JEROEN DE GUSSEM

Larger than Life?

A Stylometric Analysis of the Multi-Authored *Vita* of Hildegard of Bingen

Abstract

This article explores by aid of stylometric methods the collaborative authorship of the *Vita Hildegardis*, Hildegard of Bingen's (auto-?)biography. Both Hildegard and her biographers gradually contributed to the text in the course of the last years of Hildegard's life, and it was posthumously completed in the mid-1180s by end redactor Theoderic of Echternach. In between these *termini a quo* and *ante quem* the work was allegedly taken up but left unfinished by secretaries Godfrey of Disibodenberg and Guibert of Gembloux. In light of the fact that the *Vita* is an indispensable source in gaining historical knowledge on Hildegard's life, the question has often been raised whether the *Life of Hildegard* is – by dint of contributions by multiple stakeholders – a larger-than-life depiction of the visionary's life course. Specifically the 'autobiographical' passages included in the *Vita*, in which Hildegard is allegedly cited directly and is taken to recount biographical information in the first-person singular, have been approached with suspicion. By applying state-of-the-art computional methods for the automatic detection of writing style (stylometry), the delicate questions of authenticity and collaborative authorship of this (auto?)hagiographical text are addressed.*

1 Authorship and Canonization of Hildegard

From the late twelfth century onward, the documentation and bundling of testimonies and/or medieval authors' saintly biographies and miracles into a sort of dossier was increasingly instrumental for achieving canonization, the elevation to sainthood (Vauchez). The procedures leading towards canonization became more and more systematised and bureaucratic, meaning that the papal See's official decision rather than popular veneration through cults or local, diocesan approval was the deciding factor (Kataja-la-Peltomaa). Correspondingly, the need for such a dossier in a regulated process for the endowment of sainthood, and an awareness of its having to compete with similar-looking dossiers, justified its redaction and contribution by multiple supporters and wit-

1. Abbot Bernard of Clairvaux's (1090–53) entourage, and most notably his secretary Geoffrey of Auxerre († after 1200), may be another example of this. Bernard's multi-authored *Vita prima* may have also been part of his canonization dossier (Bredero, *Études* 147–61; Dutton).

2. Alternatively also the *Canonizatio*, sent to Pope Gregory IX in 1233 to Rome by three clerics in Mainz for approval of Hildegard's canonization (edited in the late nineteenth century by Petro Bruder). Included was Theoderic's *Vita Hildegardis* and exemplars of her work (Ferzoco 306).

3. Wiesbaden, Hessische Landesbibliothek, 2. The end redaction of Hildegard's *Opera omnia* is now generally accredited to Guibert of Gembloux, a Brabantine monk who, as we will discuss further below, played an important role in the collecting and completing of Hildegard's oeuvre near the end of her life.

nesses,[1] and in some cases required the recruitment of professional hagiographers to respond to the sensitivities of the See's procedures. The multi-authored *Vita* of the Benedictine visionary Hildegard of Bingen (1098–79), a (auto-?)hagiographical account of her life and miracles, can be taken as an example of such a project. Forming part of a dossier containing the *Acta inquisitionis*,[2] a charter concerning Hildegard's virtues and miracles petitioning her canonization, and some fragments of her visionary treatises, the *Vita* was proffered to the Curia multiple times, to Popes Gregory IX (1237), Innocent IV (1243) and John XXII (1317). Only in 2012 the canonization process was completed, which makes it the longest of its kind in the history of the Church (Ferzoco; Newman, "St Hildegard").

Hildegard's literary legacy was intensely prepared in the final years of her life, the period from which most contemporary Rupertsberg manuscripts survive. Amongst other indications, this demonstrates that her life and authorship was a project in whose success her community in the Rupertsberg cloister had stakes long after their abbess had gone. The Wiesbaden Riesencodex,[3] a monumental codex containing Hildegard's *opera omnia* of all authorised versions of her visionary writings (except for her *Liber subtilitatum diversarum naturarum creaturarum*, or *Physica* and *Cause et Cure*), has been argued to have been issued under Hildegard's command, but was completed after her death (Embach 36–65). It was Hildegard's persona rather than her historical person that was documented and 'constructed' in the Riesencodex. As will be discussed below, Hildegard's entourage, especially her secretaries, played an indispensable role in composing, conserving and 'canonizing' the image of Hildegard for posterity which they hoped future generations would keep track of. The *Vita* has been a source of fascination to scholars not only because it gives an account of Hildegard's life, but also because it contains direct citations of 'autobiographical' fragments allegedly written by the visionary herself. However, the editor of the *Vita*, Monika Klaes, was unconvinced – as opposed to others (Schrader and Führkötter 14; Dronke 144) – that these autobiographical passages contain no interpolations and are authentically of Hildegard's doing (Klaes 113*–14*). Quite recently, Van Engen confessed doubts concerning their authenticity as well: "To what degree they, even if of authentic origin, underwent redaction lies beyond our ken" ("Authorship" 339). Taking into consideration that the *Vita* was interpolated in Hildegard's Riesencodex posthumously (Klaes 157*ff.; Derolez, "Manuscript" 23), that it went through the hands of at least three biogra-

phers, and that it had the intention of promulgating Hildegard's persona with lasting impact, due concern is warranted. Is this indeed Hildegard's *Life*, or is this a larger-than-life literary construction, and how did she participate in its composition?

In order to analyse the multi-layered authorship of Hildegard's *Vita*, this article will make use of computational stylistics or stylometry, a subfield of the digital humanities (DH), based on techniques from computational linguistics and natural language processing (NLP), in which statistical methods are applied to segregate writing styles, and used as a basis for assigning anonymous documents to their authors (Daelemans; Juola; Koppel *et al.*; Stamatatos).

2 Hildegard's Collaborators

It is well known that Hildegard of Bingen's authorship has not always been regarded with the same scholarly esteem it is presently granted. Partly this is related to the fact that she composed her works in collaboration with male secretaries such as her provost and lifelong confidant Volmar of Disibodenberg († 1173), or her last secretary Guibert of Gembloux (1124–1214), and female scribes, amongst whom was her close companion Richardis of Stade (Herwegen 1904). Hildegard presented these collaborations to her readership as necessary due to her limited schooling and deficiency in speaking and writing Latin. In the nineteenth century, Wilhelm Preger took the presence of her male secretaries and self-devaluations as evidence that her entire epistolarium (both outgoing as incoming letters) was a *falsum*, and that only a man could have been capable of the intellectual accomplishments that emerge from the remaining of Hildegard's works (*Geschichte*). Consequently, her entire oeuvre was ascribed to Theoderic of Echternach, ironically enough a monk who may never even have met Hildegard in real life. That Preger believed the cards were stacked in favour of Theoderic as male forger was probably not coincidental. Theoderic is, after all, the end redactor of Hildegard's *Life*, the *Vita Hildegardis*. In Preger's eyes the orchestrator of Hildegard's 'life' was most likely to be its overall fabricator.

Having for some time been denied authorship of her works, Hildegard has slowly but surely been regaining recognition for her work since Preger. Her revaluation was gradually set in motion with the pioneering work of Herwegen ("Les collaborateurs") and Liebeschütz (*Allegorische Weltbild*) in the early twentieth century. The sem-

inal work of the two Eibingen nuns Schrader and Führkötter (*Echtheit*) confirmed the authenticity of Hildegard's visionary trilogy consisting of *Scivias*, *Liber vitae meritorum* (*LVM*) and *Liber divinorum operum* (*LDO*). The second half of that same century also saw the arrival of critical editions by Führkötter and Carlevaris (*Scivias*), Carlevaris (*LVM*), Dronke and Derolez (*LDO*), Van Acker and Klaes (*Epistolarium*), and a series of scholarly publications by Dronke (*Women Writers*), Newman (*Sister of Wisdom*) and Deploige (*In nomine femineo indocta*) that have been able to help us understand the 'Sibyl of the Rhine' as a unique author with an impressive breadth of intellectual accomplishments. She was not only a visionary writer, but also a musical composer, a playwright, a healer, a scientist, a writer of letters, an inventor of languages, and a theological thinker in her own right.

Hildegard's self-deprecations of her weaker sex (*paupercula forma*) and lack of education (*indocta*) are now taken to be of a topical nature, with the aim of asserting her humility, or in some cases to stress that her utterings are not her own but divinely inspired (Powell). Even though she never directly cites any of her main sources, one must not underestimate the extent to which her illiteracy and lack of learning are exaggerations with a strategic purpose (Dronke, "Allegorical" 14). In being prohibited to teach doctrine, Hildegard had to find a delicate balance between self-devaluation and self-authorization, between proclaiming her insufficiency whilst asserting her divinely inspired authority. The condition of male supervision – her closest secretary Volmar was the Rupertsberg's provost – undoubtedly shielded her against criticism from outside, and also kept suspicion from within the institution of the contemporaneous church at bay.

Even though Hildegard's intellectual contribution to her own works is no longer contested, it is through an interest in the historical reality of medieval authorship and collective creativity for the Middle Ages in general that the contribution of her secretaries remains a matter of debate. Despite the fact that Hildegard issued warnings that her secretaries were not to change the sense of her visions, and were to focus on formal aspects of the language such as grammar and spelling alone (Ferrante 103), scholars have emphasised by means of palaeographical, codicological and computational evidence that the influence of these secretaries (still) is a fickle point of Hildegard scholarship (Herwegen; Derolez, "Deux notes;" Kestemont, Moens and Deploige). The extensive correc-

4. Ghent, Universiteitsbibliotheek, 241.

tions by secretaries in the Ghent apograph of the *Liber divinorum operum*,[4] for instance, may make one wonder to what extent the redaction and revision process by Hildegard's secretaries profoundly changed the final outlook of the visionary's works (Derolez, *LDO* lxxxix–xcvii). Recent stylometric evidence has confirmed Guibert of Gembloux's stylistic influence on two suspect visions, *Visio ad Guibertum missa* and *De excellentia Sancti Martini*, and a somewhat more subtle yet noticeable stylistic presence in Hildegard's letters written after Volmar's death in 1173, likely to have been revised by Guibert (Kestemont, Moens and Deploige).

In this article, we mean to emphasize that Hildegard's secretaries played a vital role in shaping her image and authority, both during her life and after her death. It posits that the aspect of collective creativity is fundamental in understanding both Hildegard as figure and as author, without seeking to undermine Hildegard as figure and author. Hildegard's collaboration with secretaries does not unilaterally constitute her fabrication (as Preger argued), neither did it unambiguously signify her suppression by a male patriarchy (i.e. the willful alteration of her words by male secretaries). Neither of these poles give a satisfactory explanation of the dynamics at play, but bypass a much more complex field of constant tension and negotiation in which mutual interests are at stake. As Johnson has noted, one should not forget that Hildegard's authority as writer benefitted from the endorsement and encirclement by male clerics, an effect which she cannot but have been aware of, and therefore incorporated into her texts (Johnson 823). This kind of subtle interplay between asserting her authority and having her authority asserted through involvement of others is always present. For instance, Hildegard grants Volmar a central role in having launched her writing career at the age of forty-two, but on the other hand fails to ever mention his name (Deploige and Moens 141). Hildegard –and/or her entourage – perceptibly sought for mechanisms by which to authorize her visionary writings by involving onlookers and alliances that recognized the divine origins of her extraordinary gift. Such attestations, however, are better not always accredited at face value. Volmar is known to have concocted a false letter from Pope Eugene III, which was presented as a prestigious first letter heading her epistolarium (Van Engen, "Letters" 380).[5]

5. The spurious letter is edited by Van Acker in the edited volume of Hildegard of Bingen's *Epistolarium Pars Tertia CCLI–CCCXC,* Ep. 1, at 173.

