Deconstructing Bricolage: Interactive Online Analysis of Compiled Texts with Factotum

Tomas Zahora <tomas_dot_zahora_at_monash_dot_edu>, Monash University
Dmitri Nikulin <dnikulin_at_gmail_dot_com>, Google
Constant J. Mews <constant_dot_mews_at_monash_dot_edu>, Monash University
David Squire <david_dot_squire_at_monash_dot_edu>, Monash University

Abstract

Textual bricolage, the unacknowledged re-use of chunks of existing texts within a new composition, spans the liminal space between authorized, publicly shared, and de-authorized texts. While it can result in unique literary juxtapositions, bricolage also challenges the boundaries of authorial ownership. Understanding the methods and responses to textual bricolage reflects how a culture engages with textuality. Yet such study is often hindered by the sheer extent of compared texts. In this article we explore the potential of using Factotum, text similarity recognition software with visual interface, for analysing textual bricolage. Using examples from medieval and recent texts, we discuss different compilation techniques as well as the interaction between the notions of authorship, plagiarism and intertextuality.

Early in 2010, Helen Hegemann’s autobiographical novel of teenage angst *Axolotl Roadkill* (2010), a “ball lightning in prose form and language” as one review described it, became an instant sensation in Germany.[1] The book’s jarring format, its multiple layers and dimensions, its playful juxtaposition of the staid and known to expose fresh, unknown patterns, struck a chord among critics:

... all that which has been thought of, said, done, and accomplished a hundred times already she has absorbed, bundled, and transformed into something entirely new and unheard of, in an approach to literature that is beautiful not despite but because of its hardness, brutality, and vulgarity.[2]

The 17-year-old author was on track to receive the prestigious Leipzig Book Fair Prize when accusations of plagiarism began to surface after a reader recognized a number of passages, some amounting to entire pages, copied without acknowledgment from the novel *Strobo* by [Airen 2010]. Soon the list of unacknowledged sources grew to include a number of other novels, as well as excerpts from songs, movies, blogs, and interviews. What critics praised as a juxtaposition of seemingly different voices turned out to be precisely that: a compilation of fragments of the works of others enveloped by Hegemann’s own words. *Axolotl Roadkill* was withdrawn from prize nomination in the middle of what became one of Germany’s most involved public debates on the nature of authorship and plagiarism in the internet age [Stich 2010, 7–11].

The critical response to *Axolotl Roadkill* brought into prominence a tension inherent in artistic manipulation of existing texts. Hegemann’s opponents argued that the book’s large-scale borrowing did not acknowledge its sources and thus constituted plagiarism. By disregarding the authorial rights of others, Hegemann breached a set of conventions between writers, publishers, and distributors that have been developing since the introduction of the printing press [Thomasius 1679]; [Grafton 1990]; [Mulsow 2006]. Her supporters, on the other hand, could refer to a long tradition of unacknowledged textual borrowing: the technique of *cento*, poems composed entirely of appropriated lines, which predates Christianity; the ubiquitous unreferenced allusions to canonical texts like the Bible, Goethe’s *Faust*, or the plays of Shakespeare; or the artistic practice of textual bricolage consisting of conscious appropriation and transformation of chunks of existing texts [Meltzer 1994]; [Hume 2001]. Viewed from the latter perspective *Axolotl Roadkill* did not introduce something...
radically new, only pushed the boundaries of convention — something artists are expected to do, after all.

But the boundaries that distinguish textual appropriation, allusion, and plagiarism are not easily drawn with precision. In communities with established textual canons, the recognition by the reader of an unacknowledged canonical allusion is a mark of belonging, even status. In a work like Melville’s *Moby Dick* or Joyce’s *Ulysses*, the occasional verbatim borrowing adds another layer of meaning and pleasure to those who can recognize the appropriation. The allowable extent of such borrowing is not strictly defined, though. And even the most strictly regimented intellectual communities, such as medieval universities, could be inconsistent in applying their conventions. Medieval authors did recognize a hierarchical order of texts based on their authoritative value and proximity to truth [Minnis 2010]. By convention, the voice of an *auctor* carried what we would call authorial originality, while works of authors who did not yet reach *auctoritas* were often appropriated without acknowledgment. In reality, though, many works or their sections appeared under the names of other authors or were deauthorized altogether. Whether this happened by authorial choice or through the decision of subsequent copiers and readers, the results were not dissimilar to the intertextuality of Hegemann’s novel [Calma 2011]; [Zahora 2012b].

