni-bielefeld.de/sfb/sfb1288/>).

[10] Reproducibility / replicability is one of the core principles of scientific progress. Direct replication is the "attempt to recreate the conditions believed sufficient for obtaining a previously observed finding and is the means of establishing reproducibility of a finding with new data" [Open Science Collaboration 2015, aac4716–2]. The so-called *replication crisis* has affected many empirical fields such as medicine and psychology. While one way of addressing it has focused on scientific malpractice, for social constructivists in the sense of [Berger and Luckmann 1967] the crisis highlights precisely the scholarly practices that are constructed by dynamics of the field.

[11] <http://textometrie.ens-lyon.fr/spip.php?rubrique80>

[12] <http://ancilla.unice.fr/>

[13] TXM is in constant development and its team is taking constructive feedback to implement and complete its statistical studies.

[14] It is therefore a composition of aborted or simply abandoned former projects that literary specialists often name "cycles" (the "Stavelot cycle", the "Annie cycle", see [Debon 2008] [Décaudin 1969]).

[15] The supposed dates of writing are based on analyses of the manuscripts and the first states of publication of the poems of Apollinaire initiated with great rigor and much erudition by M. Décaudin in 1969 [Décaudin 1969], as well as in his edition of the text for the Bibliothèque de la Pléiade [Apollinaire 1994]. It is completed by the analysis of his notebooks and his correspondence kept at the Fonds Jacques Doucet (<http://www.bljd.sorbonne.fr/index.php>), which make it possible to give certain texts a quite precise drafting interval.

[16] This section took its start as a short abstract titled *Less than countless. Options to move beyond word counting in stylometry* originally provided by Mike Kestemont for the workshop. The original abstract was substantially extended and revised by Simone Reborá and the co-authors.

[17] One exception, though, might be that of "mixed methods" [Herrmann 2017], which combine wordcount with approaches such as keyness analysis to identify the linguistic phenomena that should be investigated more closely. Such methods have also been employed successfully in authorship attribution studies (cf. [Reborá et al. 2019]) and can be considered as an effort to build a bridge between raw statistics and the hermeneutic and aesthetic dimensions of literary studies. However, it should be noted that they as a rule build upon simple wordcounts (e.g., keyness analysis is generally based on raw word frequencies).

[18] <https://pan.webis.de/>

[19] <https://pan.webis.de/clef12/pan12-web/author-identification.html>

[20] <https://spacy.io/>

[21] e.g. <http://mallet.cs.umass.edu/> by [McCallum 2002]; or <https://dariah-de.github.io/TopicsExplorer/> by [Simmler et al. 2019].

[22] This problem extends to other forms of semantic text mining. For sentiment analysis, see [Kiefer 2018].

[23] [Horstmann 2018] <https://fortext.net/routinen/methoden/topic-modeling>

[24] For example, choosing a higher number of topics can improve classification scores while producing less coherent topics in terms of PMI at the same time (Keli Du, personal communication).

[25] A pandybat is a stiff leather strap with a handle that was used for punishing students in Ireland.

[26] See <https://github.com/sgsinclair/epistemologica/blob/master/Smith-Imagery.ipynb>

[27] As part of the larger project we are developing an alternative to Jupyter called Spyral as in “spiral notebook.” It is notebook environment that is also an extension of Voyant (<https://voyant-tools.org>) so you can call Voyant tools within notebooks. It is based on JavaScript rather than Python because that is what a lot of the Voyant interface is developed in. We hope it will give users of Voyant a way to extend their explorations. Spyral is available at <https://voyant-tools.org/spyral/>.

[28] ARRAS was one of the first tools to be designed to be used for interactive exploration of a text rather than in a batch mode. It influenced subsequent tools like TACT and eventually Voyant.

[29] “Humanistic understanding” may very well encompass phases of hypothesis testing and predictive methods (machine learning algorithms). Eventually, however, these steps serve to enlighten us about the conditions and effects of people’s interaction with culture, meaning, and, generally, life.

