

# Rolling Stylometry

Authorship  
attribution within  
documents

J.Nerbonne

# “Rolling” Stylometry

- Rybicki, Hoover, Kestemont (2014) “Collaborative authorship” LLC 29(3):422-431.
- Applying authorship attribution within single documents to attribute subsections
- Focus on Joseph Conrad & Ford Maddox Ford, who published *The Inheritors* (1901) together, also *Romance* (1903), & *The Nature of a Crime* (1909)
  - But who wrote what within each book?

# Other issues

- Ford claimed to have had a role in Conrad's *Nostromo*, which Conrad denied.
  - *Nostromo* regarded as one of Conrad's better works
  - Can we detect a signal of Ford in *Nostromo*?

# Important question

- Two very influential, appreciated authors
  - Conrad's *Heart of darkness*, *Lord Jim* & *The secret agent* often still listed among “the best novels”, *Heart of darkness* basis for Coppola film *Apocalypse now*.
    - But also still controversial as perhaps “colonialistic” (Achebe, 1975)
  - Ford's works also highly praised, esp. *The good soldier*
- They could have helped each other
  - Conrad established, successful, but unconfident writing English (3<sup>rd</sup> lg.)
  - Ford, younger, still mostly unread

# Burrow's $\Delta$

- Comparison of normalized relative frequencies of n most frequent words (MFW's)
  - Used successfully to identify authors in disputed cases
    - See course slides (earlier)
- Collect material from both authors
  - E.g. Conrad & Ford
- Obtain characteristic  $\Delta$ -vectors for n MFWs
- Check whether material in question is distinguished

## Yet another $\Delta$ -variant

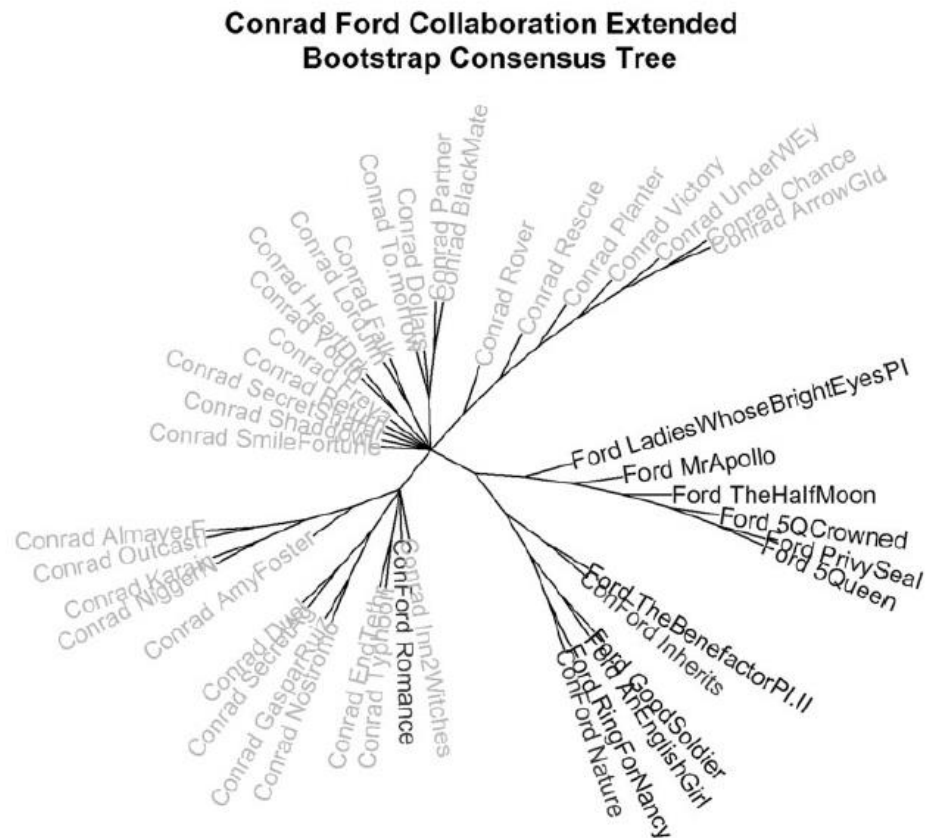
- In fact, Rybicki et al. use a slightly different  $\Delta$ , due to Argamon
- Recall standard deviation:  $sd(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m_x)^2}$ 
  - Measure of fluctuation around the mean
- Rybicki et al. use:  $\Delta_{RHK}(R, W) = \sum_{i=1}^n \frac{1}{sd_i(R)} |m_i(R) - f_i(W)|$ 
  - Where  $R$  is the reference corpus,  $W$  is material under analysis
  - So  $\Delta_{RHK}$  expresses difference in terms of  $sd$ 's (like  $z$ -scores)

# Difference between $\Delta$ -variants

- Rybicki et al. use:  $\Delta_{RHK}(D, D') = \sum_{i=1}^n \frac{1}{sd_i(D)} |m_i(D) - f_i(D')|$
- Burrows used:  $\Delta_B(D, D') = \sum_{i=1}^n |z_i(D) - z_i(D')|$ 
  - Burrows normalizes relative frequencies using sd within document
  - Rybicki et al. normalizes according to reference document alone
- No explanation in paper
- But it's very close to the  $\Delta_B$ !