Postmodern literary intertextuality has been even more playful and subversive of conventions. The erosion of the concept of a literary canon, combined with the vastness of the global cultural sphere, has brought attention to the extent to which artistic creation is in fact re-creation of existing models. As Rachel Galvin puts it, "Cast positively, cultural recycling may be viewed as homage, pastiche, parody, or ludic recombination; on the flip side, it engages in deception, forgery, theft, plagiarism, hoax, usurpation." [Galvin 2014, 25]. A work that depends almost exclusively on other texts makes the distinction difficult, leaving the reader uncertain whether the extensive dependence reflects clever artistic choice or thinly disguised theft. The belated insertion of references into editions of *Axolotl Roadkill* suggests that Hegemann’s use of *Strobo* may have been intentionally disguised — that the text was deauthorized, decontextualized, and re-used as raw material. Did Hegemann commit theft, or did she engage in cultural recycling along the lines of Barthes’ death of the author? Critics remain divided as to the appropriate moral judgment, and the text itself gives us little insight into her intentions.

While the moral judgment of Hegemann’s textual appropriation remains uncertain, the novel reflects an increased interest in textual bricolage — composition by extracting, rearranging, modifying, and compiling large chunks of pre-existing texts — facilitated by instant access to a growing multitude of freely available online texts [Pecorari 2008]; [Blum 2009]; [Fish 2010]. To a medievalist, the phenomenon brings interesting parallels with intertextual practices of the Late Middle Ages and Early Modern Era. It reminds us that conventions regarding the use of sources not only differ among textual communities, but change over time. However, an understanding of the processes inherent in bricolage and compilation is often hindered by the overwhelming extent of compared material.

In this paper we explore the possibilities of analyzing textual bricolage with text recognition software Factotum, developed by a team of researchers at Monash University in Melbourne, Australia. Originally conceived as a tool to study the early-fourteenth-century encyclopedia *Speculum morale* composed almost entirely of unacknowledged sources, Factotum has developed into a tool for visual comparison of intertextual relationships between large texts. While software like Turnitin[3] effectively identifies the sources from which an author borrows, Factotum offers a platform for identification and analysis of the compiler’s method of arrangement, and thus for recognizing the interplay of different authorial and textual voices. Accomplishing in seconds what could take even a trained reader weeks or months of meticulous work, Factotum opens up space for critical analysis of ways in which texts can be manipulated to create a significantly altered or entirely different message.

**Textual borrowing in medieval compendia**

Intuition, aided by popular representations like Umberto Eco’s *Name of the Rose*, tends to associate the Middle Ages with a scarcity of books, jealously guarded and accessible to the initiated alone. There is some truth to the stereotype as far as the situation in much of medieval Europe is concerned. But when it comes to major centers of knowledge production like Paris, the stereotype becomes too simplistic. In the thirteenth century, Vincent of Beauvais (c. 1190-c.1264), composed a three-volume encyclopedia, the *Speculum maius*,
not because of a lack but "because the multitude of books, brevity of time, and instability of memory do not allow the soul to take in equally all that is written." [4] Its three books (the Speculum historiale, the Speculum naturale and the Speculum doctrinale) provide a compendium of learning, so that one book could stand for many. Vincent was also creating a guide to sources whose use among ever-multiplying treatises and commentaries was often less than reliable. His method, expressed in words that offer an interesting comparison to Hegemann's novel, recognizes that the innovation lies not in the contents themselves but in the way they are presented:

... this work is at once both new and old, both brief and prolix: look, it is ancient in its contents and authority, but new in its compilation or arrangement of parts; it is brief inasmuch as many sayings were constrained into compact form, but nevertheless long due to the immense multitude of subjects. [5]

The degree of self-awareness and high standards of editorial practice make Vincent's work a landmark of compilation, but also a rather unique occurrence [Paulmier-Foucart & Duchenne 2004, 23–115]. Many of his contemporaries, while equally dependent on the works of others, exercised a more liberal approach to utilizing them. After the massive influx of textual material produced in the thirteenth century, succeeding generations of scholars often chose to adjust and reuse rather than compose entirely new texts. A good example is the work of Pierre d'Ailly (1351-1420), which borrows without restraint or acknowledgment from the writings of Jean de Mirecourt, Gregory of Rimini, William of Ockham or, as in this instance, the prayer of Boethius asking for divine inspiration, also imitated by Alan of Lille [Calma 2011, 565]:

|---|---|

Table 1.

