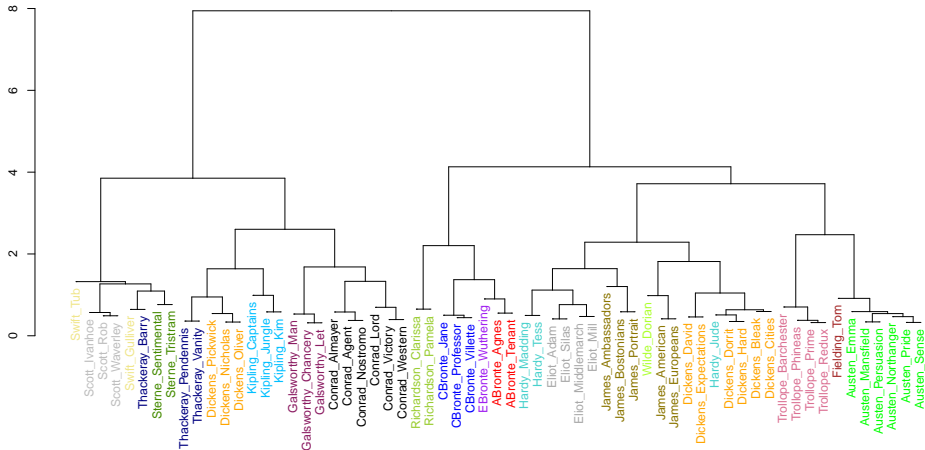


Stylometry and gender, or mind the words you use

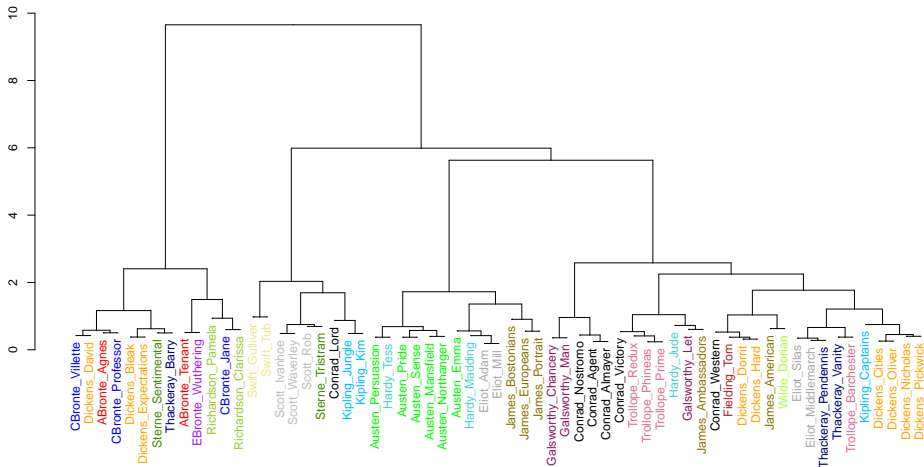
Maciej Eder

Pedagogical University in Krakow
& Polish Academy of Sciences

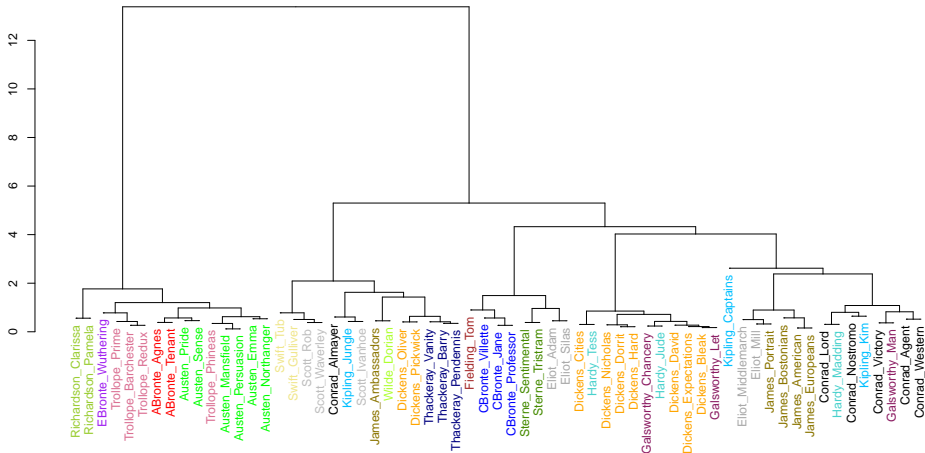
30 MFW: only three authors



I me my you your yours they their he his him she her



6 words: the a an no not nor



Simple yet effective comparisons

- Keywords analysis

- given a text and a comparison corpus,...
- ... the goal is to analyze particular word frequencies
- keyword = a word substantially more frequent in a given text