# All docs together

- All works by either author
- Differences measured by varying numbers of MFW ( $100 \leq n \leq 1000$ )
- Some culling of words (typical of single texts), e.g. pronouns
- Texts clustered different ways
- Consensus tree clusters elements that cluster  $> 50\%$  of the time
- Co-authored texts indeed mixed with others
  - *Romance* among Conrad texts
  - Other two among Ford texts
  - Other clusters “pure”

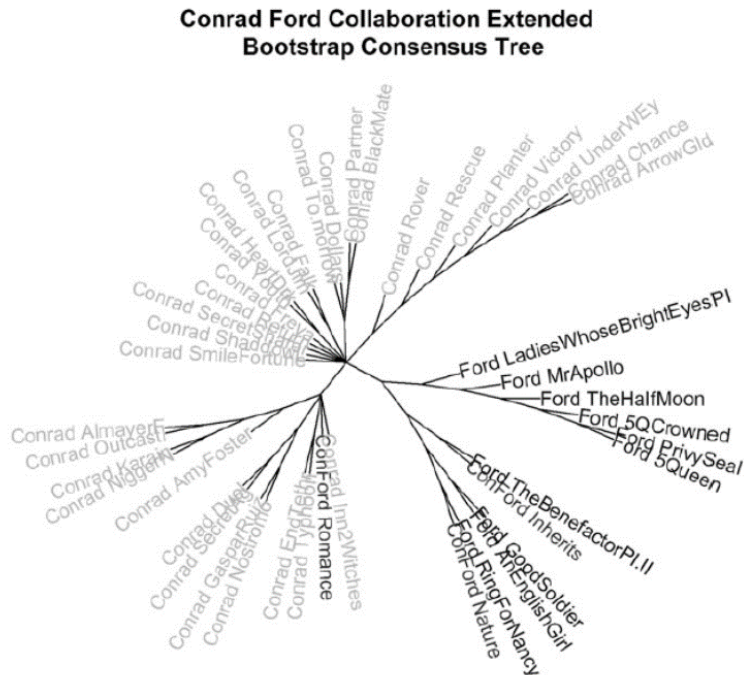


100-1000 MFW Culled @ 100%  
Pronouns deleted Classic Delta distance Consensus 0.5



# What does consensus clustering show?

- No clean separation
  - Usually we'd be disappointed at not seeing a good separation
- But the collaboration is known
- Only jointly authored texts disturb purity!



100-1000 MFW Culled @ 100%  
Pronouns deleted Classic Delta distance Consensus 0.5

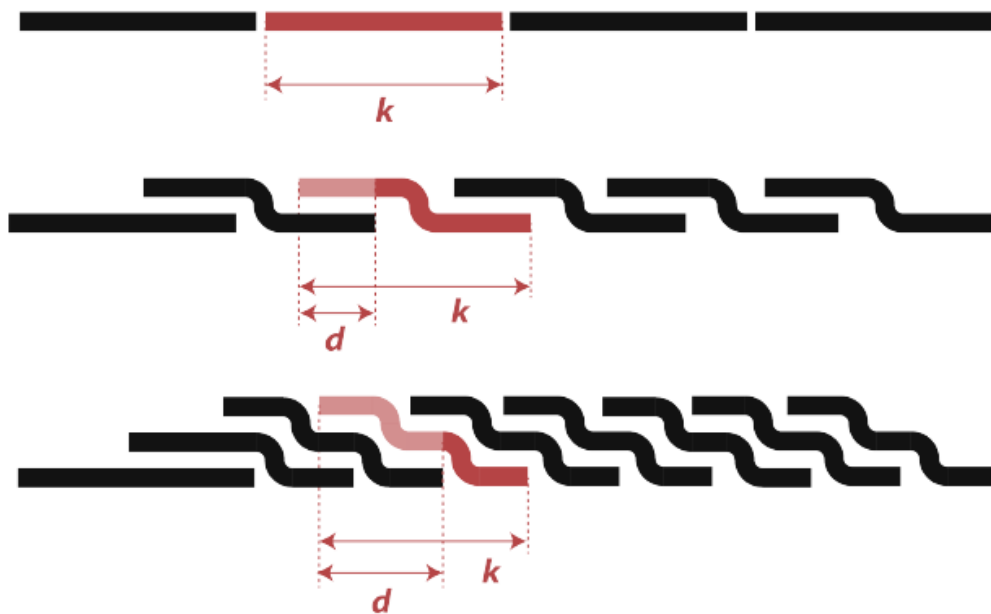
# See the collaboration more precisely?