One way of dealing with such compilations would be to classify them as plagiarism and reject them as recycled or "vulgarized" versions of more "pure" originals: an approach that has been applied to medieval thought since the sixteenth century [Zahora 2012a]. But this would not only subject these works to notions of authorship postdating their composition by several centuries; it would also prevent us from understanding the full context of reception and actual use of those works we consider original, authoritative or innovative [Ribémont 2002, 10–11]. Pierre d’Ailly was neither the only scholar resorting to such practice, nor did he or his contemporaries perceive it as problematic or culpable.

Given the shortcomings of the expression "plagiarism," Monica Calma has adopted the term bricolage, used by literary scholars, to characterize medieval compositions that closely follow an (often unacknowledged) original, alternating verbatim copying with paraphrase, rearrangement, and additional comments. Bricolage differs from allusion and direct quotation by utilizing more extensive verbatim chunks of the original, and by re-using texts from a wider spectrum of authoritative values, including those that did not achieve the status of an auctoritas [Calma 2011]. Calma’s suggestion not only makes sense of the work of Pierre d’Ailly, it also opens up the interpretation of a large body of sources hitherto underrepresented in research or ignored altogether.

Medieval preachers’ aids and collections of moral examples are among the most dependent on other works. An excellent illustration is the extensive encyclopedia of moral philosophy and theology Speculum morale [Vincent of Beauvais 1979], associated with Vincent of Beauvais but composed after 1300, at least a generation after his death. The work contains an extensive coverage of human behavior, a topic not covered by Vincent of Beauvais. Yet it has been largely ignored by scholars based on the assumption it has simply been "plagiarized" from six unacknowledged "primary" sources — with the largest borrowing from Thomas Aquinas’ (1225-1274) Summa theologiae — that alone deserve close study [Echard 1708]; [Zahora 2012a]. In fact the Speculum morale demonstrates how an unknown author has participated in the transformation of
traditional ethics by introducing a new theory of emotions formulated by Aquinas, in particular his important and innovative discussion of emotions as in themselves morally neutral. Unfortunately, such insights are difficult to extract from a work heavily dependent on verbatim transcription, in the same order, of the responses to Aquinas’ questions, as in the example below. [7]

![Figure 1. In the right column is the original text of Aquinas’ Summa. In the left column is the same text as appropriated by the Speculum morale. Highlighted bold font identifies exact verbatim match; sections of regular highlighted font mark slight alterations, such as the use of “v” for “u” or “sicut” instead of “sic.”](image)

The size and nature of the Speculum morale make thorough textual analysis a daunting task. The most recent definitive study was completed by Jacques Echard in 1708 and extends to over 600 pages. No one has attempted to replicate his efforts since — discouraged no doubt as much by the 1558 columns of the printed edition as by Echard’s conclusion that the Speculum morale is a more-or-less worthless farrago of a clumsy plagiarist [Echard 1708, 101]. And yet, as Echard himself admitted, the compiler was not entirely a mindless copier of his sources. At times he changed the order of Aquinas’ articles, paraphrased, eliminated, or inserted passages from other sources and his own observations. In other words, the Speculum morale is more than the sum of its parts. Its voices bear traces of actual use of Aquinas’s works, as well as of the interplay of theology, philosophy, and sermon source material during a period of consolidation of medieval scholasticism.

Understanding the interaction of the mass of threads that comprise the Speculum morale entails the creation of a correspondingly massive research apparatus. Echard’s linear expositions of textual parallels run to hundreds of pages of lists: our Excel spreadsheet has three tables corresponding to the three books of the encyclopedia and a total of 837 rows at 13 columns each, with additional tables for more detailed analysis.