[30] <https://www.plottingpoetry.org>

[31] <https://github.com/linhd-postdata/Network-of-ontologies>

[32] <http://nil.fdi.ucm.es/index.php?q=node/206>

[33] <http://www.transfers.ens.fr/l-oupoco-l-ouvroir-de-poesie-combinatoire>

[34] <https://www.plottingpoetry.org>

[35] <http://lremap.elra.info/>

[36] <http://dirtdirectory.org/> (offline at the time of writing, last accessed 2 July 2023)

[37] <https://dig-ed-cat.acdh.oeaw.ac.at/>

[38] <https://www.i-d-e.de/cfr-tools/>

[39] <http://tapor.ca/home>

[40] <https://www.oeaw.ac.at/acdh/events/event-series/acdh-tool-gallery-52/>

[41] <https://dls.hypotheses.org/774>

[42] The DLS-TI could also be a useful addition to the CLARIN Resource Families initiative, see <https://www.clarin.eu/resource-families>

[43] See also initiatives such as OpenMethods, which curate “descriptions of methods and tools, tool and methods critique, as well as practical and theoretical reflections about how and why humanities research is conducted digital and how the increasing influence of digital methods and tools changes scholarly attitudes and scientific practices of humanities research” <https://openmethods.dariah.eu/>

Works Cited

- Adam and Viprey 2009** Adam, J.-M. and Viprey, J.-M. “Corpus de textes, textes en corpus. Problématique et présentation”. *Corpus* 8:5–25.
- Anttila and Heuser 2016** Anttila, A. and Heuser, R. “Phonological and Metrical Variation across Genres”. *Proceedings of the Annual Meetings on Phonology. Linguistic Society of America*, Washington, DC (2016).
- Apollinaire 1994** Apollinaire, G. *Œuvres Poétiques*, M. Adéma et M. Décaudin eds., Bibliothèque de la Pléiade, Gallimard, Paris, France (1994).
- Apollinaire 2005** Apollinaire, G. *Lettres à Madeleine. Tendre Comme Le Souvenir*, Gallimard, Paris, France (2005).
- Ascher 1961** Ascher, R. “Experimental Archeology”. *American Anthropologist* 63(4):793–816. doi: <https://doi.org/10.1525/aa.1961.63.4.02a00070>
- Bacciu et al. 2019** Bacciu, A., La Morgia, M., Mei, A., Nerio Nemmi, E., Neri, V. and Stefa, J. “Cross-domain Authorship Attribution Combining Instance Based and Profile Based Features”. In: Cappellato, L., Ferro, N., Losada, D. E. and Müller, H. (eds.). *CLEF 2019 Labs and Workshops, Notebook Paper*. (2019).
- Baird 2004** Baird, D. *Thing Knowledge: A Philosophy of Scientific Instruments*, University of California Press, Berkeley (2004).
- Bal 2017** Bal, M. *Narratology: Introduction to the Theory of Narrative*, Fourth Edition., University of Toronto Press, Toronto Buffalo London (2017).
- Barbot et al. 2019** Barbot, L., Fischer, F., Moranville Y. and Pozdniakov, I.: “Which DH Tools Are Actually Used in Research?” In: *weltliteratur.net*, 6 Dec 2019. URL: <https://weltliteratur.net/dh-tools-used-in-research/>
- Beaudouin 2002** Beaudouin, V. *Mètre et Rythme Du Vers Classique*. Corneille et Racine, Honoré Champion, Paris (2002).
- Berger and Luckmann 1967** Berger, P. and Luckmann, T. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Doubleday, Garden City, NY (1967).
- Bhatia et al. 2018** Bhatia, S., Lau, J.H. and Baldwin, T. “Topic Intrusion for Automatic Topic Model Evaluation”. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium (2018), pp. 844–849.