- Where in the joint texts do we mostly see Conrad, resp. Ford?
- Can we apply Burrows's  $\Delta$  not only between but also within texts?

# Idea of “rolling delta”

- Collect reference material from both individually, Conrad material, Ford material, i.e., material without collaboration (Conrad & Ford)
- Obtain characteristic vectors for n MFWs:  $\vec{c}, \vec{f}$  (Conrad, Ford)
  - Alternative: obtain characteristic vectors per document (novel)
  - In fact, the latter was done
- Divide document to be examined into overlapping WINDOWS, e.g., of 5K words, with a step determining overlap, e.g. 100 wd
  - win1 1-5K, win2 101-5,1K, win3 201-5,2K, ...
- Measure each window  $\vec{w}$  to determine if it's closer to  $\vec{c}$  or  $\vec{f}$ 
  - Or, compare each  $\vec{w}$  to each of the novels to determine its similarity

# Windowing

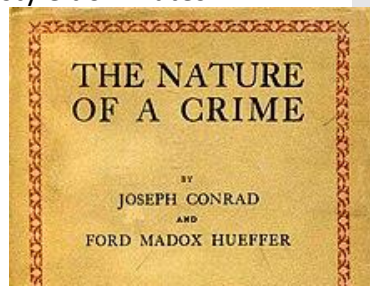
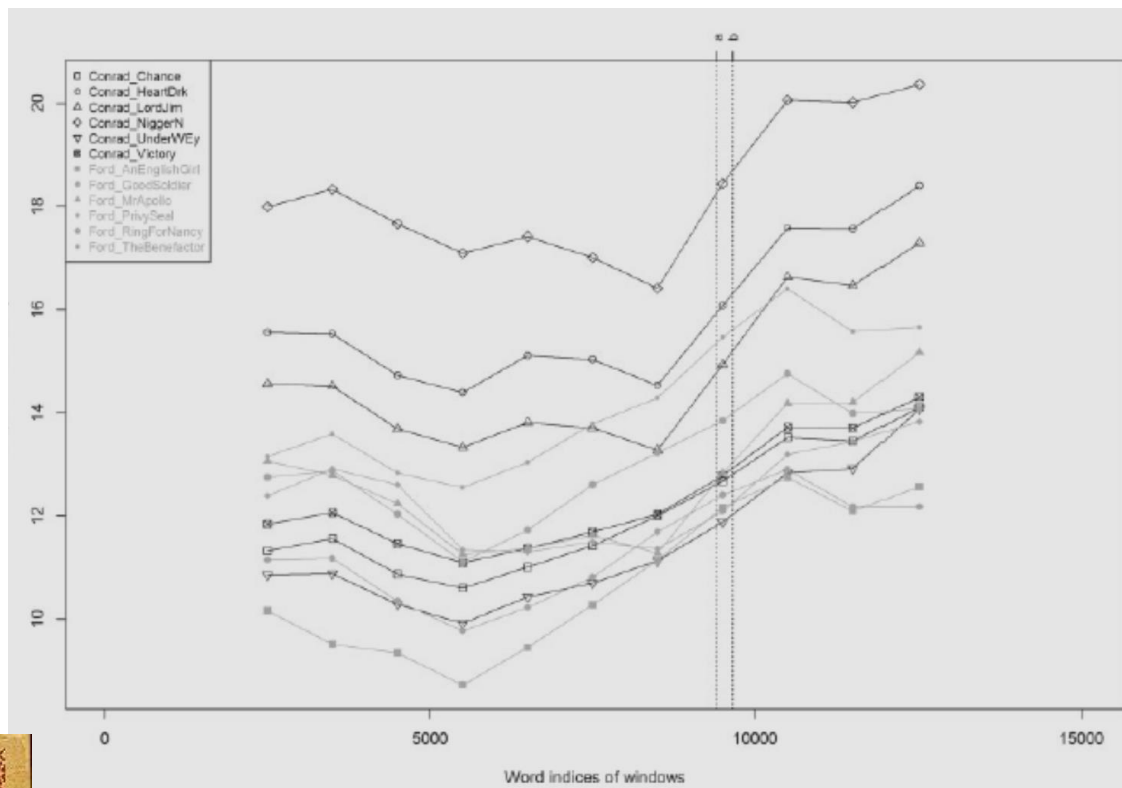