3 Composition History: The Life of a *Life*

When speaking of Hildegard's *Vita*, we are hardly speaking of a 'single' text. Multiple *Vitae Hildegardis* have circulated and co-existed. A few of them are finished works and still extant, others are fragmentary or partially recovered, and there are versions irrecoverably lost. We currently have two more or less 'final' versions of what is officially known as the *Vita Hildegardis* at our disposal.[6] The first version, widely regarded as 'canonical,' is the integral text taken up in Monika Klaes's edition. The second version is a revision of this text by Guibert of Gembloux, recoverable through Klaes' critical apparatus and an appendix in the aforementioned edition. The complex textual history of Hildegard's *Vita*, which is elaborately discussed in Klaes's lengthy introduction with a number of pages twice as long as the actual text, teaches us that its composition was accompanied by many difficulties. Despite best intentions, the project was on the verge of deferral due to a chain of unfortunate circumstances and unexpected deaths before the *Vita*'s completion, not least Hildegard's own death in 1179. Consequently, as will be further discussed below, the text has an intricate timeline, and presents an archetypal example of collaborative authorship. Five authors were (at least partially) involved, the first of which was allegedly Hildegard herself. The *Vita* contains so-called autobiographical fragments and snippets of visionary material, apparently dictated in the first person by Hildegard herself, and not repeated elsewhere in Hildegard's oeuvre. The remaining (co-)authors are her secretaries and biographers, in more or less chronological order of contribution: Volmar of Disibodenberg, Godfrey of Disibodenberg, Guibert of Gembloux and finally Theoderic of Echternach.

The academic consensus on the *Vita Hildegardis*'s composition history (treated in more detail below) makes it reasonable to suspect that the collecting of materials and the drafting of early versions of the *Vita* had already begun under Hildegard's direction (Newman, "Three-Part Invention;" "Hildegard").[7] In this she was assisted by her secretaries, probably Volmar first and – after the latter's death in 1173 – by Godfrey of Disibodenberg, who arrived in the Rupertsberg shortly after in 1174. Around 1175/6, only a year and a half later, Godfrey would also come to die (Derolez, *Epistolae* vi–vii). By the time of Godfrey's death, a partial *Vita* for Hildegard had been composed by him, a 'booklet' or *libellus*, which is believed to have been transmitted as the first book of the complete *Vita Hildegardis*. Whether or not the accounts collected in the first book of the *Vita* are indeed an in-

6. I leave aside for now the anonymous *Octo lectiones in festo Sanctae Hildegardis legendae* and the abbreviated *Vita* (*abbreviata Traiectensis*) by Guibert of Gembloux, both of which are derivations from the 'official' *Vita* that will be discussed here (texts edited in *Vita* 75–80; 83–88). I also pass over the *Acta inquisitionis* mentioned earlier in this article.

7. Amongst other reasons given further down this article, the writing of a *Vita* by Godfrey as reported by Theoderic (between 1174–75/6) and the *Vita*'s integration in Hildegard's Riesencodex (likely to have begun before her death in 1179) strengthens that hypothesis.

tegral copy of Godfrey's original booklet is unknown: other secretaries of Hildegard and her succeeding biographers will have had ample opportunity to revise the text (Klaes 91*; Newman, "Hildegard" 17).

The first of Godfrey's successors was Hildegard's last secretary Guibert of Gembloux (c. 1124–1214), a Brabantine monk who came to her aid from 1177 onwards and would assist Hildegard until she passed away two years later. During his time at the Rupertsberg, the Colognian archbishop Philip I of Heinsberg appears to have sparked in Guibert the intention to write a *Vita Hildegardis* of his own.[8] Guibert hesitated to obey Philip's request during Hildegard's lifetime, fearful of being found sycophantic by Hildegard, and it was only after her death in September 1179 –during his research for materials – that he fell upon useful sources to facilitate the task. From his description of these findings, one can suspect that Guibert had come across autobiographical memoirs of Hildegard and Volmar, and a *libellus* that might well correspond to Godfrey's first *Vita*.[9] Therefore, some time between Hildegard's death and Guibert's departure from the Rupertsberg, progress for a new *Vita* appears to have been well under way. These plans were disturbed when –after Easter 1180 – Guibert was forced to return from the Rupertsberg on appeal of the abbot of Gembloux (Moens 74). It appears that this event left Guibert unable to finish the work he had started on his *Vita*, although a fragment of his efforts at the time is presumed to have survived as an attachment to a letter addressed to his fellow monk Bovo of Gembloux.[10]

What should be emphasized here, is that Guibert's last year in Bingen (1179–80), which revolved entirely around the collecting of the prime sources of the *Vita*, left ample occasion to rework and revise Hildegard's materials –particularly the 'autobiographical' fragments –, and additionally (potential) preparatory versions by Volmar and Godfrey, of which we do not entirely know which parts were incorporated and, if so, if they were considerably amended or preserved in their original state (Klaes 58*). Former research has already indicated that the large role Guibert played in compiling and editing Hildegard's works cannot be underestimated, an activity which set out after his arrival in 1177, in the last two years of the visionary's life. Guibert supervised the scriptorium's activities at a time when Hildegard's epistolarium and the Riesencodex were in the final stages of completion (Van Acker, "Briefwechsel" 129–34). As recent research has pointed out, Guibert appears to have been granted –or appears to have taken – unprecedented liberties in editing and revising Hildegard's works, as is asserted in a letter from Hildegard which bears

8. This becomes clear from Guibert's letter to Philip I, edited as Ep. 15 in Derolez's 1988–89 edition of the *Epistolae* 210–15. The letter is difficult to date, but its contents suggest the year 1180, at which time Guibert was still in Bingen, and had just caught news of his being recalled to Gembloux (Klaes 30*–31*).

9. All of which is also described in Ep. 15 (see previous footnote). Because the *termini post* and *ante quem* for the dating of the letter lie in between Hildegard's death in 1179 and that of Philip I in 1191, the text(s) which correspond(s) to this *libellus* cannot be securely reconstructed. Either the mentioned *libellus* is exclusively Godfrey's first book of the *Vita,* or else it is the redacted version by Theoderic. The strong evidence that Guibert wrote Ep. 15 in 1180, when Theoderic of Echternach was yet to arrive in Bingen, makes the former hypothesis more likely (Klaes 30*).

10. Ep. 38 in Derolez, *Epistolae* 367–79. Guibert's *Vita* breaks off mid-sentence in the best-conserved manuscript of the letter collection (Brussels, Royal Library, 5527–34); it is presumed to have originally been longer (Klaes 42*–43*). The letter itself purports to have been composed in 1177, but in reality it must have been finished by the end of 1179 and in the beginning of 1180, when Guibert was still in Bingen.

Guibert's style completely (Kestemont, Moens and Deploige 202–4). When taking into consideration how Guibert left his stylistic mark on Hildegard's works in the final stages of her life and how he did not hesitate to alter the visionary's wordings, one may become wary concerning the monk's early involvement in the composition of the *Vita* as well (Klaes 113*).

Volmar's and Godfrey's deaths, and Guibert's commitment to new priorities on his path, left the late Hildegard still without a *Vita*. At the instigation of abbots Ludwig and Godfrey of Saint-Eucharius, the task consequently fell to Theoderic of Echternach, an unlikely candidate, as the latter may never have even met Hildegard in person (Klaes 60*–61*, 77*). Interestingly, at such a decisive moment for the project, when Theoderic followed up on a task left unfinished by Guibert, the odd fact that the monks did not cross paths and, even stranger, somehow neglected or missed out on each other's work, remains difficult to explain. Theoderic shows no familiarity with Guibert's extant fragment, and Guibert appears to have lost track of Theoderic's progress on the *Vita*. Still, there are parallels between both writers' *Vitae*, which indicate their dependence on the same pool of consulted source materials probably first collected by Guibert.[11] From the listing of sources in his preface, Theoderic indeed appears to have consulted the same sources for his *Vita* as Guibert, namely Godfrey's unfinished *libellus* and snippets of Hildegard's visions.[12] This means, as was suggested earlier, that Theoderic might have used source materials heavily revised by Guibert. Theoderic's role then, was that of editor-in-chief, a role corresponding to a kind of narrator or commentator, tying together the seemingly unrelated bits and pieces that had coincidentally fallen into his hands. The general structure of the *Vita*, then, and the purported authors of its constituents, is the following:

11. This is most clear from both *Vitae's* usage of a fragment from Hildegard's letter to Guibert, the *De modo visionis sue* (Klaes 48*). The complete letter is edited as Ep. 103 (Van Acker 2:258–65; compare with *Vita* 54–75, 260–61).

12. As can be deduced from the *Vita's* prologue: "[...] Accepi, ut post Godefridum, uirum ingenio clarum, uitam sancte ac Deo dilecte Hildegardis uirginis, quam illo honesto stilo inchoauit, sed non perfecit, in ordinem colligerem et quasi odoriferis floribus serta contexens uisiones eius gestis suis insertas sub diuisione librorum in unius corporis formam redigerem" (*Vita* 5–10, 3).

Table 1. Structure of the *Vita Hildegardis*

Author	Title (or *incipit*)	Ed. (Klaes,
Theoderic of Echternach	*Prologus in vitam*	3–4
	—*capitula*—	5
Godfrey of Disibodenberg	I. *Liber primus* (*libellus*)	6–16
	Prologus in librum secundum	17–18
	—*capitula*—	19
Theoderic of Echternach	II. *Liber secundus*	20–45
	Prologus in librum tercium	46
	—*capitula*—	47–48
	III. *Liber tercius* (*De miraculis*)	49–71

13. Especially for Hildegard's autobiographical fragments in books 2 and 3, Theoderic insists that he has left them unaltered: "in descriptione uisionum eius nullatenus mutilaretur" (*Vita* 31–2, 18). He argues the same for Godfrey's *libellus*: "nullam sue dispositionis patiatur iacturam" (*Vita* 16–17, 3).

14. "Scripsi enim de illa aliquid, ubi libenter ea inseruissem, si recolere potuissem," in Ep. 40 (*Epistolae* 20–2, 385).

That Silvas translates *textus*, which is the word Theoderic uses in his prologue when referring to the *Vita*, as "tapestry" instead of simply "text," is significant (Silvas 135). The *Vita* truly is an interwoven assembly of impressions, gathered from different sources. Theoderic combined 1. Godfrey's *libellus*, 2. a number of 'autobiographical,' memoir-like visions, and 3. a number of performed miracles, again interspersed with Hildegard's visions. In assembling the *Vita* according to this schema, the monk asserted to have changed very little to their contents,[13] although we can hardly take his word on this. Klaes, basing herself on a study of his style in his chronicle of Echternach, raised suspicions that some passages in the first book (Godfrey's *libellus*) betray his interventions and additions (Klaes, 92*–97*). Those passages that are intact from Theoderic's adjustments generally exhibit a more sober character, and a simpler syntax, features that might have been typical for Godfrey's writing, but of whom we know very little and possess no written documents. Klaes is somewhat more hesitant as to the possibility of Theoderic's alterations to Hildegard's texts. Theoderic seemed too intimidated by the density of her visions to dare make any profound changes to them (Klaes 111*).

