Stylometria: od atrybucji autorskiej do wielkoskalowej analizy tekstu

Maciej Eder

Uniwersytet Pedagogiczny w Krakowie
& Instytut Języka Polskiego PAN

8.06.2015

Akt I

Atrybucja autorska: podstawowe pojęcia

Nietradycyjna atrybucja autorska

Jeśli mamy:

- tekst o nieznanym autorstwie
- korpus porównawczy zawierający teksty znanych autorów

to czy można wśród nich odnaleźć "najbliższego sąsiada", czyli tekst najbardziej podobny spośród dostępnych w korpusie?

Gdzie można szukać stylistycznego "odcisku palca":

słowa synsemantyczne, najczęstsze słowa, pary słów, częste sekwencje literowe...

Co należy analizować? *Gdzie* jest styl?

- Idea stylistycznego odcisku palca
  - cechy języka niedostrzegalne gołym okiem
  - poza kontrolą autora (nieświadome)
  - odporne na imitację, parodię itp.
  - chętnie wybierany znacznik: częstości najczęstszych słów
- Czy styl jest jednoznacznie zależny od autora? (cf. kod DNA, odcisk palca, wzór na źrenicy, ...)

Stylistyczny odcisk palca

'If handwriting can be so exactly determined as to afford certainty as to its identity, so also with style, since style is more personal and characteristic than handwriting'

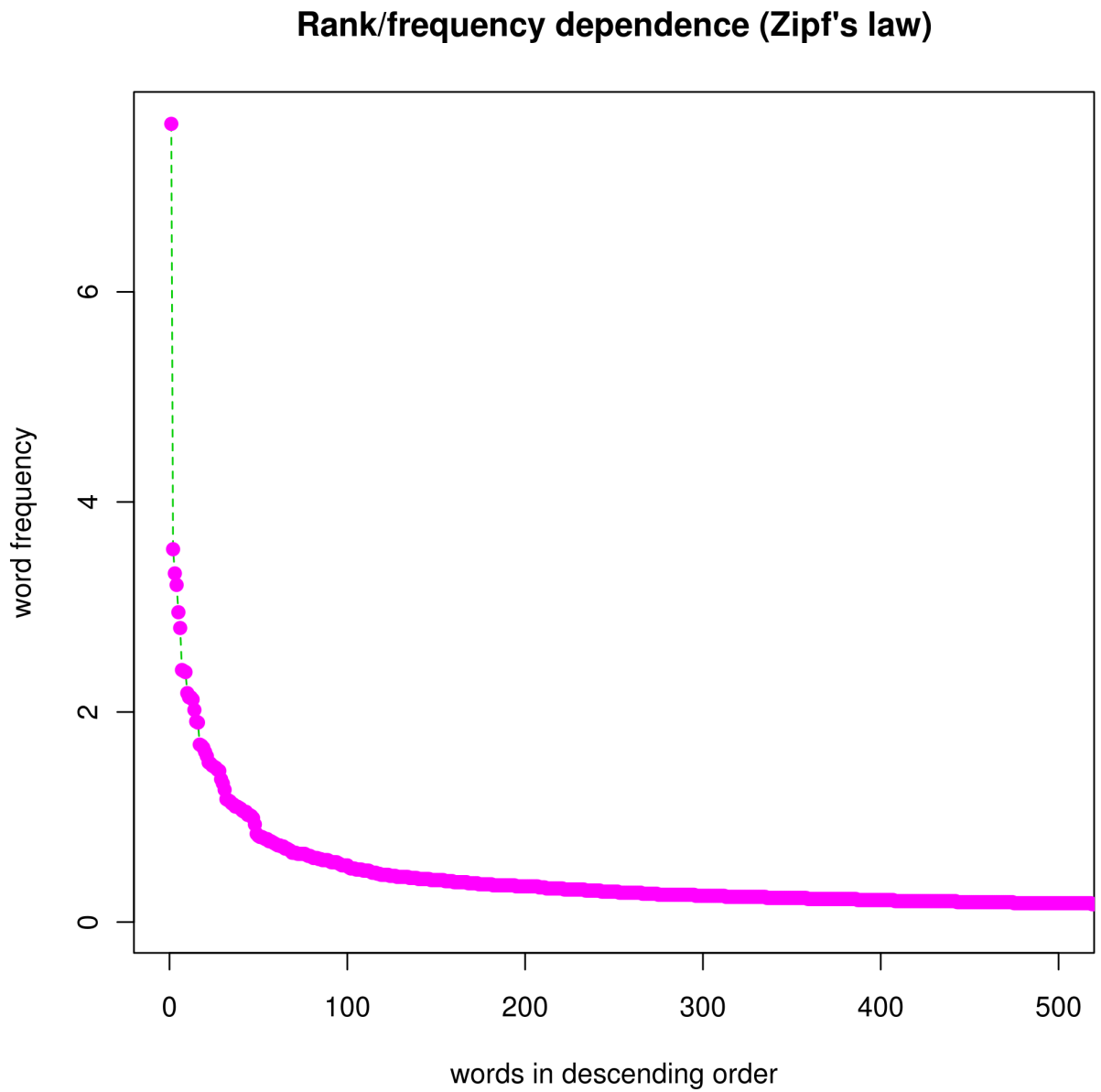
(Lutosławski 1897: 66)