- $k$  window size
- $d$  size of overlap
- Like training in n-gram models
- Result: for each window  $\vec{w}$ , a decision whether it's more like one author or the other

$$\Delta(\vec{w}, \vec{f}) < \Delta(\vec{w}, \vec{c})$$

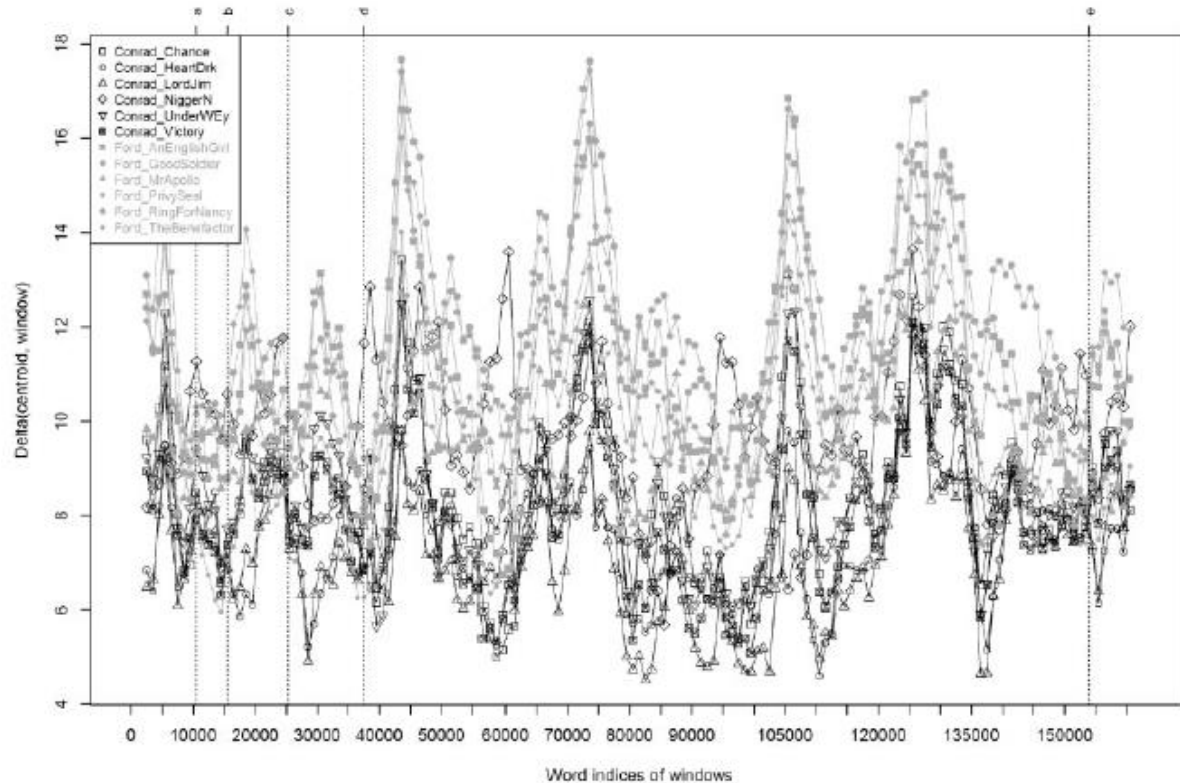
## *The nature of a crime (NoC), a collaboration*

- $x$ -axis: position in the text (*The nature of a crime*)
- $y$ -axis:  $\Delta$ -distance to NoC (see legend)
  - $y = 0$  : work's zero distance to itself
- Differences between Conrad's works (dark lines) & NoC ( $y = 0$ )
- Differences between Ford's works (light lines) & NoC ( $y = 0$ )
- The works with lower lines are closer
- Fairly mixed up to points a, b
  - Then Ford's style dominates



