Collaborative Authorship in the Twelfth Century

A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux

Jeroen Deploige / Department of History / Ghent University
Mike Kestemont / Research Foundation Flanders / University of Antwerp
In close collaboration with Sara Moens Department of History / Ghent University

Guest Lecture, AI Course, ULB, 9 May 2014
Computational Authorship Attribution

- Stylometry
- Finger print
- **Stylome** (Van Halteren et al. 2005)
- Stylistic DNA
- (Tendentious)
• Young paradigm (1960)
• Mosteller & Wallace (US)
• Federalist papers (1780s)

Innovation on 2 levels:
• Quantitative approach
• Functors
Traditional

- Guesswork
- Conspicuous features
- Odd verbs
- Checklist
- But...
  - schools, workshops, ...
  - Tradition
  - Forgeries, imitation, ...
  - ...

Mosteller & Wallace

- Inconspicuous features
- Function words or functors
  - articles (the, it, a)
  - prepositions (on, from, to)
  - pronouns (self, he)
Advantage?

Many observations
All authors, same set
Relatively content-independent
Count the number of $f$'s on the following slide...
Finished files are the result of years of scientific study combined with the experience of many years.
How many?
Do we process functors ‘unconsciously’?

Finished files are the result of years of scientific study combined with the experience of many years.
Which text is on the following slide?
PARIS IN THE SPRING
So?
Difficult error spotting...

Detection errors on the and and: Evidence for reading units larger than the word

ADAM DREWNOWSKI
Rockefeller University, New York, New York 10021
and

ALICE F. HEALY
Yale University, New Haven, Connecticut 06520
and Haskins Laboratories, New Haven, Connecticut 06511

In five experiments, subjects read 100-word passages and circled instances of a given target letter, letter group, or word. In each case subjects made a disproportionate number of detection errors on the common function words the and and. The predominance of errors on these two words was reduced for passages in which the words were placed in an inappropriate syntactic context and for passages in which word-group identification was disturbed by the use of mixed typefaces or a list, rather than a paragraph, format. These effects for the word and were not found for the control word art. These results were taken as evidence that familiarity...
Unimportant?
Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn’t mtttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.
Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn’t mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rght pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.
Parallel in art history
Morelli (1816–1891)
Function words

- Popular features in stylometry
- Frequencies as input
- Machine Learning
Hildegard of Bingen

- Influential women writer
- 1098–1179
- Germany
- Divine visions
- “Sybil of the Rhine”
Varied oeuvre

- Visions
- Music
- Scientific texts
- Recipes
- Medical treatises
- Letters (pope, emperor, ...)

[Dendermonde, St.-Pieters & Paulusabdij, Ms. Cod. 9]
Early 2012

- Sara Moens
- Jeroen Deploige
- Dept. History, UGhent
- Editing two texts
- Collaborate?
Secretaries

- Wrote in Latin...
- But was bad at it!
- No formal training as woman
- Assisted by **male secretaries**
- Gender issues...
- Dictated

[Hildegard and her 1st secretary Volmar]
Correction grammatical mistakes

(Only form, not content!)
Two shorter texts...

Visio ad Guibertum missa &

Visio de sancto Martino

• “Attributed” to Hildegard

• Opera omnia...

• But style not typical of her

• Doubts authorship?

• Last secretary....

[MS Brussels, Royal Library, 5527-34, fol. 141v.]
Guibert of Gembloux

- Monk from Brabant
- Hildegard’s last secretary
- Fascination St Martin
- Very elaborate style
- “Pushy”

[MS Brussels, Royal Library, 5527-34, fol. 141v.]
When you correct [this text], keep to this rule: that [...] you apply your skill only to make corrections where the order or the rules of correct Latin are violated. Or if you prefer – and this is something I have conceded in this letter beyond my normal practice – you need not hesitate to clothe the whole sequence of the vision in a more becoming garment of speech, preserving the true sense

Corpus

- *Corpus Christianorum* (Brepols)
- Complete materials
- *Epistolaria*
- Hildegard, Guibert
- Bernard of Clairvaux (1090-1153)
- 3x +100k tokens
“Stylometry with R”  
[Eder, Rybicki & Kestemont 2013]
Test PCA

[ss=10,000; 65 MFW; content words ‘culled’]

[Striking differences in MFW!]
Sample size?

[ss=5,000]

[ss=1,000]
Learning curve “Delta”

\[
\Delta_B^n(D, D') = \sum_{i=1}^{n} \frac{1}{\sigma_i} |f_i(D) - f_i(D')|
\]
Anonymous texts?

[ss=3,706; Bernard's *Sermo in festo sancti Martini* as “anonymous test case”; ]
Bigger picture
Synergy Hypothesis

- Pennebaker (e.g. 2011)
- *The Secret Life of Pronouns*
- Federalist papers and Beatles songs
- Collaborative writing style?
- “unlike either of one of the styles that the collaborating authors would produce on their own”
- Practical and theoretical relevance
“Hollywood version”? Online documentary
vimeo.com/70881172
References