Dziś jednak nikt nie wierzy w jednoznacznie niepowtarzalny odcisk palca; szuka się raczej statystycznie prawdopodobnych regularności

(1) Sterne, *Sentimental*, (2) Hor. *Ars*, (3) *Bartachom.*

1.	the	5.7384	et	4.69104	δ’	4.3154
2.	i	3.2138	non	1.26173	καὶ	3.09004
3.	and	3.1292	si	1.22938	ἐς	2.18434
4.	of	3.0518	in	1.13232	δὲ	1.97123
5.	to	2.7640	aut	0.905856	ἦν	1.86468
6.	a	2.4061	qui	0.841152	ὁ	1.17208
7.	it	1.9104	ut	0.776448	οὐ	1.06553
8.	in	1.8838	quid	0.744096	ἐπὶ	1.01225
9.	had	1.1583	nec	0.711744	τὸν	0.745871
10.	was	1.1462	est	0.711744	κατὰ	0.745871
11.	as	1.1100	an	0.420576	τε	0.692595
12.	my	1.0834	ad	0.420576	ἀπ’	0.639318
13.	his	1.0108	quae	0.388224	μὲν	0.639318
14.	he	0.9673	ego	0.388224	ἐπ’	0.586042
...	...	...	...	...	...	...

Prawo Zipfa


Rank/frequency dependence (Zipf's law)
word frequency
0
2
4
6
0
100
200
300
400
500
words in descending order

Gdzie stylometria (na ogół) się zaczyna: tabela frekwencji

	Milton *Samson*	Milton *Paradise*	Keats *Hyperion*	Patmore *Eros*	Browning *Bishop*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Gdzie stylometria (na ogół) się zaczyna: tabela frekwencji

	Milton *Samson*	Milton *Paradise*	Keats *Hyperion*	Patmore *Eros*	Browning *Bishop*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

. . . . ?