- Zeta method

- an extended version of keywords analysis

Example: *Gulliver's Travels* vs. Victorian novels

Swift's keywords:

majesty Yahoos emperor country their several Houyhnhnms Yahoo
island my king Glumdalclitch of whereof the master queen which
court Blefuscu feet hundred author they prince three by wherein
great sea although observed ...

Swift's negative keywords:

you her she Mr will your said Mrs he have am say dear now know
it him think woman is if says Sir not heart lady can father do thee
man herself has mother it ...

Example: *Gulliver's Travels* vs. Victorian novels

Swift's keywords:

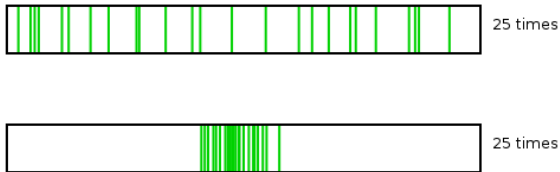
majesty Yahoos emperor country **their** several Houyhnhnms Yahoo
island **my** king Glumdalclitch **of** whereof **the** master queen which
court Blefuscu feet hundred author **they** prince three **by** wherein
great sea although **observed** ...

Swift's negative keywords:

you **her** **she** Mr will **your** said Mrs **he** have am say dear now know
it **him** think woman is if says Sir not heart lady can **father** do thee
man **herself** has **mother** it ...

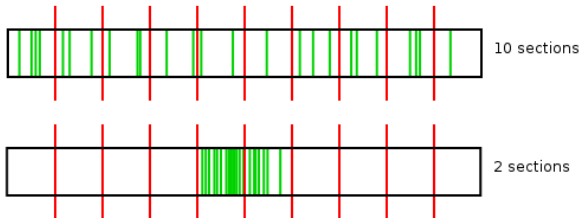
simple world ($I \neq$ they) vs. **dense social network** (you, she, him, ...)

Zeta method (originally developed by Burrows)

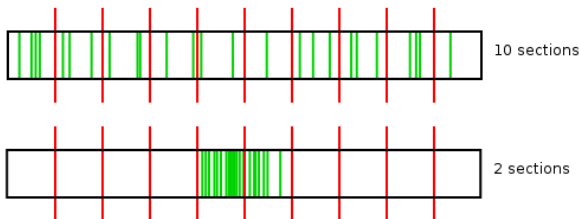


bare number of occurrences does not say the whole truth

Looking for differently distributed words



Looking for differently distributed words



$$\zeta_{(a,b)} = \left(\frac{f_{(a)} - f_{(b)}}{100} \right) + 1$$

$$\zeta_{(a,b)} = \frac{f_{(a)} - f_{(b)}}{f_{(a)} + f_{(b)}}$$

Example: Do they really differ?

Women:

A. Brontë (*Agnes Grey*, *The Tenant of Wildfell Hall*)

C. Brontë (*Jane Eyre*, *The Professor*, *Villette*)

E. Brontë (*Wuthering Heights*)

Austen (*Emma*, *Pride and Prejudice*, *Sense and Sensibility*)

Eliot (*Adam Bede*, *Middlemarch*, *The Mill on the Floss*)

Men:

Dickens (*Bleak House*, *David Copperfield*, *Hard Times*)

Fielding (*Joseph Andrews*, *Tom Jones*)

Richardson (*Clarissa*, *Pamela*)

Thackeray (*Barry Lyndon*, *The History of Pandennis*, *Vanity Fair*)

Trollope (*Barchester Towers*, *Phineas Finn*, *The Prime Minister*)

Women are from Venus...

Female words:

feelings glance effort paused feeling surprise noticed pause
consciousness dared enjoyment tone listen exclaimed features
seated continually anxiety solitude inward apparently painful
entrance respectable relief closed watching feel bent peculiar rain
suddenly cheerful clear trees aspect watched plan slight doubtless
reached smile brow vague quiet mere movement gathered suffering
entered listened observed warm exertion minutes change ...