Only in 1208/9, near the end of his life and some thirty years after Hildegard's death, Guibert of Gembloux acquaints himself with the *Vita* as redacted by Theoderic. Retired, at that particular time, to his former monastery of Florennes (Moens 77–79), he asks for Hildegard's parents' names in a letter exchange with Godfrey of Saint-Eucharius, because he is writing a "little something" on the *magistra's* life.[14] The need for refreshing his memory on Hildegard's biographical details hints at Guibert's renewed intentions of finishing the *Vita* left incomplete when he left Bingen, and which, indeed, makes no mention of Hildegard's parents' names. In response to Guibert's request, abbot Godfrey sends back Theoderic's *Vita*, and simultaneously solicits the former's corrections and additions, because still much is missing in Theoderic's impersonal account of the prophetess. Guibert answered Godfrey's request by stating that he found no fault in the work sent to him, and that his own fragmentary *Vita* could not possibly surpass a work of such great accomplishment. As far as we know, Guibert kept his word, and never completed a *Vita* of his own. But his reluctance to contribute to Theoderic's version, which features so strongly in his letter to Godfrey, appears to have been false modesty. A heavily stylistically altered version of Theoderic's *Vita* survives in both manuscripts of Guibert's letter collection. Guibert's interferences extend well into Hildegard's autobiographi-

cal fragments in book 2 and 3, which warrants concerns over the monk's general habit to revise Hildegard's texts. Especially small function words were Guibert's favourite target, pronouns such as *hic*, *ille* and *iste*, or comparative conjunctions such as *quemadmodum* and *uelut*, and so on (Klaes 155*). Meanwhile, Guibert also appears to have been sensitive to Theoderic's interventions in Godfrey's *libellus*, as becomes clear from his restructuring of the first chapter (Klaes 95*).

Figure 1. Schema of the composition stages of the *Vita Hildegardis*, visualizing its layered character and composite authorship. Full lines indicate extant works, dotted lines indicate lost works. Yellow indicates redaction by Theoderic of Echternach, blue by Guibert of Gembloux. The concentric circles at the core represent the original source material of the *Vita*, used by both Guibert of Gembloux in 1179/80 and Theoderic up until the mid-80s. These are supposedly Hildegard's 'original' memoirs, drafted or perhaps once transcribed by Volmar and/or Godfrey in a *libellus*. In the peripheries we see the two *Vitae* by Theoderic and by Guibert, reliant on the central sources yet independently composed. The blue layer separating the source material from Theoderic's *Vita* indicates the potential interferences made by Guibert, who collected the source material. Correspondingly, the blue layer wrapped around Theoderic's *Vita* on the outer edges indicates the revisions which are transmitted in MSS 5527–5534, and appended to Klaes's edition.

15. Newman found Hildegard's role in the early composition of the *Vita* very likely ("Three-Part Invention;" "Hildegard"). Whether or not the aim was to strive for Hildegard's canonization and the *Vita* as crowning piece of the *opera omnia* in the Riesencodex, is difficult to ascertain, and Klaes was rather hesitant about the idea (*Vita* 78*).
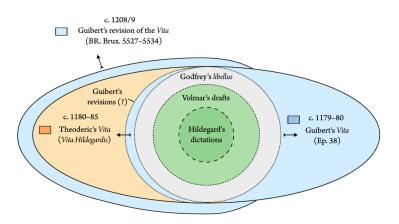
Figure 1 roughly sketches the composition history of the *Vita* as outlined in the previous section, and summarizes the (potential) zones of overlap between the different text versions. The autobiographical passages as contained within the *Vita* have especially drawn a great deal of interest in Hildegard scholarship. Their existence prior to Theoderic's integration raises a few compelling questions as to their intended form and aim. Do they indicate, for instance, that already during her life, Hildegard was consciously constructing her self-image for posterity?[15] In relation to this question, one can wonder what could have been the original connection between the individual fragments transmitted in the *Vita* as we have it, and whether or not they derive from an originally integral 'autobiographical' *Vita*. Interestingly, even Godfrey's *libellus*, a third-person account of Hildegard's life, has been hypothesized by Newman to originally have been a first-person attestation by Hildegard, a true memoir, which Godfrey then rewrote from a different focalization point (Newman, "Hildegard" 17–18). The question is if that would mean that Hildegard's original style is to a certain extent recoverable from Godfrey's transcription. On the other hand, we have the interference of a considerable number of male co-writers that had ample opportunity to rewrite and overwrite Hildegard's original text. Godfrey and Volmar present the first filters through which her signal passed. Guibert of Gembloux, then, who we know was capable of altering Hildegard's style and did not flinch from doing so

(Kestemont, Moens and Deploige), could have revised the material whilst collecting it, after which Theoderic of Echternach selected portions from it and possibly again revised all materials according to his own principles. The aim of the subsequent paragraphs is to shed more light on the extent to which these collaborators' treatment of Hildegard's text could have included the compromising of her language. More generally, our findings consequently invite reflection on Hildegard's authorship, and on the extent to which her involvement in the *Vita*'s composition is reflected stylistically in the text.

4 Computational Stylistics

The core idea of computational stylistics, as elaborately developed in the introduction to this special cluster of *Interfaces*, is that authors betray individual writing patterns which largely escape their conscious control and are therefore not easily imitated. Often these patterns lurk in the frequencies of 'stylistic features,' such as short marker words, particles, or parts of words (*n*-grams). By implication, this means that every author has a so-called 'stylome,' (van Halteren) or attests to a stylistic DNA which can be harvested with statistical analysis. That the method works well for Latin, and for medieval Latin specifically, has been demonstrated by an increasing number of scholars in the past years (Kestemont, Moens and Deploige; Eder, "*Chronica Polonorum*;" Vainio *et al.*; Downey *et al.*; De Gussem). Particularly popular are function words, grammatical or syntactical markers which are used frequently in natural languages but do not have a strong semantic meaning to their users, such as conjunctions, pronouns, prepositions, adverbs and particles. From a practical point of view, function words are useful thanks to their high frequency, which offers a distribution which can be analysed and harvested on a statistical basis. From a theoretical point of view they provide alluring prospects because they are argued to be applied unconsciously and escape the attention of both writer and reader. They are like stylistic fingerprints that are difficult to imitate, and are taken to be relatively constant in frequency across different genres of texts written by one and the same author.[16]

Computational stylisticians often emphasize the advantages of 'distant reading' as opposed to 'close reading' (Moretti) when it comes to scope –analysing texts by the dozens – and circumventing researchers' subjectivity. It is argued that in doing so stylometrists offer an objective and uncompromised viewpoint on which autho-

16. Recent scholarship on computational authorship attribution has shown that a text's topic or genre has a far less explicit impact on frequency of function words than individual authorship. The latter could be argued to be the predominant, primary signal it captures, and the secondary signal of genre or topic does not necessarily jeopardise the validity of that other signal or vice versa. For the 'genre effect,' see Jockers 63–104 (especially 80–81); for a kind of 'genre effect' at work in Hildegard of Bingen's oeuvre specifically, arguably demonstrating her diachronic, intellectual development, see De Gussem and Wouters, 31–60.

rial profile provides the best match for an anonymous or falsely ascribed text. The objections one may raise to objectivity claims such as these are numerous, and I will not repeat all of them here. For now it serves to emphasise that computational stylistics offers an additional tool –aside from those traditionally familiar to medievalists, such as palaeography, stemmatology, stylistics, etc. – for attempting to identify the provenance of historical documents and formulate assessments as to their authorship on the basis of statistical observations.

Before the computational analysis of style can take place, a document's word order and symbolic appearance are abandoned, as the text is encoded into a representation of its contents in terms of numbers. These lists of numbers –called 'vectors' in data analysis – summarize relevant information of the text's lexical properties, disregarding its former orderly principle or linguistic logic. Consider this intuitive example:

Table 2. Intuition of bag of words and vectorisation — 1.

vocabulary	*sua*	*nos*	*honestum*	*est*	*quod*	*vi*	*trahit*	*et*	*dignitate*	*allicit*
frequencies	2	2	1	1	1	1	1	1	1	1

Honestum est quod sua vi nos trahit et sua dignitate nos allicit.

The array of numbers in row 2 of this table would constitute a vector, a series of frequencies of encountered words or 'tokens' that summarises the contents of a short sentence such as the one above. Note that the tokens *sua* and *nos*, which are our so-called 'features,' occur twice in the sentence but only once in the vocabulary (row 1). In practice, computational stylistics performs exactly the same routine for documents that are much longer than a single sentence, wherefore the vocabulary and frequencies will considerably expand, returning arrays such as the following:

Table 3. Intuition of bag of words and vectorization — 2.

	est	*et*	*in*	*enim*	*non*	*autem*	*quod*	*que*	*cum*	*ad*		...
text A	30	29	13	9	8	7	11	10	3	4		...
text B	35	23	22	14	15	11	6	7	14	8		...

Once documents are translated to vector representations, such as A and B in the example above, they become comparable on a numerical basis, where their 'difference' or 'distance' can be measured quite intuitively. Especially the top frequency strata of the texts' vectors are significant. Aside from a number of recurrent keywords, this top stratum contains function words such as *et*, *in*, *enim*, *non* and *autem*, mentioned above, which have, ever since Mosteller and Wallace's fundamental study of 1964 (Mosteller and Wallace), been shown to be extremely effective for establishing a text's authorship. If so desired, one may opt to leave out all semantically charged content words and analyse texts by the function word only.

Once texts have been converted into numerical data, and are no longer human-readable, there are different kinds of statistical techniques by which to analyse the gathered frequencies and look for authorial patterns on a very large –or distant – scale. To this end I will apply a number of state-of-the-art methods for computational analysis of style in this article, such as Principal Components Analysis (PCA), Support Vector Machine (SVM) and the *impostors* method, which I will explain in more detail throughout.

5 Candidates, Corpus and Preprocessing

Assessing the authorship(s) of the *Vita* is a complex matter, because not all of the candidates have left independent writings of their own which facilitate a direct basis of comparison with a sample from the *Vita*. This is true for secretaries Volmar and Godfrey of Disibodenberg, or other potentially involved assistants whose contributions we lack evidence of. As for the remaining authors for whom we do have the ability to assemble a background corpus – Hildegard, Guibert and Theoderic –, issues of reliability and incompatibility with the *Vita*'s genre and style are at stake. The background corpus is therefore small and comes with its own insecurities. Guibert of Gembloux's *Epistolae* (± 124,500 words) and his very short *De combustione* (± 1,000 words) were included (Derolez; Pertz), as were Theoderic of Echternach's chronicles (Weiland). For Hildegard of Bingen, her *Vitae* of Saint Disibod (± 7,500 words) and Saint Rupert (± 4,200 words) were integrated in the corpus (edited by Evans), which only seemed reasonable considering that these works best represent her handling of the hagio- and biographical genre to which the *Vita Hildegardis* belongs. Nevertheless, these texts' brevity required the involvement of her visionary treatises as well: *Scivias*, *Liber vitae meritorum* and *Liber divinorum operum*. It is on the basis of these works that a training corpus was assembled that could best represent the stylistic profiles of these three important candidates. Hildegard's texts, which should provide the 'gold standard' of her style, can obviously not be strictly separated from the potential influences of Volmar. Theoderic's chronicles, the *Chronicon Epternacense* and the *Libellus de libertate Epternacensi propugnata*, contain genre-specific qualities that might destabilize a firm basis for comparison. The texts and the Python programming code[17] that derive from the analysis in this article are openly accessible on GitHub,[18] but some of the texts in the repository have been camouflaged so as to respect the copyright laws protecting the editions.[19]

17. Python is a programming language popularly used for Natural Language Processing-related tasks such as stylometry. Naturally, it is also possible to replicate the experiments in this article with more user-friendly stylometric toolkits, such as Stylometry with R (Eder, Rybicki and Kestemont).