Gdzie stylometria (na ogół) się zaczyna: tabela frekwencji

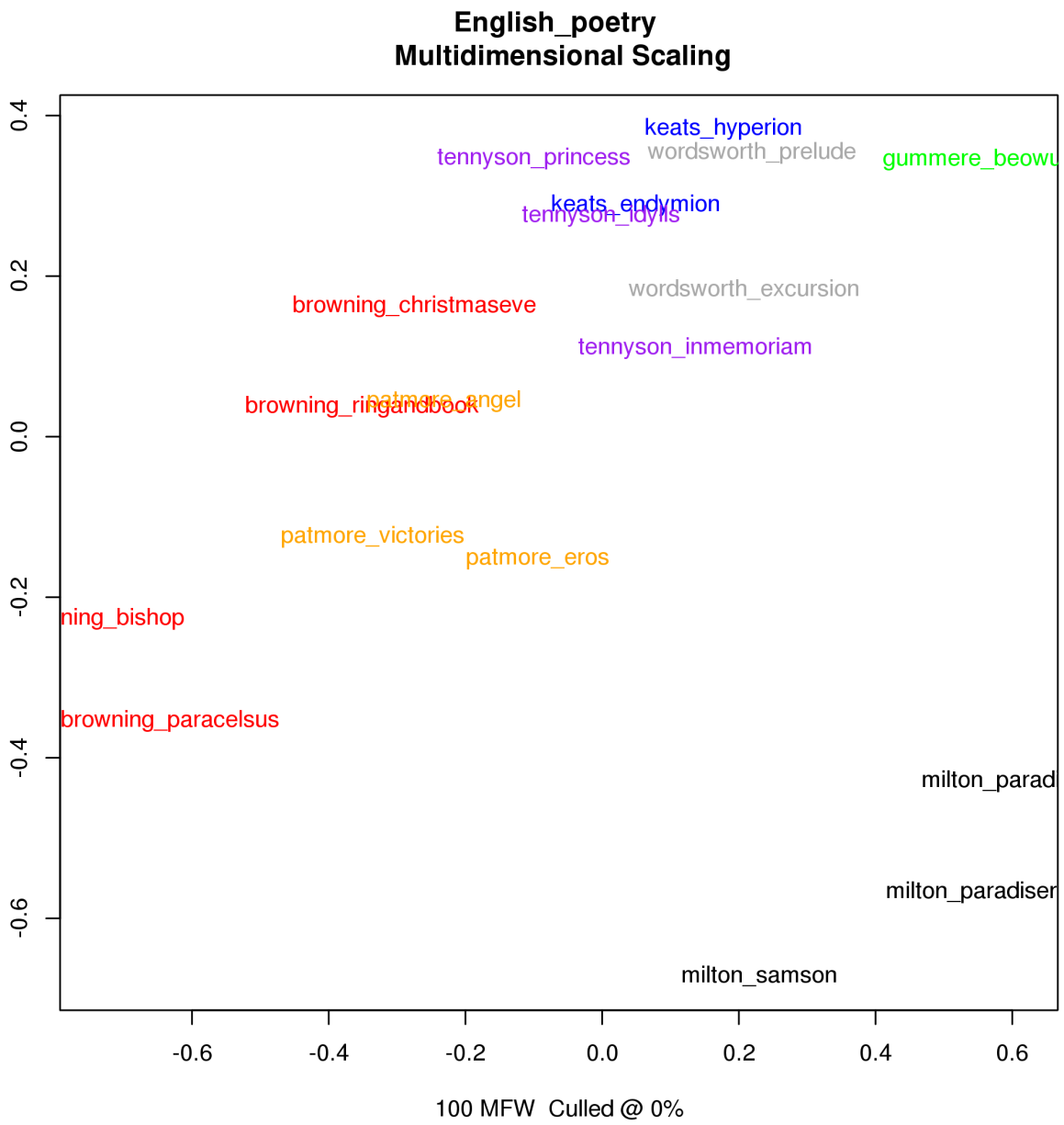
	Milton *Samson*	Milton *Paradise*	Keats *Hyperion*	Patmore *Eros*	Browning *Bishop*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

. . . . ? . . . . . ???

Tabela odległości między próbkami

	Milton *Samson*	Milton *Paradise*	Keats *Hyperion*	Patmore *Eros*	Browning *Bishop*	...
Milton *Samson*	0	0.9839	1.12	1.0493	1.0864	...
Milton *Paradise*	0.9839	0	1.0891	1.089	1.1047	...
Keats *Hyperion*	1.12	1.0891	0	1.128	1.11	...
Patmore *Eros*	1.0493	1.089	1.128	0	1.1128	...
Browning *Bishop*	1.0864	1.1047	1.11	1.1128	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Metody nienadzorowane: skalowanie wielowymiarowe



Znaczniki stylu: czy tylko MFWs?

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Znaczniki stylu: czy tylko MFWs?