For a close reading, rather than relying like Echard on a superb memory and intimate familiarity with French archives, the Monash team has approached the task by using the experience in electronic plagiarism detection and the increasing availability of electronic texts [Mews et al. 2010]. The result of the efforts is text recognition software Factotum. Factotum enables us to move away from the plagiarism mindset by mapping the mutual relationships of the sources of Speculum morale identified by Echard, and by focusing directly on places where the original material is transformed or rearranged, rather than on the overwhelming bulk of verbatim transcriptions from other sources.

**Deconstructing bricolage with Factotum**

Factotum (http://webfactotum.com) combines experience with plagiarism detection software Damocles (http://viper.csse.monash.edu.au/damocles/about/) created by David McG. Squire, the specific needs of historians analyzing compiled texts, and expertise in algorithms suitable for advanced text recognition. It
allows for comparison of texts within its own database, or of user-contributed texts against all other texts within the program. The outcome can be simple identification (e.g. recognition that the *Speculum morale* has textual similarities with the Vulgate version of the Bible or with the *Summa theologiae* of Thomas Aquinas), close comparison of two chosen texts, or deep analysis in which all textual parallels are explored among a number of selected texts.

The program operates as a web application written in Scala (http://www.scala-lang.org/) using the Lift web framework (http://www.liftweb.net/), and is built around two core elements: algorithms and data structures relevant to analysis and interaction tasks, and a graphic interface that addresses the needs of analyzing large texts. Factotum is designed to process raw textual material into a canonical form that can be readily matched, to align inexact matches over sub-sequences of billions of words of text, to link together separate matching components of a passage that may be split over multiple source texts, and to reconcile the entire match data set into an HTML display that the user navigates interactively without ever having to download the full text or match data set. After selecting search criteria like the number and order of words in a searchable sequence, the user is able to select matches to record in permanent reports that are also stored and displayed on Factotum, and can be exported into LaTeX for easy inclusion in publications.[8]

An essential characteristic of Factotum, key to its thinking beyond plagiarism, is the availability of visual representations that allow the user not only to identify the location of source texts within a document (overall view), but also to study their transformation and mutual relationships (scrollable side-by-side parallel view).

The overall view, a visual display of the presence of a source text within another, is a graph that bears some similarity to DNA sequencing images. For the purposes of an instant overview, each text is divided into horizontal segments of 400-point width, where each point represents a cluster of matchable content which and becomes highlighted in the instance of a positive match. For example, the dependence of the first book of *Speculum morale* on the *Summa theologiae* below reveals the heavy presence of the *Summa* in the *Speculum morale*, but also its spatial distribution within the *Speculum*’s 1049 clusters (top part of the graph) along with the extensive recourse to the innovative central part, *Prima and Secunda Secundae*, of Aquinas’ *Summa* (the higher density of highlighted areas in the central portion of the lower part of the graph).

---

**Figure 2.**

The overall view display allows the researcher to gauge the extent of potential textual coincidence, as well as to analyze a particular part of the document by selecting a highlighted segment.

While the horizontal layout of the overall view provides a global survey of the works’ mutual relationship, the vertical layout of the scrollable side-by-side parallel view allows for close textual comparison of individual passages. The two columns let the reader navigate through the entirety of either text, marked with highlights identifying parallel passages. Numbered hyperlinks at the end of each matching passage identify the locations of textual parallels in the related document. By clicking on the hyperlink, the texts will align at the location of the match Figure 3.
In the case of the *Speculum morale*, the visual display allows easy access to the compiler's method. We have already seen an example of pure copying from the *Summa*; the example in Figure 3 shows another common method, a combination of text from arguments of Aquinas’ *Summa theologiae* followed by a more-or-less verbatim citation of the conclusion.