- Binder and Jennings 2014** Binder, J.M. and Jennings, C. "Visibility and meaning in topic models and 18th-century subject indexes". *Literary and Linguistic Computing* 29(3):405–411. doi: <https://doi.org/10.1093/lc/fqu017>
- Birnbaum and Thorsen 2015** Birnbaum, D.J. and Thorsen, E. "Markup and meter: Using XML tools to teach a computer to think about versification". *Proceedings of Balisage: The Markup Conference*. (2015). doi:10.4242/BalisageVol15.Birnbaum01
- Blei 2012** Blei, D.M. "Probabilistic Topic Models". *Communication of the ACM* 55(4):77–84. doi: <https://doi.org/10.1145/2133806.2133826>
- Bobenhausen and Hammerich 2015** Bobenhausen, K. and Hammerich, B. "Métrique littéraire, métrique linguistique et métrique algorithmique de l'allemand mises en jeu dans le programme Metricalizer". *Langages* 199(3):67–88. doi: <https://doi.org/doi:10.3917/lang.199.0067>
- Bode 2018** Bode, K. *A World of Fiction: Digital Collections and the Future of Literary History*, University of Michigan Press., (2018).
- Bories 2020** Bories, A.-S. Des Chiffres et Des Mètres. *La Versification de Raymond Queneau*. Honoré Champion, Paris, France (2020).
- Bories et al. 2021** Bories, A.-S., Purnelle, G. and Marchal, H. *Plotting Poetry: On Mechanically-Enhanced Reading*, Presses Universitaires de Liège, Liège (2021).
- Bories et al. 2023** Bories, A.-S., Plecháč, P. and Ruiz Fabo, P. *Computational Stylistics in Poetry, Prose, and Drama*, De Gruyter (2023).
- Brard and Ferrari 2015** Brard, B. and Ferrari, S. "Des Vers et des Mesures : Détection des Noyaux Vocaliques". *Langages* 199(3):107–124. doi: <https://doi.org/doi:10.3917/lang.199.0107>
- Brunet 1989** Brunet, É. "Hyperbase: Logiciel Documentaire et Statistique pour l'Exploitation des Grands Corpus". *Tools for humanists*, p. 33–36.
- Bubenhofer and Dreesen 2018** Bubenhofer, N. and Dreesen, P. "Linguistik als antifragile Disziplin? Optionen in der digitalen Transformation". *Digital Classics Online*, pp. 63–75. doi: <https://doi.org/10.11588/dco.2017.0.48493>
- Buchanan n.d.** Buchanan, J. "Collaborating With the Dead: Revivifying Frank Benson's Richard III". Booklet published by Silents Now.
- Burrows 2002** Burrows, J. "'Delta:' a Measure of Stylistic Difference and a Guide to Likely Authorship". *Literary and Linguistic Computing* 17(3):267–287. doi: <https://doi.org/10.1093/lc/17.3.267>
- Calzolari et al. 2012** Calzolari, N., Del Gratta, R., Francopoulo, G., Mariani, J., Rubino, F., Russo, I. and Soria, C. "The LRE Map. Harmonising Community Descriptions of Resources". *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation*. Istanbul, Turkey (2012), pp. 1084–1089.
- Carter 2012** Carter, R. "Coda: Some Rubber Bullet Points". *Language and Literature* 21: 106–114. doi: <https://doi.org/10.1177/0963947011432048>
- Collins 1975** Collins, H.M. "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics". *Sociology* 9(2):205–224. doi: <https://doi.org/10.1177/003803857500900202>
- Da 2019** Da, N.Z. "The Computational Case against Computational Literary Studies". *Critical Inquiry* 45(3):601–639. doi: <https://doi.org/10.1086/702594>
- Daelemans et al. 2019** Daelemans, W., Kestemont, M., Manjavacas, E., Potthast, M., Rangel, F., Rosso, P., Specht, G., Stamatatos, E., Stein, B., Tschuggnall, M., Wiegmann, M. and Zangerle, E. "Overview of PAN 2019: Bots and Gender Profiling, Celebrity Profiling, Cross-Domain Authorship Attribution and Style Change Detection". In: Crestani, F., Braschler, M., Savoy, J., Rauber, A., Müller, H., Losada, D. E., Heinatz Bürki, G., Cappellato, L. and Ferro, N. (eds.). *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Springer International Publishing, Cham (2019), pp. 402–416.
- Debon 2008** Debon, C. *'Calligrammes' Dans Tous Ses États - Édition Critique Du Recueil de Guillaume APOLLINAIRE*, éditions Calliopées, Paris (2008).
- Delente and Renault 2015** Delente, É. and Renault, R. "Projet Anamètre : Le Calcul du Mètre des Vers Complexes". *Langages* 199(3):125–148. doi: <https://doi.org/doi:10.3917/lang.199.0125>
- Descartes 1959** Descartes, R. *Règles pour la Direction de l'Esprit* (3rd Ed.; J. Sirven, Ed.). Brin, Paris (1959).