18. GitHub is a collaborative platform specifically designed for developers, programmers and researchers to share, replicate and contribute to data in online repositories.

19. GitHub. Only function words – which are highly successful for distinguishing writing styles – were retained in their original position and form. All the remaining, content-loaded tokens were substituted by asterisks, rendering the text illegible. This means that some experiments in the current article which relied on content words in addition to function words will not be replicable by relying solely on the text data as available on GitHub. To replicate these experiments as well, one may request access to the electronic versions of the editions referred to by contacting the publishers in charge.

Before handling medieval Latin texts, stylometrists need to ascertain that the data is comparable. Prior to analysis the Latin texts have to be transformed to adopt the same formatting norms so as to allow a reliable basis for comparison. This entails some minor interventions in the texts which commonly fall under the header of 'preprocessing,' where irrelevant material is removed and the divergent orthographical forms due to various manuscript witnesses or editorial practices are normalized, such as such as <j>'s to <i>'s, <v>'s to <u>'s, <ae>'s to <e>'s, etc. For the *Vita Hildegardis* specifically, it should be noted that the *Capitula* after each prologue were removed.

6 Experimental Set-Up: Rolling SVM-*Impostors*

Having taken into consideration the training corpus at our disposal and this case's particular challenges, the authorship problem of the *Vita Hildegardis* does not, strictly speaking, allow for a traditional, 'closed' attribution set-up, in which the most suitable author is chosen from among a set of candidates. Two challenges arise.

1. The authorship(s) of the *Vita Hildegardis* presents a layered and complex attribution problem of dual and/or mixed authorship, which requires a fine-grained means to assess style changes over short passages that distinguish between Hildegard and her biographers, but also indicates transitional stages in the text or passages that are the result of collaborations of multiple authors.
2. Not all potential contributors to the text (Volmar and Godfrey, to name but two) have a reference corpus. In some cases, we will want to assess if the stylistic profile of a particular passage in the *Vita* can even be matched to one of our candidates at all. In stylometric practice, assessing if the latter condition holds true or false is called 'authorship verification': we want to know if our candidate is simply not included in our corpus.

When it comes to the first challenge, the concept of sampling briefly needs to be addressed. In computational stylistics it is considered good practice to slice up the text into smaller segments or 'text samples.' Such samples need to be short enough in word count so as to allow for a fine-grained comparative analysis with segments from both other- and same-authored works, but simultaneously lengthy enough so that a passage's stylistic composition can be captured re-

liably. The current state of the art for computational stylistics in Lat-
in has come to a consensus that the minimal working length for sam-
ples is 2,500 to 3,000 word tokens (Kestemont, Moens and Deploige;
Eder, "Does Size Matter"). Any experiment carried out on document
segments beneath that length are considered less secure, although
some methods have been demonstrated to be sufficiently accurate at
sample lengths as short as only 500 words (Koppel and Winter 178).
There are, however, two kinds of sampling: 'discrete' and 'rolling'
sampling (illustrated in Figure 2). 'Discrete' sampling entails all the
steps explained in the previous paragraph, where the text is sampled
in discrete chunks according to the analyst's preset ranges. Rolling sam-
pling methods, on the other hand, have the original text sequentially
sliced up into non-identical, partially overlapping windows according
to a step size. The advantage is that the original sample size can be re-
tained, that more information is gathered as to how the text develops
sequentially, and that higher peaks of 'unexpected' stylistic patterns
can be localized. One can think of this as taking up a magnifying glass,
scanning the text linearly, and registering how it changes from the very
first to the very last word. In the experiments below, we will slide over
the text by processing it 500 words at a time, and gradually proceeding
onto the next sample by a step size of 100 words.

Figure 2. Intuition of discrete sampling vs. a rolling sampling method. The original text is sequentially sliced up into overlapping windows according to a step size.

To tackle the second challenge, an 'open' authorship verification
technique ("is the candidate author amongst the candidates – yes or
no?") called the *impostors* method (Koppel and Winter) is used in
tandem with a 'closed' attribution technique called SVM classifica-
tion (Diederich *et al.*): "amongst a closed set of candidates – which
candidate yields the best match?"

  Both SVM and the *impostors* method are derived from state-of-the-
art text classification techniques that are heavily indebted to the (still on-
going) rise of machine learning (ML) in computer science. ML here cor-
responds to a set of smart computer algorithms that are better equipped
to exhaustively search many different parameters and evaluate each of
these combinations' efficiency for detecting stylistic similarities or solv-
ing a particular text classification problem such as attribution.

The difference between classification and verification is the following. A classifier such as SVM is a learning algorithm that takes securely attributed texts as 'training data,' learns to tell apart the stylistic patterns particular to each of the involved authors (Hildegard, Guibert and Theoderic), and departs from this basis to make predictions on insecurely attributed texts or 'test data,' here: the *Vita Hildegardis*. Verification methods, on the other hand, such as the *impostors* method, defend themselves from coincidental attributions by maintaining that even though two documents might well attain a degree of similarity in some representation or another, security can only be established if the match is validated for k iterations (usually k=100). During these iterations, the algorithm changes the documents' vector representations by randomly selecting 50% of the original feature set, and by introducing so-called 'impostors' from a large background dataset of texts (listed in the Appendix, Table 7). The proportion of times that some candidate is the top match is consequently tested against a benchmark threshold σ^*, which, if not surpassed, lets the impostors algorithm output the category label 'None of the above.' Anything below that threshold outputs 'Uncertain.'

An advantage of both SVM and the *impostors* method is that they have a learning phase, in which their effectiveness in segregating authorial styles can be evaluated and expressed in percentages. This entails that different parameters are trained, tested and evaluated, and that the best-scoring model, considered most apt for segregating 'classes' (here: our authors), is applied to the problem.

20. Tfidf stands for 'term frequency inverse document frequency.' It divides all feature values by the number of documents that respective feature appears in. As a consequence, less common features receive a higher weight, which prevents them from sinking away (and losing statistical significance) amidst more common features.
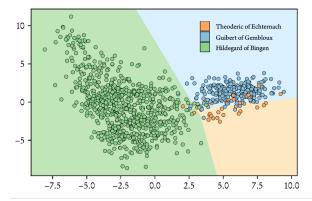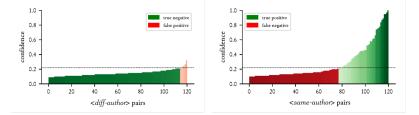
Figure 3. PCA plot giving a two-dimensional intuition of the decision boundaries drawn between Theoderic, Guibert and Hildegard by the best-scoring SVM classifier. Each of the coloured dots in the plot is a simplified representation of the vector information of an author's text segment projected in a two-dimensional space, where the PCs demonstrate the greatest variance in the data and help to inspect the most conspicuous distances (or differences) between the candidates' respective samples. When samples cluster together, they contain resembling writing patterns. Settings: 500 most-frequent words. Sample length: 500 word tokens. Standard-scaled tfidf-weighted[20] raw frequencies. Explained variance is 5.20%.

Figure 3 shows, in an intuitive PCA plot (Binongo and Smith), how an SVM classifier learns to demarcate the clusters of pre-labelled text samples, and draws a decision boundary that reflects which regions ('vector spaces') associate with the stylistic behaviour of candidate authors Hildegard, Theoderic and Guibert. Once the model has learned these boundaries, it can make predictions on a test sample of unknown authorship, for instance a passage in the *Vita Hildegardis*. Note that Theoderic and Guibert's works appear somewhat tricky to distinguish. Bringing in an additional, third component, as we will see in PCA plots further down this article (figure 6), helped to make more nuanced distinctions between both authors, who apparently have quite a few stylistic aspects in common.

Similarly, the *impostors* method goes through a learning phase, in which the threshold σ* is established on the basis of a pre-labelled training corpus. The *impostors* method relies on learned instances of known labelled pairs of <*same-author*> or <*different-author*> as training data. For both types of classes we ultimately collect percentages indicating how many times out of k=100 times correct and false attributions were made. This mean percentage can consequently function as a learned threshold. The *impostors* method can be a very powerful and meticulous method when it trains well on the authors under scrutiny. If a well-balanced σ* can be established, this grants great confidence to any test attribution following it, especially if this attribution has a firm and high confidence score. On the other hand, authorship verification remains an extremely difficult and unsolved problem, exploring the limits of what is feasible in the current-day landscape of computational stylistics.

Figure 4. Intuition of *impostors* method thresholding. The y-axis indicates the confidence by which the attribution was made, whereas the x-axis indicates the sample index number (there were 120 samples in total). Note that the number of paired samples per class (<*same-author*> and <*diff-author*>) is equal. The trained threshold σ* (= 0.22) is visualised by the horizontal, dotted line. Its purpose is to avoid as many false decisions as possible (red) in favour of correct ones (green).

The σ* threshold of 0.22 was in fact quite stable for distinguishing between Hildegard, Guibert and Theoderic, yielding quite promising figures despite the short sample length (table 3).

Table 4. Evaluation metrics (accuracy, precision, recall and f1) after training the *impostors* method on Theoderic of Echternach, Guibert of Gembloux and Hildegard of Bingen, at the best-performing threshold of 0.22. Typically, *impostors* method thresholds have a high precision and low recall (fewer acceptances and fewer mistakes).

	dev set	test set
accuracy	0.66	0.62
precision	0.88	0.89
recall	0.39	0.27
f1	0.54	0.41

Figure 5. Rolling SVM-impostors method for the *Vita Hildegardis*. The x-axis shows the progression of the sliding windows throughout the *Vita* by taking a step of 100 words at a time. The corpus development line at the very bottom of the figure indicates by the chapter numbers in Klaes's edition which part of the *Vita* is treated. In total, attributions were made for 140 partially overlapping text samples consisting of 500 words. Bars above the x-axis indicate the prediction of the impostors method, bars below the x-axis indicate the prediction of the majority of SVM classifiers. They y-axis (the height of the bars) indicates the confidence score (between 0 and 1) for both the impostors method and the SVM classifiers (longer bars with higher colour intensity indicating more confident attributions). The indexes 1–15 given below the figure are referenced in table 4.

7 Results

The results of performing the rolling SVM-*impostors* method on the *Vita* are given in figure 5. The x-axis shows the gradual progression of the sliding windows throughout the *Vita* by taking a step of 100 words at a time. The 'Corpus development' line at the very bottom of the figure indicates by chapter number in Klaes's edition which part of the *Vita* is treated per sample. Bars above the x-axis indicate the prediction of the *impostors* method, bars below the x-axis indicate the prediction of the majority of SVM classifiers. The y-axis (the height of the bars) indicates the confidence score (between 0 and 1) for both the *impostors* method and the SVM classifiers (longer bars with higher colour intensity indicating more confident attributions).

The indexes 1–15 indicated below the figure are referenced in table 4. They mainly correspond (with a few exceptions discussed below) to Hildegard's autobiographical fragments, of special interest to us here. Immediately, it appears that despite the short sample length the combined method (*impostors*-SVM) recognizes the eight visions and other passages in which Theoderic cites Hildegard as strongly Hildegardian. The *impostors* method is –as was to be expected –

Table 5. Contents of fragments of the *Vita Hildegardis* indexed in figure 5, complete with a description of their contents, incipit, and reference to the edition.