The bolding of matching text and highlighting of related passages facilitates instant recognition not only of textual borrowing but also of lacunae and paraphrase, as in the section on human works in the *Speculum morale* (Figure 4).
peccatum” (Therefore in the first way, sensible things appearing on the outside move the will of man towards sin, but in the second and third way either the devil or even man himself can incite to sin) belongs to the section on sin, hence the focus on the different channels of temptation. But the author of the Speculum morale used Aquinas’ text in another setting: the principal origin of human works, both good and bad. In the compiled version, much of Aquinas’ wording was preserved, but also amended to point out the usefulness of the created world: "Therefore in the first way, sensible things appearing on the outside move the will of man toward the good or the bad, just as contemplation of created things moves and induces man to praise the wisdom, honor the omnipotence, and love the goodness of the creator..."

The changes made by the compiler are relatively insignificant in terms of word count, but they, along with the methods of rearranging sources, make the Speculum morale a work with its own character. Factotum allows the reader to recognize the extent of borrowing, observe the modes of textual rearrangement, and finally focus on passages where the transformation of material or addition of unidentified passages becomes interesting. Instead of months of reading and comparison, the task of identification and highlighting of parallels is reduced to seconds.

Selecting different keyword cluster lengths, searching for word clusters in random order within larger groups of words, or expanding the search by a number of words before and after matched cluster results in a variety of results that can suit a wide range of texts. Figure 5 shows the search set at six keywords exactly, a convenient setting for a text heavily dependent on verbatim borrowing such as the Speculum morale. Choosing lower numbers, or ignoring word order, will result in a larger numbers of false positives but may be suitable for the study of paraphrase, more subtle textual borrowing, or for shorter texts. [9]

![Match definition](http://www.digitalhumanities.org/dhq/vol/9/1/000203/000203.html)

**Figure 5.**

**Bricolage in Moby Dick and Breivik’s Manifesto**

Because Factotum does not rely on a specific dictionary, it can be used to uncover several layers of intertextual relationships in any language. Here we present two examples in English, one identifying the presence of the King James Bible in Herman Melville’s Moby Dick, and the other examining the extent of borrowing from Ted Kaczynski’s Unabomber Manifesto in Anders Breivik’s manifesto 2083 A European Declaration of Independence.

While Moby Dick [Melville 1962] can hardly be called a bricolage, Melville’s incorporation of exact quotations from the King James Bible along with paraphrase and allusions offered a great test case for Factotum. On the most basic level, the program identified verbatim quotations, such as the presence of the text of Jonah in the chapter called The Sermon. The visual presentation reveals Melville’s elaboration of Jonah 1.8-1.10, as well as the insertion of the reference to the hand of God, with links to multiple occurrences of the phrase in the Bible.
In another case, the "darkness, and the weeping and wailing and teeth-gnashing" was matched against Matthew 8:12 (link 2602 in the left column below), but also against Matthew 22:13 (link 2657) and 25:30 (link 2672). This case shows both the strengths and the limitations of Factotum. The reference to Matthew is not noted in some editions, as the Mansfield-Vincent edition of 1952. In this case Factotum did identify an unnoticed textual coincidence. But at the same time the program failed to recognize the word Tophet — which the edition does comment on — because the smallest cluster recognizable by Factotum consists of three words [Melville 1962, 604].

Apart from clearly identifiable biblical matches, Factotum identified a number of loose matches — possible paraphrase, analogy, or false matches. In chapter three, "The Spouter Inn," an interesting parallel with Luke 17:34 appeared in the context of two men sleeping in one bed Figure 8. We will leave aside the question of whether Melville was thinking of the Gospel of Luke or simply describing a scene. What the example does demonstrate is Factotum's processing of loosely analogous or rephrased passages. As the search settings move away from exact matching, the number of potential matches will increase.

The flexibility of search settings and precision of matching can result in concordance-like comparison of the occurrence of a particular passage in another text. Thus Factotum picked up the parallels between Melville's repeated use of the word seven hundred and seventy-seven, and its occurrence in a number of places in the King James Bible, two of which are shown in Figure 9.

Figure 6. Textual parallels between Herman Melville’s Moby Dick (left column) and the King James Bible (right column).
Depending on the settings, Factotum can identify dozens or hundreds of matches of varying faithfulness to the original wording of the text. Instant access to results allows the researcher to overview different matches, reject false or marginal matches, and focus on the most promising ones. While the probability of discovering hitherto unrecognized biblical parallels in a text studied as closely as Moby Dick is minimal, the different recognition filters allowed by the settings allow for the study of the different layers of textuality present in the novel, and offer a layer of macro-perspective on the heteroglossia present in Melville’s work [Alter 2010, 44–45].