- Drucker 2012** Drucker, J. "Humanistic Theory and Digital Scholarship". In: Gold, M. K. (ed.). *Debates in the Digital Humanities*. University of Minnesota Press, Minneapolis, MN (2012). Retrieved from <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfd1e/section/0b495250-97af-4046-91ff-98b6ea9f83c0>
- Du 2019** Du, K. "A Survey On LDA Topic Modeling In Digital Humanities". *Proceedings of the 2019 Digital Humanities Conference*. Utrecht (2019).
- Décaudin 1969** Décaudin, M. *Le Dossier d'alcools*, Droz, Paris (1969).
- Eder 2015** Eder, M. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem". *Literary and Linguistic Computing* 30(2):167–182. doi: <https://doi.org/10.1093/lilc/fqt066>
- Eder 2017** Eder, M. "Visualization in Stylometry: Cluster Analysis Using Networks". *Digital Scholarship in the Humanities* 32(1):50–64. doi: <https://doi.org/10.1093/lilc/fqw061>
- Eder et al. 2016** Eder, M., Rybicki, J. and Kestemont, M. "Stylometry with R: A Package for Computational Text Analysis". *The R Journal* 8(1):107–121.
- Erlin et al. 2021** Erlin, M., Piper, A. Knox, D., Pentecost, S., Drouillard, M., Powell, B. and Townson, C. "Cultural Capitals: Modeling Minor European Literature". *Journal of Cultural Analytics* 6 (1). <https://doi.org/10.22148/001c.21182>
- Evert et al. 2017** Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C. and Vitt, T. "Understanding and Explaining Delta Measures for Authorship Attribution". *Digital Scholarship in the Humanities* 32(suppl_2):ii4–ii16. doi: <https://doi.org/10.1093/lilc/fqx023>
- Flanders and Jannidis 2019** Flanders, J. and Jannidis, F. *The Shape of Data in Digital Humanities: Modeling Texts and Text-Based Resources*, Routledge, Taylor and Francis Group, London; New York (2019).
- Follet 1987** Follet, L. "Apollinaire Entre Vers et Prose, de 'L'Obituaire' à la 'Maison des Morts'", *Semen* 3, février 1987. doi: <http://semen.revues.org/5523>
- Franzini 2012** Franzini, G. "Catalogue of Digital Editions". doi: <https://doi.org/10.5281/zenodo.1161425>
- Frontini et al. 2017** Frontini, F., Boukhaled, M.-A. and Ganascia, J.-G. "Mining for Characterising Patterns in Literature Using Correspondence Analysis: an Experiment on French Novels". *Digital Humanities Quarterly* 11(2).
- Fusi 2015** Fusi, D. "A Multilanguage, Modular Framework for Metrical Analysis: It Patterns and Theoretical Issues". *Langages* 199(3):41–66. doi: <https://doi.org/doi:10.3917/lang.199.0041>
- Genette 1972** Genette, G. *Figures III*, Éditions du Seuil, Paris (1972).
- Gius and Jacke 2017** Gius, E. and Jacke, J. "The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis". *International Journal of Humanities and Arts Computing* 11(2):233–254. doi: <https://doi.org/10.3366/ijhac.2017.0194>
- Gius et al. 2019** Gius, E., Reiter, N. and Willand, M. "Foreword to the Special Issue 'A Shared Task for the Digital Humanities: Annotating Narrative Levels'". *Journal of Cultural Analytics*. doi: <https://doi.org/10.22148/16.047>
- Guiraud 1953** Guiraud, P. *Index Des Mots d'Alcools de G. Apollinaire (Index Du Vocabulaire Du Symbolisme. 1)*, Klincksieck, Paris (1953).
- Hayles 2012** Hayles, N. K. *How We Think: Digital Media and Contemporary Technogenesis*. University of Chicago Press, Chicago (2012).
- Heiden et al. 2010** Heiden, S., Magué, J.-P. and Pincemin, B. "TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement". *JADT* 2010. (2010), pp. 1021–1032.
- Henny et al. 2018** Henny, U., Betz, K., Schlör, D. and Hotho, A. "Alternative Gattungstheorien. Das Prototypenmodell am Beispiel hispanoamerikanischer Romane". *Proceedings of the DHd 2018 Conference*. Cologne (2018).
- Herrmann 2017** Herrmann, J.B. "In a Test Bed with Kafka. Introducing a Mixed-Method Approach to Digital Stylistics". *Digital Humanities Quarterly* 011(4).
- Herrmann and Lauer 2018** Herrmann, J.B. and Lauer, G. "Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne". *Osnabrücker Beiträge zur Sprachtheorie (OBST)* 92:127–156.
- Herrmann et al. 2019** Herrmann, J.B., Woll, K. and Dorst, A.G. "Linguistic Metaphor Identification in German". *MIPVU in*