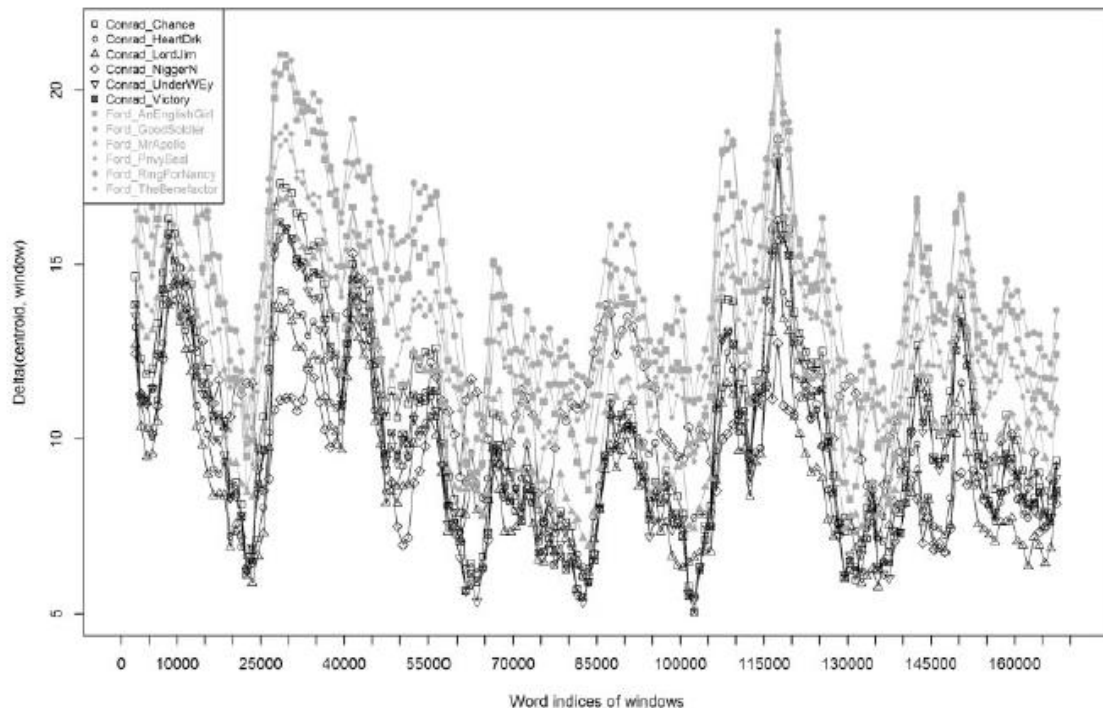
## ***Romance, also collaborative***

- Dominated by dark lines (Conrad)
- Conrad's influence dominates throughout the novel
- Aside: lots of fluctuation through the book!
- Lots of passages unlike either author's earlier work



## What about *Nostromo*?

- Regarded as one of Conrad's best
- Published solely under Conrad's name, but Ford later claimed credit for a "fragment"
- Conrad's dark lines dominate throughout
  - Conrad's style dominates