Index	Fragment	Incipit	Ed. (Klaes, *Vita*)
1	Letter to Guibert	"Deus inquit ubi …"	§1.8–9, 14–15.
2	*Prima visio*	"In mystica inquit …"	§2.2, 21–24.
3	*Secunda visio*	"Quodam inquit tempore …"	§2.5, 27–30.
4	*Visio tertia*	"Vidi in visione …"	§2.7, 31–32.
5	*Visio quarta*	"In lectum egritudinis …"	§2.9, 33–35.
6	The philosopher	"Quidam phylosofus de …"	§2.12, 37–38.
7	*Visio quinta*	"In vera inquit …"	§2.14, 38–41.
8	*Visio sexta*	"Tres turres in …"	§2.15, 42–43.
9	*Visio septima*	"Subsequenti demum tempore"	§2.16, 43–44.
10	Sigewize	"Posteaquam me visio …"	§3.20, 56–57.
11	Letter to Gedolphus	"G. ecclesie Brunwilarensis …"	§3.21, 60–62.
12	Possessed woman in the Rupertsberg	"De adventu inquit …"	§3.22, 64–65.
13	Account of her illness	"Post hec inquit …"	§3.23, 66.
14	*Visio octava*	"Pulcherrimus inquit et …"	§3.24, 67–68.
15	*Mulierem inquiunt* and *Cum beata*	"Mulierem inquiunt" | "Cum beata"	§3.26–27, 68–70.

somewhat more severe in its prediction. When thrown in an 'open setting' (*impostors* method) the autobiographical samples are struggling far more to beat the competition by authors from the benchmark corpus. In a 'closed setting' (SVM classifier) they univocally adhere to Hildegard's style. 78 out of 140 samples assign Hildegard as a candidate for the *Vita*'s authorship. This is an extensive and con-

vincing contribution to the text as a whole, and may counter scepticism on the authenticity of these passages because of their similarity to Hildegard's canonical works, which we presume to be the most reliable specimen of what constitutes Hildegard's style. In other words, the *Vita* may rightfully be designated 'autobiographical,' and Guibert's and Theoderic's influences –to which we will return below – remain limited. A fascinating instance of an extremely small (!) portion of the *Vita* similar to Hildegard's style appears at the close of book 3's series of *miracula*. It concerns two letters, indexed as 14 and 15 in figure 5, written by Hildegard's sisters:

> His - prout possibilitas ingenioli suppetebat - a nobis digestis calamum ad uerba sanctarum filiarum eius uertamus, et que de ipsa memoratu digna scripserunt, maxime de beato transitu eius, sicut uiderunt et audierunt et manibus suis tractauerunt, adiuuante Domino fideliter et ueraciter huic operi annectamus. (*Vita* §3.26, 10–4, 68)

> Now that we have edited everything as far as the capacity of our limited talent allows, let us turn our pen to the words of her holy daughters, who have written worthily of her memory. With the help of the Lord let us append to this work faithfully and truthfully what they saw and heard, especially concerning her blessed passage from this life, which they have written down with their own hands. (*Life of Hildegard* 208)

The first account of the sisters is an anecdotal and concise summary of miraculous deeds performed by Hildegard (*Mulierem inquiunt*). The second is referenced in Table 4 as *Cum beata*, and is preoccupied with Hildegard's illness and her death at the age of eighty-two, which is portended by the apparition of a glowing red cross at the firmament. After these two passages, at the very ending of the *Vita*, an additional, short, unintroduced passage on Hildegard's burial occurs, for which Theoderic mentions no source in the text. It recounts the miraculous benefits that visitors had gathered from venerating Hildegard's grave. In the course of these final passages, the rolling SVM-*impostors* algorithm signalizes a lot of 'Hildegardian' material in the language. Evidently this poses a problem, as the passages include a description of events not only before but also after the author's death. Also on closer inspection of *Mulierem inquiunt* and *Cum beata*, one gains a strong impression that Hildegardian language is present. In

21. Levenshtein distance is a very simple operation for measuring the difference between two string sequences. A low Levenshtein distance means a close match between two word groups or sentences.

22. As one perceptive (but regrettably anonymous) peer reviewer has remarked, some of the matches are strictly speaking not 'Hildegardian,' but rely on an authoritative corpus of biblical-patristic tradition. I consent to this, and in fact I am inclined to extend the same argument to Hildegard's larger oeuvre and to the majority of medieval authors operating in similar contexts. Nevertheless, the preliminary experiments in this paper have clearly indicated that medieval authors' engagement with a reservoir of Latin *auctoritates* does not jeopardise the observation that they dispose of an individual style as well. I am therefore inclined to add that Hildegard's proper selection of *topoi* available to her –if I may call them such – is exactly part of the fabric of her style. Indeed, the decisions she makes as to what she includes and excludes is part of her stylistic profile as well. However, I admit that the field of stylometry would be better off if the impact of intertextuality on medieval Latin literary style was more systematically assessed — a considerable task and challenge.

23. "But God showed clearly in her passing what standing she had before him" (*Life* 209).

24. "Two men who made bold to touch her holy body recovered from a severe illness" (*Life* 210).

25. "Exequiis igitur uenerabiliter a reuerendis uiris celebratis in uenerando loco est sepulta, ubi meritis eius omnibus pio corde querentibus prestantur beneficia multa" (*Vita*, §3.27, 39–41, 70–71).

the majority of turns of phrase one finds her preferred syntactic constructions and (occasionally biblical and patristic) imagery, especially of *Scivias* and the *Liber divinorum operum*. I have appended a more detailed study of corresponding passages in tables 5 and 6 in the Appendix, which were automatically searched by using Levenshtein distance.[21] A large number of sentences has parallels with passages in Hildegard's writings, especially the description of Hildegard's death in *Cum beata*, in which the red cross illuminates the sky into a colourful and dizzying spectacle.[22]

We might want to pause briefly at what is happening here. One should not forget that Hildegard's *Vita* is classified as an autohagiography, a genre heavily based upon literary precedents. This text, which can be held to inflate or even distort factual reality from a modern point of view, would be considered 'true' by virtue of its ethical-exemplary function by a medieval audience (Greenspan). Hildegard's death, – and the events leading up to it – would have, to a large extent, been pre-written according to the rules of the genre. The depiction of a saint's death was a literary *topos*, invoked with a specific purpose: the ultimate authentication of the saint's holiness. The conventional nature of death passages in female saints' hagiographical literature is an important point emphasized by Garay and Jeay in their recent piece exploring the "stages and staging of holy women's death" ("Sanctification" 139) By discussing the death passages of female mystics such as Elisabeth of Schönau, Douceline of Digne († 1274), Marie of Oignies († 1213) and Lutgard of Aywiéres († 1246), they stress that "death is the moment which epitomizes the heroic life of women who have been chosen for the vocation of sainthood" (ibid.). Many of the aspects Garay and Jeay attend to in order to expose the constructed nature of these death passages may well be shown to apply to Hildegard's *Mulierem inquiunt* and *Cum beata* as well. One is, for instance, the divine endorsement crucial to legitimating the saintly status: "Deus uero, cuius meriti apud se esset in transitu suo euidenter declarauit" (*Vita* §3.27, 16–17, 70).[23] Another is Hildegard's performance of "posthumous appearances and miracles" (Garay and Jean 139): "Nam duo homines, qui sanctum corpus eius spe bona tangere presumpserunt, a graui infirmitate conualuerunt" (*Vita* §3.27, 37–39, 70).[24] Thirdly, the ending of the *Vita* allocates a large role to the participation of Hildegard's community. The posthumous miracles lead up to her enshrinement "in a venerable place," which draws pilgrims for its "many benefits [...] available to all who come seeking them with devout heart."[25] All of

these elements make the depiction of Hildegard's death symbolically coincide with a wider involvement of the members of her community, for whom the cultivation of her person and the tradition she had founded becomes paramount. The death passage was, in other words, the apogee of the narrative, with a lasting importance for Hildegard's remembrance and perhaps even for the economical survival of the Rupertsberg. It was, in Dalarun's interpretation of Max Weber, "the transition of her personal charisma to a durable institution" ("La mort" 194).[26]

26. "Ce qui se joue en effet au moment précis du transitus du saint fondateur, c'est non seulement le passage attendu de l'ici-bas à l'au-delà, c'est aussi le passage périlleux d'un charisme personnel à une institution faite pour durer, d'un idéal toujours et oujours plus idéalisé par l'hagiographie à une pratique quotidienne, à une nécessaire insertion dans l'Église et dans la société" (Dalarun 194).

Taking this all into account, there is evidently more than one hypothesis which could account for why Hildegardian language appears here. The boldest one is to believe that Hildegard described the miracles in *Mulierem inquiunt* herself, and prophesied on the events of her death in *Cum beata* and ultimately arranged for the texts to be incorporated in her autohagiographical *Vita*, all of which occurred under her own authority and by her own hand. One may invoke one or two reasons in this hypothesis's defence. The *Vita* portrays Hildegard as prescient of the conditions by which she was to die, and, most importantly, as portending this course of events to her fellow sisters – to whom Theoderic emphatically attributes the authorship of the current passages.[27] The 'fabricating' of death stories has precedents in the twelfth century. One could think, for instance, of Geoffrey of Auxerre's death letter of Bernard of Clairvaux, composed in order to recuperate the saint's authority and authorize Arnaud of Bonneval's contribution to Bernard's *Vita* (Bredero, "Der Brief").

27. The Saint's prescience on his or her death is, however, an often encountered hagiographical trope (Boglioni 189). The original passage in Latin says "spiritu prophetie ei reuelauit, quem et sororibus predixit (*Vita* §3.27, 7–8, 69). Theoderic's abridged version states "Hec obitum suum longe ante presciens et sororibus predicens," (*Octo lectiones* §8, 15–16, 79); and Guibert's revised version – with his words indicated between angle brackets – says: "Cum beata inquiunt mater <regi et dominatori omnium> multis laborum <et dolorum> certaminibus deuote militasset, presentis uite tedio affecta <ad gaudia summe beatitudinis anhelans> dissolui et esse cum Christo cottidie cupiebat. <Quapropter> Deus <hanc dilectam suam a bono desiderio suo fraudari nolens diutius> finem <mortalis uite, quem ad ipsum suspirando optauerat,> spiritu prophetie ei reuelauit, quem et <filiabus suis in breui futurum esse sepe dixit>" (*Vita retractata* §3.26, 1–8, 106).

However, it is a curious theory, and without proved precedent in the hagiographical genre, to believe that Hildegard deliberately sat down to write about her own death (amongst other matters), with her secretaries as accomplices to what can arguably be called a very bizarre undertaking indeed. A more acceptable hypothesis is that the algorithm picks up on the fact that Hildegard's fellow sisters were trained extensively to imitate their *magistra*, and made an express effort to conjure up her style and tone in a passage with such great symbolic significance. Hildegard's words reverberate almost literally (again, I refer the reader to the Appendix where tables 5 and 6 show the correspondences). Another hypothesis could be that the passage is a revision of authentic Hildegardian materials, recycled to an extent sufficient enough to fool the *impostors* method. By principles similar to those of end redactor Theoderic of Echternach, her sisters loosely collected some of Hildegard's remaining

writings after her death in the Rupertsberg scriptorium and cobbled them together. Considering how the *Vita's* composition process was one of recuperating materials that coincidentally happened to be at the composers' disposal, this may well be feasible.