- Multiple Languages*. John Benjamins, Amsterdam / Philadelphia (2019).
- Herrmann et al. 2021** Herrmann, J.B., Jacobs, A. and Piper, A. "Computational Stylistics". In D. Kuiken and A. Jacobs (Eds.), *Handbook of Empirical Literary Studies*, pp. 451-486. Berlin: De Gruyter.
- Herrmann forthcoming** Herrmann, J.B. *Externalizations. Data-Driven Literary Studies*.
- Herschberg-Pierrot 2005** Herschberg-Pierrot, A. *Le Style En Mouvement. Littérature et Art, Sup/Lettres*, Berlin (2005).
- Herschberg-Pierrot 2006** Herschberg-Pierrot, A. "Style, Corpus et Genèse". *Corpus* 5.
- Hirst and Feiguina 2007** Hirst, G. and Feiguina, O. "Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts". *Literary and Linguistic Computing* 22(4):405-417. doi: <https://doi.org/10.1093/lc/fqm023>
- Horstmann 2018** Horstmann, J. "Topic Modeling" In: *forTEXT. Literatur digital erforschen*. Available at: <https://fortext.net/routinen/methoden/topic-modeling>[Accessed: 3 December 2019].
- Jacobs 2018** Jacobs, A.M. "The Gutenberg English Poetry Corpus: Exemplary Quantitative Narrative Analyses". *Frontiers in Digital Humanities* 5. doi: <https://doi.org/10.3389/fdigh.2018.00005>
- Jacquot 2012** Jacquot, C. "Le Poulpe, une Figure de la 'Plasticité' d'Apollinaire?", *Apollinaire* 11 (2012), pp. 35-43.
- Jacquot 2014** Jacquot, C. *Plasticité de l'Écriture Poétique d'Apollinaire. Une Articulation du Continu et du Discontinu*. Thèse de doctorat en langue française sous la direction de J. Dürrenmatt, Université Paris IV-Sorbonne (2014).
- Jacquot 2017** Jacquot, C. "Corpus Poétique et Métadonnées: la Problématique de la Datation dans les Poèmes de Guillaume Apollinaire". *Modèles et Nombres En Poésie*. Champion, Paris (2017), pp. 81-103.
- Jenny 2011** Jenny, L. *Le Style En Acte. Vers Une Pragmatique Du Style*, MétisPresses, Genève (2011).
- Jones 2018** Jones, M. L., "How We Became Instrumentalists (Again)". *Historical Studies in the Natural Sciences*, 48(5): 673-684. doi: <https://doi.org/10.1525/hsns.2018.48.5.673>