A compilation like Anders Breivik’s manifesto, on the other hand, is an ideal subject for Factotum. The right-wing extremist’s 2083: A European Declaration of Independence [2011] is a compilation of acknowledged and unacknowledged sources ranging from St Bernard’s founding manifesto of the Knights Templar In praise of a new knighthood, to contemporary extremist conservative blogs, and Ted Kaczynski’s Unabomber Manifesto [1995]. Breivik’s approach to his sources is similar to that of the compiler of the Speculum morale, in that he appropriates entire passages with minimal author input. Yet as van Gerven Oei [2011] has pointed out the changes to the text, minute in extent, are important to understanding Breivik’s motives and mindset.

Factotum comparison of 2083: A European Declaration of Independence and the Unabomber Manifesto (Figure 10) reveals both the extent and nature of Breivik’s borrowing [Breivik 2011]; [Kaczynski 1995].

In scope, the borrowing from Kaczynski is relatively minor, amounting to only a few of the work’s 1500 pages. Breivik draws from the opening passages of the Unabomber manifesto and concentrates them into two distinct sections, clearly visible in the graph above. Visual comparison of the two texts, displayed in Figure 11, reveals near-verbatim reproduction of Kaczynski’s text by Breivik, with subtle changes confirming the observation made by van Gerven Oei that Breivik appropriated a text referring to specifically American conditions and recontextualized it in European settings.
Kaczynski, writing about "leftists," locates political correctness among "university professors, who have secure employment with comfortable salaries, and the majority of whom are heterosexual, white males from middle-class families." Breivik incorporates the accompanying text, but his version refers to "cultural Marxists," and adds government employees, politicians, journalists, and publishers in government broadcasting companies to Kaczynski's university professors, while replacing "white males" with "Ethnic Europeans." In a similar manner, Kaczynski's list of women, American Indians, and homosexuals as groups with which adherents of political correctness identify is transformed in 2083 by replacing American Indians with "so-called oppressed minorities," and "inferior" with "other groups in the victim hierarchy."

Our example replicates van Gerven Oei's findings as an illustration of Factotum's ability to identify matching passages, additions, lacunae, and rearrangement of originals, along with providing a global visual overview of the relationship of the compiled text to its source. The instant access to potentially significant passages spares the researcher wading through a mass of text, much of which is copied verbatim, and focus on material that can produce a more solid analysis.

Due to copyright restrictions we did not scan and process Hegemann's Axolotl Roadkill [2010] against Airen's Strobo [2010] in their entirety. The situation has its ironies, but even an analysis of select passages through Factotum does suggest that Hegemann was engaging in more than simple cut-and-paste bricolage. For example, she slightly reworded clusters of dramatic scene descriptions, as in the discussion of sex in Figure 12.

In another scene, in which drugs play an essential role, she doubled Airen's quarter (Viertel des Pulvers) of powder to a half (Hälfte), before swerving the discussion in a new direction (Figure 13).
In all the textual parallels we identified, clusters of Airen’s text were adjusted by slight rewording and recontextualization. They are distinctly Airen’s, yet they appear as such only to those familiar with the original, making the result similar to students’ unconventional use of texts by paraphrase, usually classified as plagiarism [Pecorari & Shaw 2012]. To the unaware reader the chunks, borrowed from another text, simply vanish and become, as the reviewer noted, “absorbed, bundled, and transformed into something entirely new and unheard of.” With Factotum, they — and the transformations they have undergone — become visible again.