In the spirit of gradually moving from one inference to the next, we may make the assumption – based on the rolling SVM-*impostors* method above – that the autobiographical passages are genuinely Hildegard's. This is further confirmed in the PCA plot in figure 6, where the original *Vita Hildegardis* was divided into two distinct batches: Hildegard's visions versus all non-Hildegardian fragments of the text. The division into batches also allowed this additional verification to work with 1,000-word instead of 500-word samples. Again, Hildegard's autobiographical passages can patently be shown to be Hildegard's (red triangles), and are clearly distinguishable from remaining samples of the *Vita*, namely Theoderic's commentaries (dark gray) and Godfrey's *libellus* (purple). The behaviour of these remaining samples of the *Vita*, traditionally believed to have been the work of Godfrey of Disibodenberg and Theoderic of Echternach, prove far more difficult to categorize. If we revisit the predictions of the rolling SVM-*impostors* method earlier (figure 5), and combine them with the PCA plot in figure 6, the following indications are given:

1. The PCA plot in figure 6 has Theoderic's commentaries (gray) and Godfrey's *libellus* cluster predominantly on the right end of the figure, alongside the works of Guibert of Gembloux (coloured blue).
2. In a closed setting (SVM in figure 5), these samples – be it hesitantly – sympathize with Guibert as well.
3. The *impostors* method (figure 5), on the other hand, refuses to become very confident, and makes few to almost no attributions to either Theoderic or Guibert which surpasses the σ* threshold.
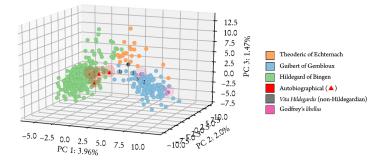
Figure 6. PCA plot containing Hildegard's, Theoderic's and Guibert's training texts (green, orange, blue). Theoderic's commentaries in the *Vita* are coloured gray, and come annotated with sample indices indicating a rough order of appearance (1–5). Also included in separate colouring are Godfrey of Disibodenberg's *libellus* or book 1 of the *Vita* (purple), and the autobiographical passages of Hildegard (red triangles), giving further evidence in support of figure 5 that Hildegard's visions are fully Hildegardian. Settings: 500 most-frequent function words. Sample length: 1,000 word tokens. Standard-scaled tfidf-weighted[28] raw frequencies. Explained variance is 7.43%.

28. See above, p. 140.

That Theoderic's commentaries fail to cluster with any of his chronicles is particular, and calls for some additional analysis. The three PCA plots given in figure 6 leave out Hildegard's works, and benchmark test documents Godfrey's libellus and Theoderic's commentaries against exclusively Theoderic's and Guibert's training texts (both individually and together). Here again, one gains the impression that the remaining samples of the *Vita* are inbetweeners, with a more explicit preference to side with Guibert, be it never quite convincing. Klaes's suspicions that chapters §1.8–9 of the *Vita* (containing the letter to Guibert) testify more to Theoderic's style, is not confirmed.[29]

29. These chapters are collected under sample 2 (purple). For Klaes's suggestions of Theoderic's authorship, see her introduction to the *Vita* at 92*–94*.

Neither of these candidate authors are very convincing, and it turns out that Guibert is systematically the best guess, if guessing is at all allowed in this scenario. Even Theoderic's first prologue, in which he explicitly announces his presence and informs his readers which source materials were used (without mentioning Guibert), turns out to be more like Guibert than like Theoderic. Guibert's (quite extensive?) stylistic influence on the *Vita* as we have it is problematic, for it does not appear compatible with the commonly accepted timeline of the *Vita*'s composition.

The prologues are important in establishing the chronology, for they give firm evidence of the current *Vita*'s completion by Theoderic, at a time definitely after Hildegard's death († 1179). It has com-

monly been assumed that Theoderic and Guibert just missed each other at the Rupertsberg. Theoderic arrived in the early 1180s shortly after Guibert departed for Gembloux. Consequently the two biographers are thought to have been unaware of each others' *Vitae* until Guibert coincidentally discovered Theoderic's in corresponding with Godfrey of Saint-Eucharius (all of which was already explained above at 8), after which he revised it c. 1208/9 before including it in his own epistolarium. In their current form it must stand beyond doubt that the *Vita*'s prologues' *terminus a quo* is 1181, when Godfrey became abbot of Saint-Eucharius, and their *terminus ante quem* before Ludwig of Saint-Eucharius passed away, in 1187.[30]

In other words, that Guibert's presence is suggested even in those passages of the text which have always been thought to have been exclusively Theoderic's additions is problematic. I see two (maybe three) plausible hypotheses for explaining it, but hard historical evidence for either of them is lacking. Either the last redactor of the *Vita* was Guibert instead of Theoderic, or else Theoderic's reliance upon Guibert's source materials is far more extensive than has hitherto been presumed. Before discussing the pros and cons of either of them somewhat more extensively, it should be noted that both hypotheses are weakened by Guibert's seemingly genuine surprise upon learning in 1208/9 from abbot Godfrey that there was an extant *Vita* of Hildegard. If the *Vita* sent to him by Godfrey had been a work largely reliant upon his own text, then Guibert shows no sign of indignation or familiarity, no reaction at all really. Instead, Guibert responds to Godfrey's request for corrections by remarking that "I have nothing in memory to infer or add, nor can I find anything superfluous which I would remove, nor anything ineptly placed that I would correct."[31] It is peculiar to believe Guibert is talking about a *Vita* he had at hand himself, unless we assume that these lines are intended to be tongue-in-cheek or slightly smug, or that he no longer recognized his work after some thirty years, or that there were reasons for him to conceal his former contribution to this older *Vita*.

30. The prologues testify of Godfrey and Ludwig's simultaneous abbacy (Silvas 121).

31. "[...] Non habens pre memoria quid inferrem uel adderem, nec inueniens in ea quicquam superfluum quod demerem, neque aliquid inepte positum quod corrigerem [...]" (*Ep.* 42, 117–19, 394). My translation.
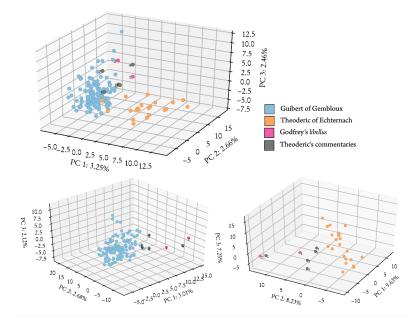
Figure 7. PCA plots of Theoderic's commentaries (gray) and Godfrey of Disibodenberg's *libellus* or book 1 of the *Vita* (purple). Theoderic's commentaries and the *libellus* are annotated with sample indices (1–5 and 1–2), indicating a rough order of appearance. Settings: 500 most-frequent function words. Sample length: 1,000 word tokens. Standard-scaled tfidf-weighted raw frequencies. Explained variances given alongside axes of subplots.

Let us explore the first hypothesis' presumptions and its merits. It presupposes that some time after Theoderic's completion, Guibert saw a chance to extensively revise an earlier version of the *Vita* by the former's hand, which is now lost. This poses numerous problems. Firstly, the manuscript on which Klaes's edition is based has firmly been retraced to Echternach, and has on palaeographical grounds been shown to contain handwriting similar to that of Theoderic. In other words, our best manuscript of the *Vita* is an autograph by Theoderic.[32] One could always assume that Guibert visited the Rupertsberg while Theoderic was working on the *Vita* in the early 1180s, but this seems unlikely considering Guibert's recent departure and busy schedule, including a pilgrimage to Tours (Moens 76). In his letter to abbess Ida, written around 1185, Guibert seems to be apologising for his longstanding silence toward the Rupertsberg community after his departure, which he defends by stating that he had been victim of false accusations and jealousy toward him after his stay there.[33] Considering Guibert's close involvement in collecting the source materials, one might wonder why neither Ida nor any other Rupertsberg sister felt it necessary to inform Guibert on a new *Vita* in the works, or send it to him if it happened to be finished around that

32. Wien, ÖNB, 624. "Die eigentliche Provenienz des Kodex ist aber das Kloster Echternach, wo er vom Autor der *Vita S. Hildegardis*, dem Echternacher Mönch Theoderich, selbst aufgezeichnet wurde" (Klaes 158*).

33. See Guibert's Ep. 32 (100–113, 336), the first of a longer letter exchange with the Rupertsberg (Epp. 32–37, 333–65).

time. Aside from these problems, the hypothesis that the *Vita* as we have it has known revisions by Guibert that postdate Theoderic's version becomes difficult in light of the fact that we already have a revision by Guibert, the *Vita sanctae Hildegardis retractata*. This would lead to the conclusion that Theoderic's *Vita* contains Guibert's first revision, and that Guibert's *Vita retractata* is the revision of the revision.

The second hypothesis holds that we have underestimated the degree to which also Theoderic's interbeddings are heavily indebted to the text prepared by Guibert between Hildegard's death and the latter's arrival at the Rupertsberg (1179/8). Guibert had been in close contact with Hildegard from 1175 onward and had become her closest secretary in 1177. Being closely involved in the composition of her epistolarium and the completion of the Riesencodex, the *Vita* which Theoderic found upon arrival might have looked very similar to the one lying before us today. After all, who else would have found it more necessary to extend Godfrey's *libellus* with Hildegard's letter to Guibert, the *De modo visionis sue*, than Guibert himself (indexed as 1 in figure 5)? Also in his fragmentary *Vita* sent to Bovo (Ep. 38 of his letter collection), Guibert included this letter from Hildegard, which shows the importance he attached to it. Following this train of thought, we may assume that Theoderic made subtle stylistic amendments, perhaps inserted references to Echternach's well-known abbot Thiofrid (Klaes 84*), but in reality heavily relied –including even large parts of the three prologues to the individual books – on an architecture formerly constructed by Guibert (and, perhaps, also Hildegard). Theoderic but had to score out Guibert's name, insert the necessary realia, and assemble the entire work under his name so as to finish the task. That Theoderic did not name Guibert as his predecessor is reminiscent of how Guibert had himself erased the existence of his predecessor Godfrey, so as to enhance his position as direct successor of Volmar (Schrader and Führkötter 147–50).

From what the sources tell us when it comes to Theoderic's final redaction, which was to assemble the pre-existent material, this second hypothesis wins my personal favour, although substantial weaknesses remain. The question arises why Guibert would have left Theoderic a *Vita* in such an advanced stage of completion, although 'completion,' of course, is a relative term in the Middle Ages. To him, whatever work he left behind in the Rupertsberg had been unfinished.[34] Here again, it has always been assumed that whatever work Guibert had started on a *Vita* during his time at the Rupertsberg is contained within the fragment sent to Bovo (Ep. 38). Why are The-

34. So much becomes clear from his letter to abbot Godfrey: "opus ceptum imperfectum reliquisse" (Ep. 42, 144, 394).

oderic's *Vita* and Guibert's fragmentary *Vita* so dissimilar, if we suspect that Guibert was at the origin of both of them? And finally: if Guibert's influence on the whole was as extensive as I am insinuating, then why does he appear – as the current experiments have shown – to have remained loyal to Hildegard's source material instead of extensively revising it?

These questions are bound to remain open for now. One might be excused for asking how far one is willing to go in speculation, if these results do not provide a better timeline than that of Klaes, or if they might simply be confronting us with the limits of what is methodologically feasible. Perhaps we are handling a collaborative style so far advanced that stylometry abandons us. The *impostors* method's suggestion is better taken seriously: there are simply no favourite candidates amongst all the authors included in the benchmark corpus. The involvement of many hands in a *Vita* undoubtedly important for many of Hildegard's close followers might defy the detection of single-author stylistic elements.

8 Conclusion

Whereas we have begun this article by questioning Hildegard's autobiographical fragments in the *Vita*, we have instead ended with new questions concerning Theoderic's and Guibert's respective contributions to the interbedding commentaries, where much remains unclear. It turns out that Hildegard's autobiographical fragments appear uncorrupted despite their transmittal through the hands of multiple biographers. This is the only result in this article that I believe can stand as conclusive. The experiments' remaining results, however, mainly give indications toward further investigation.