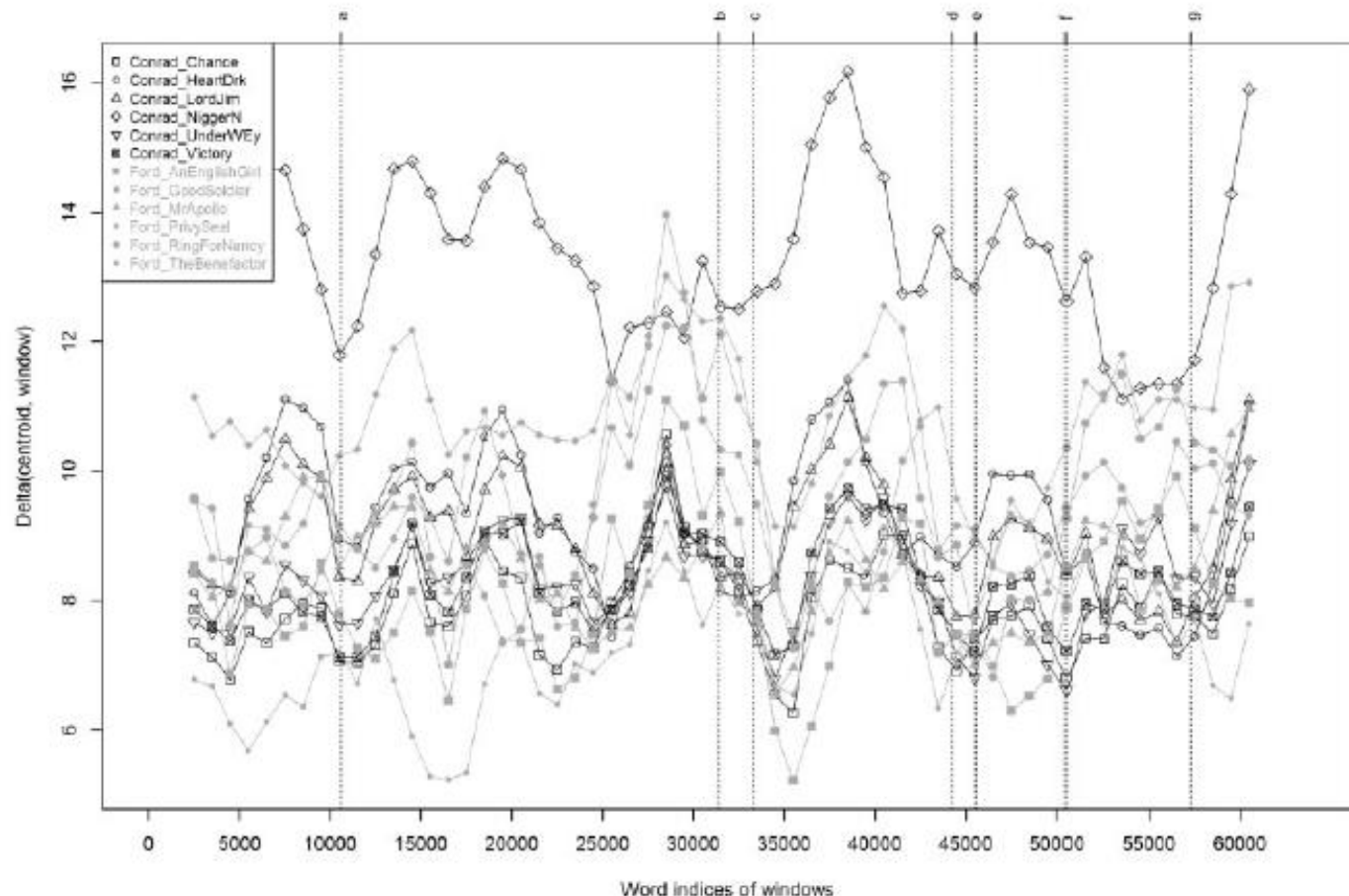
# Examining *The inheritors*

- Rybicki et al. compare *The inheritors* to six novels each of Conrad & Ford
- Again: 1,000 MFWs, window 5K words, step (increment) 1K wd.



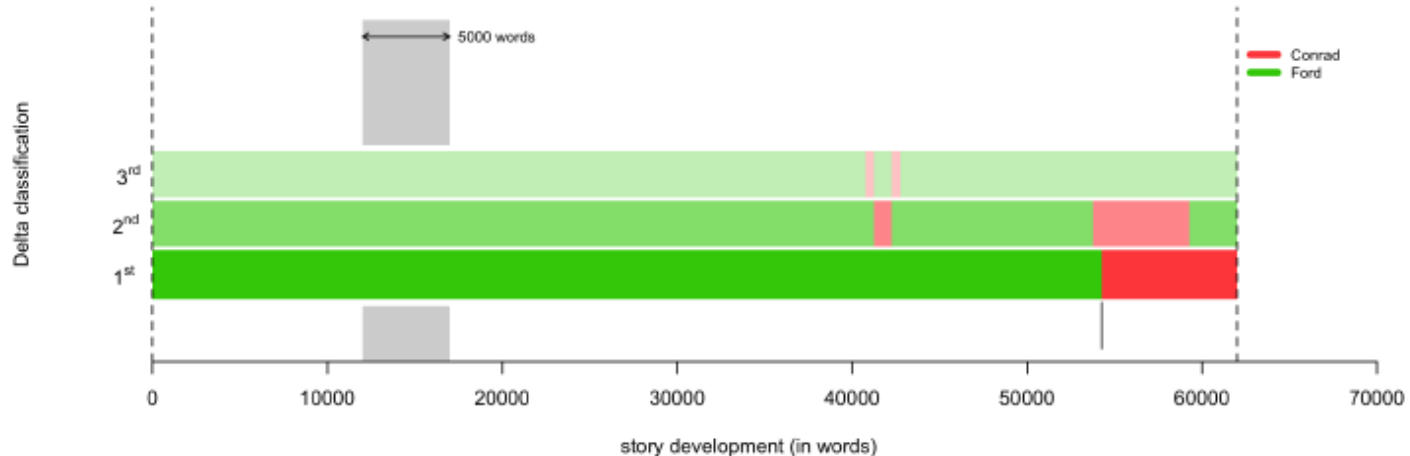


- Mostly Ford!
  - (Closest)
- Points a-e (above) closer to Conrad!
- But note that the relation to Conrad's style is more consistent
- Maybe “rolling” needs to look at references together?