- Joyce 1968** Joyce, J. *A Portrait of the Artist as a Young Man, Text, Criticism, and Notes*. Edited by Chester G. Anderson. The Viking Press, New York (1968).
- Kestemont et al. 2018** Kestemont, M., Tschuggnall, M., Stamatatos, E., Daelemans, W., Specht, G., Stein, B. and Potthast, M. "Overview of the Author Identification Task at PAN-2018: Cross-Domain Authorship Attribution and Style Change Detection". *Working Notes Papers of the CLEF 2018 Evaluation Labs. CEUR Workshop Proceedings* (2018), pp. 1-25.
- Kiefer 2018** Kiefer, K. "Tool Criticism on emotional text analysis". *Proceedings of the EADH Conference*. (2018).
- Koolen et al. 2018** Koolen, J., van Gorp, M. and van Ossenbruggen, J. "Lessons Learned from a Digital Tool Criticism Workshop". *Proceedings from DH Benelux 2018*. Amsterdam, The Netherlands (2018).
- Kuhn 2019** Kuhn, J. "Computational Text Analysis Within the Humanities: How to Combine Working Practices from the Contributing Fields?". *Language Resources and Evaluation* 53(4): 565-602. doi: <https://doi.org/10.1007/s10579-019-09459-3>
- Lafon 1980** Lafon, P. "Sur la Variabilité de la Fréquence des Formes dans un Corpus". *Mots. Les Langages du Politique* 1(1): 127-165. doi: <https://doi.org/10.3406/mots.1980.1008>
- Lafon and Muller 1984** Lafon, P. and Muller, C. *Dépouillements et Statistiques En Lexicométrie, Travaux de linguistique quantitative*, Slatkine/Champion, Genève/Paris (1984).
- Lau and Baldwin 2016** Lau, J.H. and Baldwin, T. "The Sensitivity of Topic Coherence Evaluation to Topic Cardinality". *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California (2016), pp. 483-487.
- Martínez Cantón et al. 2017** Martínez Cantón, C.I., Ruiz Fabo, P., González-Blanco García, E. and Poibeau, T. "Automatic enjambment detection as a new source of evidence in Spanish versification". *Plotting Poetry: On Mechanically-Enhanced Reading / Machiner La Poésie: Sur Les Lectures Appareillées*. Basel (2017).
- McCallum 2002** McCallum, A.K. *MALLET: A Machine Learning for Language Toolkit*, (2002). <http://mallet.cs.umass.edu>
- McCarthy 2005** McCarthy, W. *Humanities Computing*, Palgrave Macmillan UK (2005).