One of them is the observation that the two letters by Hildegard's sisters reporting on her death and appended to the *Vita's* third and final book, *Mulierem inquiunt* and *Cum beata,* are heavily indebted to Hildegard's wording and imagery. Either the sisters of the Rupertsberg meant to resuscitate Hildegard's tone and authority at the very end of her *Vita*, and perhaps even drew on Hildegard's materials so as to literally invoke her style, or perhaps Hildegard may even have had a hand in them herself. Undoubtedly, Hildegard's style was imitated at the Rupertsberg, where multiple of her assistants had been in the front row in learning to imitate and conjure up the visionary's style. Then again, that Hildegard was somehow involved herself is not impossible per

se. She is known to have participated in collecting and revising her *opera omnia* during the last years of her life, which had the aim of representing her image for posterity, and the saint's death is a crucial culmination point if a canonization project was envisioned, in which Hildegard was meant to be depicted in a larger-than-life, hagiographical fashion. For what it is worth, the *Vita* itself also reported on how Hildegard dictated the events surrounding her death to her sisters.

Paradoxically, whereas Hildegard's authority was not undermined in the autobiographical fragments, the largest tussle for stylistic dominance appears to have taken place in the commentaries guiding them. These have commonly been taken to have been written by end redactor Theoderic of Echternach. However, Guibert of Gembloux's style appears present in a few of them, which might lead one to suspect that either Guibert had opportunity to revise the *Vita* at a time when Theoderic was (near to) completing it, or else that Theoderic largely relied on preparations carrying Guibert's mark. The first argument is hardly sustainable when based on Guibert's whereabouts during the time of Theoderic's ending of the *Vita*, but there is something to be said for the latter hypothesis. Then again, to my knowledge there is no additional evidence to support it aside from the statistical suggestions in this article, wherefore the question remains open and no conclusive answer is possible yet.

Bibliography

Binongo, José Nilo G., and M. Wilfrid A. Smith. "The Application of Principal Components Analysis to Stylometry." *Literary and Linguistic Computing* 14.4 (1999): 446–66.

Boglioni, Pierre. "La mort dans les premières hagiographies latines." *Le sentiment de la mort au moyen âge*. Ed. Claude Sutto. Montreal: L'aurore, 1979. 185–210.

Bredero, Adriaan Hendrik. "Der Brief des heiligen Bernhard auf dem Sterbebett: eine authentische Fälschung." *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986*. Vol. 5.

MGH Schriften 33. Hannover: Hahnsche Buchhandlung, 1988. 201–24.

---. *Études sur la "Vita prima" de saint Bernard*. Rome: Editiones Cistercienses, 1960.

Bruder, Petro. *Acta inquisitionis de virtutibus et miraculis S. Hildegardis*. Analecta Bollandiana 2 (1883): 116–29.

Daelemans, Walter. "Explanation in Computational Stylometry." *Computational Linguistics and Intelligent Text Processing* 7817. Ed. Alexander Gelbukh. Berlin, 2013. 451–62.

Dalarun, Jacques. "La mort des saints fondateurs. De Martin à François." *Les fonctions des saints dans le monde occidental (IIIe-XIIIe*

siècle). Actes du colloque de Rome (27–29 octobre 1988). Rome: École Française de Rome, 1991. 193–215.

De Gussem, Jeroen. "Bernard of Clairvaux and Nicholas of Montiéramey: Tracing the Secretarial Trail with Computational Stylistics." *Speculum* 92.S1 (2017): S190–S225.

De Gussem, Jeroen and Dinah Wouters. "Language and Thought in Hildegard of Bingen's Visionary Trilogy: Close and Distant Readings of a Thinker's Development." *Parergon* 36.1 (2019): 31–60.

Deploige, Jeroen. *In nomine femineo indocta: Kennisprofiel en ideologie van Hildegard van Bingen (1098-1179)*. Hilversum: Verloren, 1998.

Derolez, Albert. "Deux notes concernant Hildegarde de Bingen." *Scriptorium* 27 (1973): 291–99.

---. "The Manuscript Transmission of Hildegard of Bingen's Writings: The State of the Problem." *Hildegard of Bingen: The Context of Her Thought and Art*. Ed. Charles S.F. Burnett and Peter Dronke. Warburg Institute Colloquia 4. London: Warburg Institute, 1998. 17–27.

Diederich, Joachim, Jörg Kindermann, Edda Leopold and Gerhard Paass. "Authorship Attribution with Support Vector Machines." *Applied Intelligence* 19 (2003): 109–23.

Downey, Sarah, Michael D.C. Drout, Veronica E. Kerekes and Douglas C. Raffle. "Lexomic Analysis of Medieval Latin Texts." *The Journal of Medieval Latin* 24 (2014): 225–75.

Dronke, Peter. "The Allegorical World-Picture of Hildegard of Bingen: Revaluations and New Problems." *Hildegard of Bingen: The Context of Her Thought and Art*. Ed. Charles S.F. Burnett and Peter Dronke. Warburg Institute

Colloquia 4. London: Warburg Institute, 1998. 1–16.

---. *Women Writers of the Middle Ages: A Critical Study of Texts from Perpetua († 203) to Marguerite Porete († 1310)*. Cambridge, UK: Cambridge University Press, 1984.

Dutton, Marsha L. "A Case for Canonization: The Argument of the *Vita Prima Sancti Bernardi*." *Cistercian Studies Quarterly* 52.2 (2017): 131–60.

Eder, Maciej. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Digital Scholarship in the Humanities* 30.2 (2013): 167–82.

---. "In Search of the Author of the *Chronica Polonorum* Ascribed to Gallus Anonymus: A Stylometric Reconnaissance." *Acta Poloniae Historica* 112 (2015): 5–23.

---. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 32.1 (2017): 50–64.

Eder Maciej, Jan Rybicki and Mike Kestemont. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8 (2016): 107–21.

Embach, Michael. *Die Schriften Hildegards von Bingen: Studien zu ihrer Überlieferung und Rezeption im Mittelalter und in der Frühen Neuzeit*. Erudiri sapientia: Studien zum Mittelalter und zu seiner Rezeptionsgeschichte 4. Berlin: Akademie Verlag, 2003.

Ferrante, Joan. "*Scribe quae vides et audis*. Hildegard, Her Language, and Her Secretaries." *The Tongue of the Fathers: Gender and Ideology in Twelfth-Century Latin*. Ed. David Townsend, Taylor Andrew and Ruth Mazo Karras. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1998. 102–35.

Ferzoco, George. "The Canonization and Doctorization of Hildegard of Bingen." *A Companion to Hildegard of Bingen*. Ed. Beverly Mayne Kienzle, Debra L. Stoudt and George Ferzoco. Leiden: Brill, 2014. 305–16.

Flanagan, Sabina. "*Spiritualis amicitia* in a Twelfth-Century Convent? Hildegard of Bingen and Richardis of Stade." *Parergon* 29 (1981): 15–21.

Garay, Kathleen and Madeleine Jeay. "Sanctification of the Body: The Stages and Staging of Holy Women's Death in High Medieval Europe." *Heroes and Saints: The Moment of Death in Cross-Cultural Perspectives*. Ed. Phyllis Granoff and Koichi Shinohara. Newcastle-upon-Tyne: Cambridge Scholars, 2007. 136–58.

Godfrey of Saint-Disibodenberg and Theoderic of Echternach. *The Life of Hildegard*. In *Jutta and Hildegard: The Biographical Sources*. Ed. and transl. Anna Silvas. Medieval Women: Texts and Contexts 1. Turnhout: Brepols, 1998. 118–210.

---. *Vita sanctae Hildegardis virginis*. Ed. Monika Klaes. *CC CM* 126. Turnhout: Brepols, 1993.

Greenspan, Kate. "Autohagiography and Medieval Women's Spiritual Autobiography." *Gender and Text in the Later Middle Ages*. Ed. Jane Chance. Gainesville: University Press of Florida, 1996. 216–36.

Guibert of Gembloux. *De combustione monasterii Gemblacensis*. Ed. Georg Heinrich Pertz. MGH Scriptores 8. Hannover, 1848. 563–64.

---. *Epistolae quae in codice B.R. BRUX. 5527–5534 inueniuntur*. Ed. Albert Derolez. *CC CM* 66–66A. Turnhout: Brepols, 1988–89.

---. *Vita sanctae Hildegardis retracta-ta*. Ed. Monika Klaes. *CC CM* 126. Turnhout: Brepols, 1993. 93–106.

Herwegen, Hildephonse. "Les collaborateurs de sainte Hilde-garde." *Revue Bénédictine* 21.1 (1904): 192–204; 302–16 (*Suite*); 381–403 (*Fin*).

Hildegard of Bingen. *Epistolarium Pars Prima I–XC*. Ed. Lieven Van Acker. *CC CM* 91. Turnhout: Brepols, 1991.

---. *Epistolarium Pars Secunda XCI–CCL^R*. Ed. Lieven Van Acker. *CC CM* 91A. Turnhout: Brepols, 1993.

---. *Epistolarium Pars Tertia CCLI–CCCXC*. Ed. Lieven Van Acker and Monika Klaes-Hachmöller. *CC CM* 91B. Turnhout: Brepols, 2001.

---. *Liber divinorum operum*. Ed. Albert Derolez and Peter Dronke. *CC CM* 92. Turnhout: Brepols, 1996.

---. *Liber vite meritorum*. Ed. Angela Carlevaris. *CC CM* 90. Turnhout: Brepols, 1995.

---. *Scivias*. Ed. Adelgundis Führköt-ter and Angela Carlevaris. *CC CM* 43. Turnhout: Brepols, 1978.

---. *Visiones Hildegardis a Guiberto Gemblacensi exaratae*. Ed. Jeroen Deploige and Sara Moens. *Hildegardis Bingensis opera minora* II. Ed. Jeroen Deploige, Michael Embach, Christopher P. Evans, Kurt Gärtner and Sara Moens. *CC CM* 226A. Turnhout: Brepols, 2016. 135–236.

---. *Vita Sancti Disibodi episcopi*. Ed. Christopher P. Evans. *Hildegardis Bingensis opera minora* II. Ed. Jeroen Deploige, Michael Embach, Christopher P. Evans, Kurt Gärtner and Sara Moens. *CC CM* 226A. Turnhout: Brepols, 2016. 59–87.

---. *Vita Sancti Ruperti confessoris*. Ed. Christopher P. Evans. *Hilde-gardis Bingensis opera minora* II.

Ed. Jeroen Deploige, Michael Embach, Christopher P. Evans, Kurt Gärtner and Sara Moens. *CC CM* 226A. Turnhout: Brepols, 2016. 91–108.

Jockers, Matthew Lee. *Macroanaly-sis: Digital Methods and Literary History*. Champaign, IL: Universi-ty of Illinois Press, 2013.

Johnson, Lynn Staley. "The Trope of the Scribe and the Question of Literary Authority in the Works of Julian of Norwich and Margery Kempe." *Speculum* 66.4 (1991): 820–38.

Juola, Patrick. "Authorship Attribu-tion." *Foundations and Trends in Information Retrieval* 1.3 (2006): 233–334.

Katajala-Peltomaa, Sari. "Recent Trends in the Study of Medieval Canonizations." *History Compass* 8.9 (2010): 1083–92.

Kestemont, Mike, Sara Moens and Jeroen Deploige. "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux." *Digital Scholarship in the Humanities* 30. 2 (2013): 199–224.