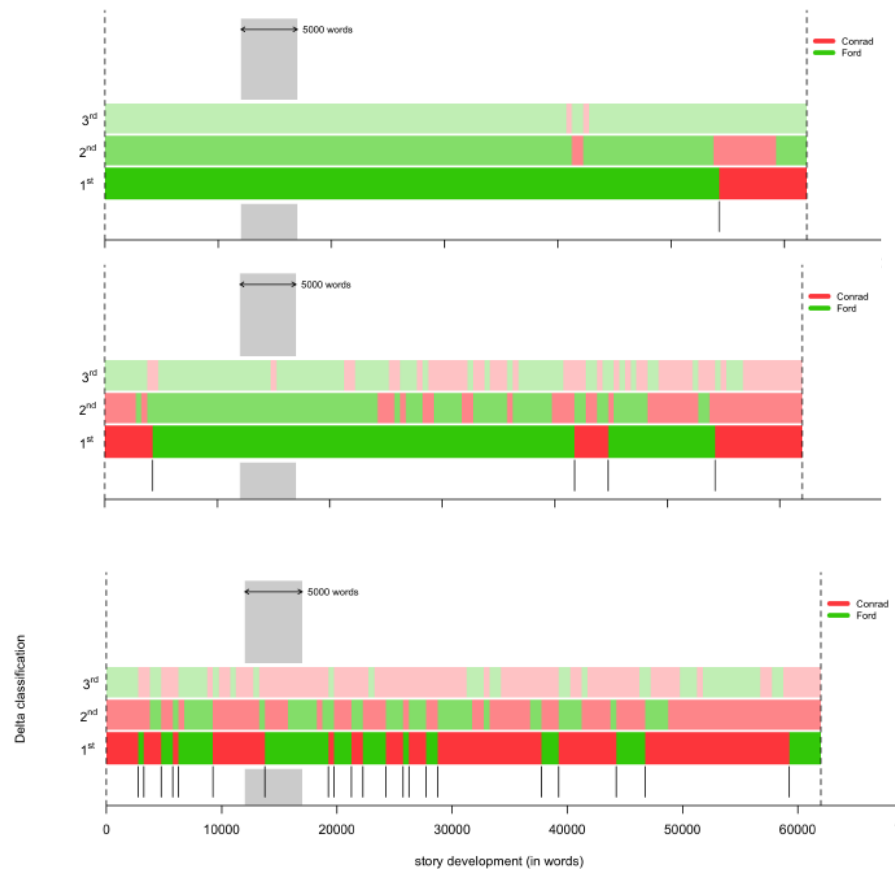


# Replication (*Inheritors*) by Maciej Eder

- Eder, M. "Rolling stylometry" *Dig. Scholarship in the Humanities* 31.3 (2016): 457-469.
- Similar setup, but varying n MFW's: 1000, 500, 100
- Graph shows author of training document closest style (bottom bar)
- Middle bar – 2<sup>nd</sup> candidate, top bar – 3<sup>rd</sup> candidate

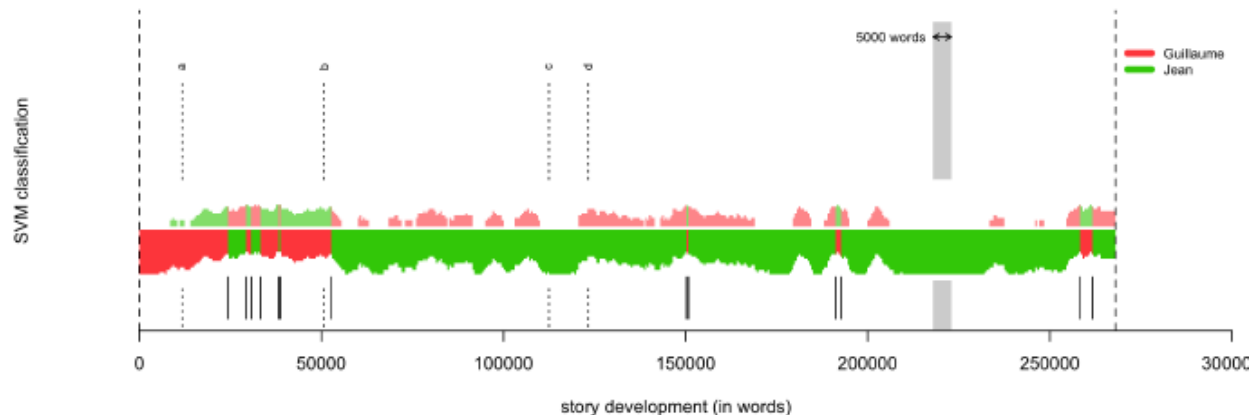


- Bottom bar shows closest style
  - Red (Conrad), green (Ford)
- Top graph 1K MFWs, middle 500, bottom 100
- Top graph confirms Rybicki et al.'s results
- Fewer MFW – noisier results?