Beyond plagiarism

In their defense strategy, Helen Hegemann and her lawyers argued that the author was not plagiarizing but recasting, in the manner of a visual-arts collage, elements of other material to create an authentic work with integrity of its own. Airen, whose novel Strobo she used as raw material, rejected such a possibility outright: “This [Strobo] is not a novel; it is my life as I lived it. I have not made it up. Helen Hegemann hasn’t experienced it. It is I who have experienced it this way.”[10]

Airen’s complaint is shared by the critics of the writer Kathy Acker as much as by Jacques Echard who denied any integrity to the compilation Speculum morale. To them, words based not on actions but on other words can only result in a simulacrum, a vague approximation of the “real” experience whose source must be acknowledged and given preference. But with the abandonment of a rigid definition of originality by literary and cultural studies, and the recognition that permutations of existing discourse can create new meaning, authenticity may be too vague a factor in examining textual bricolage. For all intents and purposes, Hegemann’s novel creates not only an illusion but the reality of literary truth — a mark, at least according to some, of its integrity [Woolf 1989, 72].

By consciously reusing unacknowledged texts, bricolage spans the liminal space between authorized, publicly shared, and de-authored texts. Understanding how some discourses are endowed with the author function while others are deprived of it [Foucault 1998, 211] can offer us a map of how a culture engages with textuality. In the medieval academic world, texts of scholastic authors who had not yet reached the rank of auctoritas or authority (and sometimes even afterwards) became the common currency of their near-contemporaries who amended and transformed their work without acknowledgment. A similar situation is found among today’s university students whose practice of weaving commonly used definitions from Wikipedia into their essays is often tacitly ignored by lecturers [Pecorari & Shaw 2012]. On the other hand, unreferenced use of authorized research sources, or extensive verbatim borrowing, will result in penalty. Hegemann’s belated retraction can be taken as an acknowledgement that she has crossed the currently accepted boundaries of convention. However, as our example of the Speculum morale has shown, the boundaries between acceptable and unacceptable intertextuality are not constant. In this light, the proliferation of re-use of readily available online texts represents another stage in an identifiable evolutionary process.

The visual access to detailed analysis of textual parallels that Factotum offers can lead to a better
understanding of how borrowed texts are shaped and reorganized, and how they interact with others. In that, it can provide a counterpoint to the observations of medievalists and literary scholars who have shown that a subtle authorial voice can be identified even in texts that defy conventional notions of authorship [Hume 2001]. Factotum has already contributed to the discovery of a new source of the Speculum morale, and has led to an understanding of the compiler’s technique that would take years to achieve using manual methods.

Even if mechanical identification of textual similarity remains a relatively crude tool in the recognition of complex textual interplay in novels like Moby Dick, by identifying matching texts and allowing the reader to judge not only the extent of borrowing but also the secondary author’s method of recasting the sources, Factotum opens the space for a meta-analysis of textual relationships. The experience of making the invisible seams of intertextual composition visible, and comparing similar texts side by side, can lead to more subtle observations than an outlook dominated by the concept of plagiarism.

There is a sense in which scholarly attitudes towards textual bricolage and plagiarism are already changing. Rather than focusing on the moral implications of allegedly timeless and universal principles, scholars like Stanley Fish have called for an open discussion of conventions that guide professional practice [Fish 2010]. The traditional way of dealing with textual parallels has been greatly improved by the searching capabilities of Google and proprietary databases like Brepols’ Library of Latin Texts,[11] as well as by progress in textual processing. Even so, analysis of textual bricolage remains time consuming, open to errors, and invites a very rigid, text-based approach: the memory, burdened by close study, may end up like the mind of Borges’ Funes who could remember every piece of text but was incapable of analytical thought and comparison. We are hopeful that Factotum is a step in the transformation of the way such texts are studied.

Notes

[1]”Ein Kugelblitz in Prosaform und Prosasprache,” [Stich 2010, 7].

[2] “… all das, was schon hundertmal gedacht, gesagt, getan und getragen wurde, hat sie aufgesogen, gebündelt und in etwas ganz Neues. Unerhört verwandelt, in den Ansatz zu einer Literatur, die nicht trotz, sondern wegen ihrer Härte, Brutalität und Vulgarität schön ist,” [Stich 2010]


[4] “Quoniam multitudo librorum et temporis breuitas memorie quoque labitas non patiuntur cuncta que scripta sunt, pariter animo comprehendi …” [Vincent of Beauvais 1979, 115].


[7] The text in the image is taken from an analysis by the Factotum program. It uses a transcription of the Speculum morale owned by the Monash Factotum team and a publicly available text of the Summa theologiae (http://www.corpusthomisticum.org).


Works Cited