Koppel, Moshe and Yaron Winter. "Determining If Two Documents Are Written by the Same Author." *Journal of the Association for Information Science and Technolo-gy* 65.1 (2014): 178–87.

Liebeschütz, Hans. *Das allegorische Weltbild der Heiligen Hildegard von Bingen*. Leipzig: Teubner, 1930.

Moens, Sara. "De horizonten van Guibertus van Gembloers (ca. 1124–1214). De wereld van een benedictijns briefschrijver in tijden van een verschuivend religieus landschap." Ph.D. Dissertation, Ghent University, 2014.

Moretti, Franco. *Distant Reading*. London: Verso, 2013.

Mosteller, Frederick and David L. Wallace. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. 2nd ed. Springer Series in Statistics. New York: Springer Verlag, 1984 (1964).

Newman, Barbara. "Hildegard and Her Hagiographers. The Remak-ing of Female Sainthood." *Gen-dered Voices. Medieval Saints and Their Interpreters*. Ed. Catherine M. Mooney. The Middle Ages Series. Philadelphia: University of Pennsylvania Press, 1999. 16–34.

---. *Sister of Wisdom. St Hildegard's Theology of the Feminine*. Berkeley and Los Angeles: University of California Press, 1987.

Powell, Morgan. "*Vox ex negativo*. Hildegard of Bingen, Rupert of Deutz and Authorial Identity in the Twelfth Century." *Unver-wechselbarkeit. Persönliche Identität und Identifikation in der vormodernen Gesellschaft*. Ed. Peter von Moos. Norm und Struktur 23. Cologne: Böhlau, 2004. 267–95.

Preger, Wilhelm. *Geschichte der deutschen Mystik im Mittelalter*. Leipzig: Dörffling und Franke, 1874.

Schrader, Marianna and Adelgundis Führkötter. *Die Echtheit des Schrifttums der Heiligen Hildegard von Bingen*. Cologne / Graz: Böhlau Verlag, 1956.

Stamatatos, Efstathios. "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology* 60.3 (2009): 538–56.

Theoderic of Echternach. *Chronicon Epternacense*. Ed. Ludwig Wei-land. MGH Scriptores 23. Hanno-ver, 1874. 38–64.

---. *Libellus de libertate Epternacensi propugnata*. Ed. Ludwig Weiland.

MGH Scriptores 23. Hannover, 1874. 64–72.

---. *Octo lectiones in festo sancte Hildegardis legende*. Ed. Monika Klaes. *CC CM* 126. Turnhout: Brepols, 1993. 75–80.

Van Acker, Lieven. "Der Briefwechsel der heiligen Hildegard von Bingen Vorbemerkungen zu einer kritischen Edition (*Fortsetzung*)." *Revue Bénédictine* 99.1–2 (1989): 129–34.

Van Engen, John. "Authorship, Authority, and Authorization: The Cases of Abbot Bernard of Clairvaux and Abbess Hildegard of Bingen." *Shaping Authority: How Did a Person Become an Authority in Antiquity, the Middle Ages and the Renaissance?* Ed. Shari Boodts, Johan Leemans, and Brigitte Meijns. Lectio 4. Turnhout: Brepols Publishers, 2016. 325–62.

---. "Letters and the Public Persona of Hildegard." *Hildegard von Bingen in ihrem historischen Umfeld. Internationaler wissenschaftlicher Kongreß zum 900jährigen Jubiläum, 13.–19. September 1998, Bingen am Rhein*. Ed. Alfred Haverkamp and Alexander Reverchon. Mainz: Verlag Philipp von Zabern, 2000. 375–418.

van Halteren, Hans *et al.* "New Machine Learning Methods Demonstrate the Existence of a Human Stylome." *Journal of Quantitative Linguistics* 12.1 (2005): 65–77.

Vainio, Raija *et al.* "Reconsidering Authorship in the Ciceronian Corpus Through Computational Authorship Attribution." *Ciceroniana On Line* 3.1 (2019).

Vauchez, André. *La sainteté en Occident aux derniers siècles du Moyen Age: d'après les procès de canonization et les documents hagiographiques*. Rome: École française de Rome, 1981.

Appendix

Table 5 (Appendix). Word correspondences between *Mulierem inquiunt* (*Vita* §3.26, 15–38, 68–69) and Hildegard's canonical works, corresponding passages automatically searched by using Levenshtein distance. Latin text of the *Vita* in left column edited by Monika Klaes.

Mulierem inquiunt	Matches with passages in Hildegard's oeuvre
Mulierem inquiunt quandam acriter a demonio muto uexatam, **super quam** et fratres de Lacu plurimum laborauerant,	... ut sol rem aliquam calefacit **super quam...** (*Scivias*, *Protestificatio* 29)
cum ad se **magno** labore uirorum in	... **cum magno** sacramento incarnationis sue (*Scivias*, II.3.34, 737) ...
lecto deportata esset, pia mater **audacie et presumptioni** demonis […]	... **audatie et presumptionis** temeritatem ... (*LDO*, I.3.14, 1)
benedictionibus **non cessauit, quousque** per gratiam […]	... **nec** in hoc **cessabit, quousque** numerus ... (LDO, III.4.9, 10)
que propter furorem insanie diris	... Homo autem, **qui propter** timorem ... (*LDO*, I.2.35, 24)
[…] **orationes, uigilias** et **ieiunia**,	... que **cum uigiliis et ieiuniis** ac **orationibus** ... (*LDO*, II.1.39).
ad perceptionem quoque **sacramentorum** […]	... **ad percipiendum** idem **sacramentum** ... (*Scivias*, II.6, 243) ...
Inter **que etiam ita** eam	ager fructum proferens **qui etiam ita** Deo est consecratus ... (*Scivias*, II. 5.48, 1494)
afflixit, quod nomina et	... in tantum **affligitur, quod** ... (*LDO*, I.4.64, 27)
aspectum **quorundam hominum** et animalium **in tantum** abhorrebat, **quod** ipsis uisis	... cerebrum **quorundam hominum** igneum et siccum est ... (LDO, I. 2.32, 166) ... **in tantum** affligitur, **quod** ... (*LDO*, I.4.64, 27).
uel auditis horribili uoce per longam horam perstrepebat. Hec a priore et conuentu cum litteris ad sanctam uirginem missa ab ea et **confortata** et a	**confortare** is an often used word in Hildegard's oeuvre.
[…] cum esset **paupercula** et **ceca**,	... Ego igitur **paupercula** et inbecillis ... (*LDO*, *Prologus*, 27)
in elemosinam eius recepta in spirituali habitu uitam **feliciter consummauit**	… sed tamen ea bono fine **feliciter consummauit** … (*Vita sancti D.*, 10, 136, 63)

Table 6 (Appendix). Word correspondences between *Cum beata* (*Vita* §3.27, 4–35, 69–70) and Hildegard's canonical works. Latin text of the *Vita* in left column edited by Monika Klaes.

Cum beata	Matches with passages in Hildegard's oeuvre
Cum beata inquiunt mater Domino multis laborum certaminibus deuote militasset, uite presentis **tedio affecta**	... labore et **tedio affectum** ... (*LDO*, I.2.35, 12)
dissolui et esse cum Christo cottidie **cupiebat**. Cuius **desiderium** Deus exaudiens finem suum,	... in candore tantum celestis **desiderii querit dissolui et esse cum Christo** ... (*Scivias*, III. 10.22 648)
sicut ipsa preoptauerat, **spiritu prophetie**	... in **spiritu prophetiae** cognouerunt ... (*LVM*, 2.30, 545)
ei reuelauit, quem et sororibus predixit. Aliquamdiu **itaque infirmitate laborans** octogesimo secundo **etatis sue** anno	**itaque** ... **infirmitate laborasset**, uicesimo **etatis sue** anno ... (*Vita Sancti Ruperti*, 11, 358, 103.)
[...] conferendis non dubitarent, **propter** discessionem tamen eius,	... **propter** sanguinem agni, **per quem** ...(*LDO*, III.5.37, 25)
per quam semper consolabantur,	... pugna, **per quam** semper ... (*Scivias*, III.6.30, 750)
[...] lucidissimi et **diuersi coloris** arcus	... et **diuerso colore** depicta ... (*Scivias*, III.10.9, 433)
in firmamento apparuerunt, qui ad	... magna **in firmamento** discurrunt, sic ...(*LDO*, I.4.51, 14)
magnitudinem magne platee **se dilatauerunt**	... **se dilatauerant**, in hac palude ... (*LVM*, I.121, 1852)
in quatuor partes terre se **extendentes**,	... **in quattuor partes** se diuiserunt... (*Scivias*, II.7, 84)
	... **ad quattuor** plagas orbis **extendentem**, ... (*Scivias*, I.4.9, 398)
	... super **quatuor partes terrae** ... (*LDO* II.1.8, 65)
quorum alter ab **aquilone ad austrum**,	... ad **austrum** et altera **ad aquilonem** ... (*LDO*, I.4.49, 66)
alter ab **oriente** ad occidentem **procedebant**.	... ad **orientem procedebant** rami a ... (*Scivias*, III.4, 59)
	... **procedebant**, se in altitudine ... (*Scivias*, III.6.35, 1010)
At **in summitate**, ubi hi	... **in** cuius **summitate**, ubi locus ... (*LDO*, III.4.1, 10)
[...] se protendens **tenebras noctis** ab habitaculo	... **tenebras noctis** cum mala ... (*LVM*, III.28, 527)
depellere **uidebatur**. **In** hac	... excellere **uideretur**; **in** quo ... (*LDO*, III.1.1, 5)
[...] **quam innumerabiles** uarii	... **qui innumerabiles in numero** ... (*LVM*, I.49, 777)
coloris circuli, **in quibus** singulis singule rutilantes crucicule oriebantur, cum	... **uarii coloris** induta est, ... (*LVM*, 2.47, 935)
	... **circuli**, **in quo** similitudo ... (*LDO*, I.2.1, 160)

circulis suis crescentes priore tamen minores conspiciebantur.	... se **in firmamento** distantes signati ... (*LDO*, I.4.22, l.1)
Et cum he **in firmamento** se	... **latitudinem** multam habentem, ac ... (*LVM*, IV.50, 1090)
dilatassent, **latitudine** sua	... et **ad orientem uersam** ... (*LDO*, III.5.2, l.19)
ad orientem magis pertingebant et ad terram **uersus** domum in	
qua sancta uirgo transierat, declinare uise **totum montem clarificabant**.	... **totum mundum illuminaret** ... (*LDO*, III.5.9, 100) ... **signis** se **ostendunt**; quoniam ... (*LDO*, III.1.5, 4)
Et credendum quod **hoc signo** Deus	... **tanta claritate** fulgebat, ut ... (*LDO*, III.3.1, 7)
ostendit, **quanta claritate** dilectam suam in celestibus illustrauerit.	

Table 7 (Appendix). Authors contained in the benchmark corpus. More detailed information (word count and document titles) is provided on GitHub.

Alan of Lille	Hugh of Saint-Victor
Anselm of Canterbury	Ivo of Chartres
Anselm of Laon	John of Salisbury
Bernard of Clairvaux	Odo of Deuil
Bruno of Cologne	Peter Abelard
Gerhoh of Reichersberg	Peter of Celle
Walter of Châtillon	Peter Damian
William of Conches	Peter Lombard
William of Saint-Thierry	Peter the Venerable
Hildebert of Lavardin	Rupert of Deutz
Honorius of Regensburg (Autun)	Suger of Saint-Denis