- Miall 2018** Miall, D.S. "Reader-Response Theory". *A Companion to Literary Theory*. John Wiley and Sons, Ltd (2018), pp. 114–125. doi: <https://doi.org/10.1002/9781118958933.ch9>
- Millson 2010** Millson, D. (ed) *Experimentation and Interpretation: The Use of Experimental Archaeology in the Study of the Past*. Oxbow Books (2010).
- Mitrofanova 2015** Mitrofanova, O. "Probabilistic Topic Modeling of the Russian Text Corpus on Musicology". In: Eismont, P. and Konstantinova, N. (eds.). *Language, Music, and Computing*. Springer International Publishing, Cham (2015), pp. 69–76.
- Moore 1995** Moore, C. *Apollinaire en 1908, la Poétique de l'Enchantement: Une Lecture d'Onirocritique*, Paris, Minard (1995).
- Noble 2018** Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York University Press, New York University.
- Open Science Collaboration 2015** Open Science Collaboration. (2015). "Estimating the reproducibility of psychological science". *Science*, 349(6251), aac4716–aac4716. doi: <https://doi.org/10.1126/science.aac4716>
- Outram 2008** Outram, A.K. "Introduction to Experimental Archaeology". *World Archaeology* 40(1):1–6. doi: <https://doi.org/10.1080/00438240801889456>
- O'Neil 2016** O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1 edition., Crown, New York (2016).
- Pavlova and Fischer 2018** Pavlova, I. and Fischer, F. "Topic Modeling 200 Years of Russian Drama". *Proceedings of the EADH Conference* (2018).
- Percillier 2017** Percillier, M. "Creating and Analyzing Literary Corpora". *Data Analytics in Digital Humanities*. Springer, Cham (2017), pp. 91–118.
- Pilshchikov and Starostin 2015** Pilshchikov, I. and Starostin, A. "Reconnaissance Automatique des Mètres des Vers Russes: Une Approche Statistique sur Corpus". *Langages* 199(3):89–106. doi: <https://doi.org/10.3917/lang.199.0089>
- Pincemin 2011** Pincemin, B. "Analyse Stylistique Différentielle à Base de Marqueurs et Textométrie". In: Garric, N. and Maurel-Indart, H. (eds.). *Vers Une Automatisation de l'analyse Textuelle. texto! Textes and Cultures*, Volume XV - n°4 et XVI - n°1 (2011), pp. 54–61.
- Piotrowski 2019** Piotrowski, M. "Historical Models and Serial Sources". *Journal of European Periodical Studies* 4(1):8–18. doi: <https://doi.org/10.21825/jeps.v4i1.10226>
- Piper 2017** Piper, A. "Think Small: On Literary Modeling". *PMLA* 132(3):651–658. doi: <https://doi.org/10.1632/pmla.2017.132.3.651>
- Piper 2018** Piper, A. *Enumerations*, The University of Chicago Press, Chicago (2018).
- Plecháč and Kolár 2017** Plecháč, P. and Kolár, R. *Kapitoly z Korpusové Versologie*, Akropolis, Prague (2017).
- Plecháč et al. 2021** Plecháč, P., Kolár, R., Bories, A.-S. and Říha, J. "Tackling the Toolkit. Plotting Poetry through Computational Literary Studies", Prague: ICL CAS. <https://doi.org/10.51305/ICL.CZ.9788076580336>
- Popper 2002** Popper, K.R. *Conjectures and Refutations: The Growth of Scientific Knowledge*, 3rd ed. revised., Routledge and Kegan Paul, London (2002).
- Queneau 1961** Queneau, R. *Petite Cosmogonie Portative*, Gallimard, Paris (1961).
- Rapp 2017** Rapp, A. "Manuelle und automatische Annotation". In: Jannidis, F., Kohle, H., Rehbein, M. (eds.) *Digital Humanities*. J.B. Metzler (2017) https://doi.org/10.1007/978-3-476-05446-3_18
- Rastier 2002** Rastier, F. "Enjeux épistémologiques de la linguistique de corpus". In: G. W. (ed.). *Deuxièmes Journées de La Linguistique de Corpus*. Presses Universitaires de Rennes, Lorient, France (2002), pp. 31–46.
- Rebora et al. 2019** Rebora, S., Herrmann, J.B., Lauer, G. and Salgaro, M. "Robert Musil, a war journal, and stylometry: Tackling the issue of short texts in authorship attribution". *Digital Scholarship in the Humanities* 34(3):582–605. doi: <https://doi.org/10.1093/llc/fqy055>
- Rockwell and Sinclair 2016** Rockwell, G. and Sinclair, S. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, MIT Press (2016).