# Classifiers other than $\Delta$

- Eder also examines Support Vector Machines (SVM)
- This adds a confidence value for the classification, also visualizable
- *Roman de la rose* 13<sup>th</sup> cent. Guillaume de Lorris or Jean de Meun?

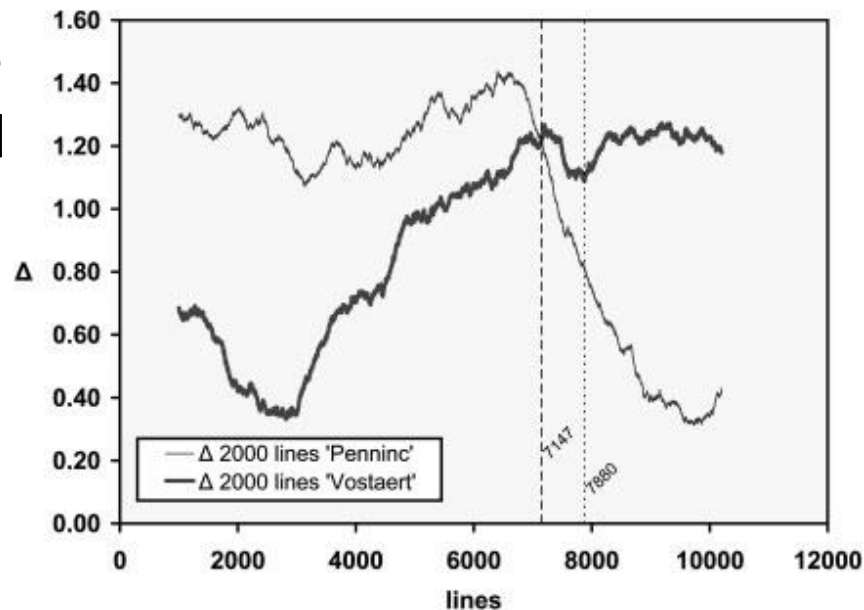


# Medieval epic poem *Walewein*

- LLC paper by van Dalen-Oskam & van Zundert (2007)
- It was known that two authors collaborated on the Middle Dutch epic poem *Walewein* (ca. 1260)
  - Penninc started, Vostaert completed (reported in poem)
- But no reference material was available, so no “rolling  $\Delta$ ”
- vD-O&vZ exploited the report that one author started and the second completed

# Walewein

- vD-O&vZ exploit report that one author started, other completed
  - They used the initial and final sections of the poem as “references”!
  - Comparing via  $\Delta$
- l.7880 old scholarly consensus
  - vD&vZ see earlier break



# Prospects

- Some posthumous novels are “finished” by people who haven’t written too much similar material
  - Case: F.Scott Fitzgerald’s *The Last Tycoon* edited by Edmund Wilson
    - Wilson was a literary critic and a college friend of Fitzgerald’s, but he wrote little fiction himself
  - No reference texts
  - Can rolling delta still apply? –Maybe we can detect breaks in style

## if there's time ....

- It turns out that Stylo does offer a way to find most contrastive features `oppose` (see documentation)
- Assuming that `CorpusH` (containing only Hamilton's texts) and `CorpusM` (containing only Madison's) are both with the working directory, we can invoke  
`oppose(primary.corpus.dir="CorpusH", secondary.corpus.dir="CorpusM")`
- This seeks words that contrast based on Craig's zeta measure



# Questions from class (2022)

- Is there work on stylometry & gender?
  - Verhoeven et al. (2016) "Twisty: a multilingual twitter stylometry corpus for gender and personality profiling." *Proc 10th LREC*.
- Have deep learning techniques been applied?
  - Emmery et al. (2021) "Adversarial stylometry in the wild: Transferable lexical substitution attacks on author profiling." *arXiv preprint arXiv:2101.11310* Also EACL. 2022, 2388-2402

# Craig's $\zeta$ (zeta)

- We contrast one author with another, Hamilton vs Madison
- We first partition all the documents into equal-sized slices
- For a feature/word  $i$  we count the number of slices it appears in to obtain a DOCUMENT PROPORTION:
  - $dp_i(H) = \frac{df_i(H)}{n(H)}$ , where  $df_i(H)$  is number of slices where  $i$  occurs,  $n(H)$  is the number of slices
- $\zeta(i) = dp_i(H) - dp_i(M)$
- Emphasis on regular occurrence, not absolute frequency

- Hamilton's papers are the reference
- *upon* is most strongly preferred
- *former* deprecated
- Not based on **total** number of occurrences, but rather on regular occurrence throughout document set ( $\zeta$ )