Male words:

story although lord bosom honour honest duke parliament city
score enemy coach coat inn thousand breast bill dozen lordship
guilty court ain legs bottle captain fight pen battle sum
nevertheless reader virtue order innocent condition infinite castle
widow england accident readers laws fellows hundred service king
stories persons ladyship fly street dearest honours member fortune
government wig drank papers wretch described honourable pocket

Women are from Venus...

Female words:

feelings glance effort paused feeling surprise noticed pause
consciousness dared enjoyment tone listen exclaimed features
seated continually anxiety solitude inward apparently painful
entrance respectable relief closed watching feel bent peculiar rain
suddenly cheerful clear trees aspect watched plan slight doubtless
reached smile brow vague quiet mere movement gathered suffering
entered listened observed warm exertion minutes change ...

Male words:

story although lord bosom honour honest duke parliament city
score enemy coach coat inn thousand breast bill dozen lordship
guilty court ain legs bottle captain fight pen battle sum
nevertheless reader virtue order innocent condition infinite castle
widow england accident readers laws fellows hundred service king
stories persons ladyship fly street dearest honours member fortune
government wig drank papers wretch described honourable pocket

Women are from Venus...

Female words:

feelings glance effort **paused** feeling surprise **noticed** pause
consciousness **dared** enjoyment tone **listen** **exclaimed** features
seated continually anxiety solitude inward apparently painful
entrance respectable relief **closed** watching **feel** bent peculiar rain
suddenly cheerful clear trees aspect **watched** plan slight doubtless
reached smile brow vague quiet mere movement **gathered** suffering
entered **listened** **observed** warm exertion minutes change ...

Male words:

story although lord bosom honour honest duke parliament city
score enemy coach coat inn thousand breast bill dozen lordship
guilty court ain legs bottle captain fight pen battle sum
nevertheless reader virtue order innocent condition infinite castle
widow england accident readers laws fellows hundred service king
stories persons ladyship fly street dearest honours member fortune
government wig drank papers wretch **described** honourable pocket

Chawton House, formerly owned by J. Austen's brother



Chawton House corpus of women's writings

- Dates: 1723–1830
- 12 anonymous novels
- 34 novels by women known by name
 - 5 authors of 2 novels
 - 24 authors of 1 novel

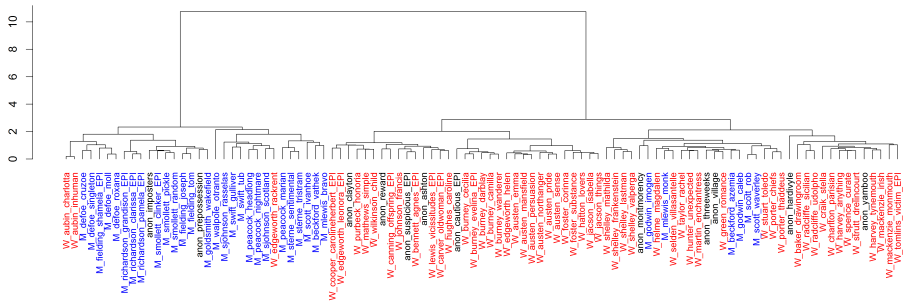
comparison corpus:

- 22 novels written by women (Austen, Radcliffe, Shelley, ...)
- 33 novels written by men (Defoe, Swift, Johnson, Sterne, ...)

THE words used as input data

	ABronte <i>Agnes</i>	Austen <i>Emma</i>	CBronte <i>Jane</i>	Conrad <i>Lord</i>	Dickens <i>Bleak</i>	...
"feelings"	0.0307	0.0636	0.0257	0.0077	0.0084	...
"anxiety"	0.0102	0.0118	0.0043	0.0061	0.0053	...
"conscious"	0.0000	0.0069	0.0032	0.0031	0.0045	...
⋮	⋮	⋮	⋮	⋮	⋮	...
"thousand"	0.0073	0.0100	0.0128	0.0054	0.0107	...
"honour"	0.0073	0.0168	0.0053	0.0138	0.0219	...
⋮	⋮	⋮	⋮	⋮	⋮	...

Chawton House in Context Men Women Cluster Analysis



Thank you!