- Schöch 2017** Schöch, C. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama". *Digital Humanities Quarterly* 11(2).
- Simmler et al. 2019** Simmler, S., Vitt, T. and Pielström, S. "Topic Modeling with Interactive Visualizations in a GUI Tool". *Proceedings of the 2019 Digital Humanities Conference*. Utrecht (2019).
- Simpson 2004** Simpson, P. *Stylistics. A Resource Book for Students*, Routledge English language introductions, Routledge, London [u.a.] (2004).
- Smith 1973** Smith, J.B. "Image and Imagery in Joyce's Portrait: A Computer-Assisted Analysis". *Directions in Literary Criticism: Contemporary Approaches to Literature*. The Pennsylvania State University Press, University Park, PA (1973), pp. 220–227.
- Smith 1978** Smith, J.B. "Computer Criticism". *Style* XII(4):326–356.
- Smith 1980** Smith, J.B. *Imagery and the Mind of Stephen Dedalus: A Computer-Assisted Study of Joyce's A Portrait of the Artist as a Young Man*, Bucknell University Press, Lewisburg, PA (1980).
- Smith 1984** Smith, J.B. "A New Environment For Literary Analysis". *Perspectives in Computing* 4(2/3):20–31.
- Stamatatos et al. 2014** Stamatatos, E., Daelemans, W., Verhoeven, B., Potthast, M., Stein, B., Juola, P., Sanchez-Perez, M.A. and Barrón-Cedeño, A. "Overview of the Author Identification Task at PAN 2014". *Working Notes Papers of the CLEF 2014 Evaluation Labs. CEUR Workshop Proceedings*, 877-897 (2014).
- Steen et al. 2010** Steen, G.J., Dorst, A.G., Herrmann, J.B., Kaal, A.A., Tina, Krennmayr. and Pasma, T. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*, John Benjamins, Amsterdam and Philadelphia (2010).
- Steyvers and Griffiths 2007** Steyvers, M. and Griffiths, T. "Probabilistic Topic Models". *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum (2007), pp. 424–440.
- Suppes 1968** Suppes, P. "The Desirability of Formalization in Science". *The Journal of Philosophy*, 65(20), 651–664 (1968).
- Svensson 2015** Svensson, P. "Sorting Out the Digital Humanities". *A New Companion to Digital Humanities*. John Wiley and Sons, Ltd (2015), pp. 476–492. doi: <https://doi.org/10.1002/9781118680605.ch33>
- Terras et al. 2013** Terras, M., Vanhoutte, E. and Nyhan, J. *Defining Digital Humanities: A Reader*, Routledge, London/New York (2013).
- Traub and van Ossenbruggen 2015** Traub, M.C. and van Ossenbruggen, J. *Workshop on Tool Criticism in the Digital Humanities - Report*, CWI Techreport, Amsterdam (2015).
- Underwood 2019** Underwood, T. *Distant Horizons: Digital Evidence and Literary Change*, First edition., University of Chicago Press, Chicago (2019).
- Unsworth 2000** Unsworth, J. "Scholarly Primitives: What Methods do Humanities Researchers Have in Common, and How Might our Tools Reflect This". *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. King's College, London. (2000).
- Weizenbaum 1984** Weizenbaum, J. *Computer Power and Human Reason: From Judgment to Calculation*. Penguin, Harmondsworth. (1984).
- Winko 2015** Winko, S. "Zur Plausibilität als Beurteilungskriterium literaturwissenschaftlicher Interpretationen". *Theorien, Methoden und Praktiken des Interpretierens*. De Gruyter, Berlin, Boston (2015), pp. 483–511. doi: <https://doi.org/10.1515/9783110353983.483>
- van Cranenburgh 2012** van Cranenburgh, A. "Literary Authorship Attribution With Phrase-Structure Fragments". *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Association for Computational Linguistics, Montréal, Canada (2012), pp. 59–63.
- van Es et al. 2018** van Es, K., Wieringa, M. and Schäfer, M.T. "Tool Criticism: From Digital Methods to Digital Methodology". *Proceedings of the 2nd International Conference on Web Studies*. ACM, New York, NY, USA (2018), pp. 24–27. doi: <http://doi.acm.org/10.1145/3240431.3240436>