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Co właściwie powinniśmy mierzyć:

Znaczniki stylu: czy tylko MFWs?

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Co właściwie powinniśmy mierzyć: najczęstsze słowa?

Znaczniki stylu: czy tylko MFWs?

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Co właściwie powinniśmy mierzyć: najczęstsze słowa? słowa synsemantyczne?

Znaczniki stylu: czy tylko MFWs?

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"the"	4.57	4.24	4.25	4.19	4.47	...
"to"	3.11	3.29	3.43	3.14	3.71	...
"and"	3.19	3	3.08	2.85	2.81	...
"of"	2.6	3	2.63	2.43	2.86	...
"I"	2.17	2.2	2.13	2.42	2.22	...
"a"	2.24	1.92	1.92	2.21	1.92	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Co właściwie powinniśmy mierzyć: najczęstsze słowa? słowa synsemantyczne? inne elementy języka?

Przeróżne znaczniki stylu: przykład

‘It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.’

(J. Austen, *Pride and Prejudice*)

Przeróżne znaczniki stylu: przykład

pojedyncze słowa:
‘it’ ‘is’ ‘a’ ‘truth’ ‘universally’ ‘acknowledged’ ‘that’ ‘a’ . . .

leksykalne bi-gramy:
‘it-is’ ‘is-a’ ‘a-truth’ ‘truth-universally’ ‘universally-acknowledged’
. . .

literowe bi-gramy:
‘it’ ‘t-’ ‘-i’ ‘is’ ‘s-’ ‘-a’ ‘a-’ ‘-t’ ‘tr’ ‘ru’ ‘ut’ ‘th’ ‘h-’ ‘-u’ . . .

leterowe tri-gramy:
‘it-’ ‘t-i’ ‘-is’ ‘is-’ ‘s-a’ ‘-a-’ ‘a-t’ ‘-tr’ ‘tru’ ‘rut’ ‘uth’ ‘th-’ . . .

literowe heksa-gramy:
‘it-is-’ ‘t-is-a’ ‘-is-a-’ ‘is-a-t’ ‘s-a-tr’ ‘-a-tru’ ‘a-trut’ ‘-truth’ . . .

Tabela frekwencji (literowe tri-gramy)

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"-th"	1.4901	1.3523	1.5002	1.8030	1.5110	...
"the"	1.2484	1.0411	1.1717	1.4986	1.1899	...
"he-"	0.9835	1.0204	1.1500	1.4710	1.0677	...
"-an"	0.8949	0.7362	0.8166	0.6324	0.8086	...
"nd-"	0.8195	0.6858	0.8380	0.6497	0.8069	...
"er-"	0.6849	0.7428	0.6683	0.4968	0.6282	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Tabela frekwencji (literowe tri-gramy)

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"-th"	1.4901	1.3523	1.5002	1.8030	1.5110	...
"the"	1.2484	1.0411	1.1717	1.4986	1.1899	...
"he-"	0.9835	1.0204	1.1500	1.4710	1.0677	...
"-an"	0.8949	0.7362	0.8166	0.6324	0.8086	...
"nd-"	0.8195	0.6858	0.8380	0.6497	0.8069	...
"er-"	0.6849	0.7428	0.6683	0.4968	0.6282	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Niebanalne pytanie: Ile jednostek językowych powinniśmy brać pod uwagę?

Tabela frekwencji (literowe tri-gramy)

	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"-th"	1.4901	1.3523	1.5002	1.8030	1.5110	...
"the"	1.2484	1.0411	1.1717	1.4986	1.1899	...
"he-"	0.9835	1.0204	1.1500	1.4710	1.0677	...
"-an"	0.8949	0.7362	0.8166	0.6324	0.8086	...
"nd-"	0.8195	0.6858	0.8380	0.6497	0.8069	...
"er-"	0.6849	0.7428	0.6683	0.4968	0.6282	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Niebanalne pytanie: Ile jednostek językowych powinniśmy brać pod uwagę?

Tabela frekwencji (literowe tri-gramy)

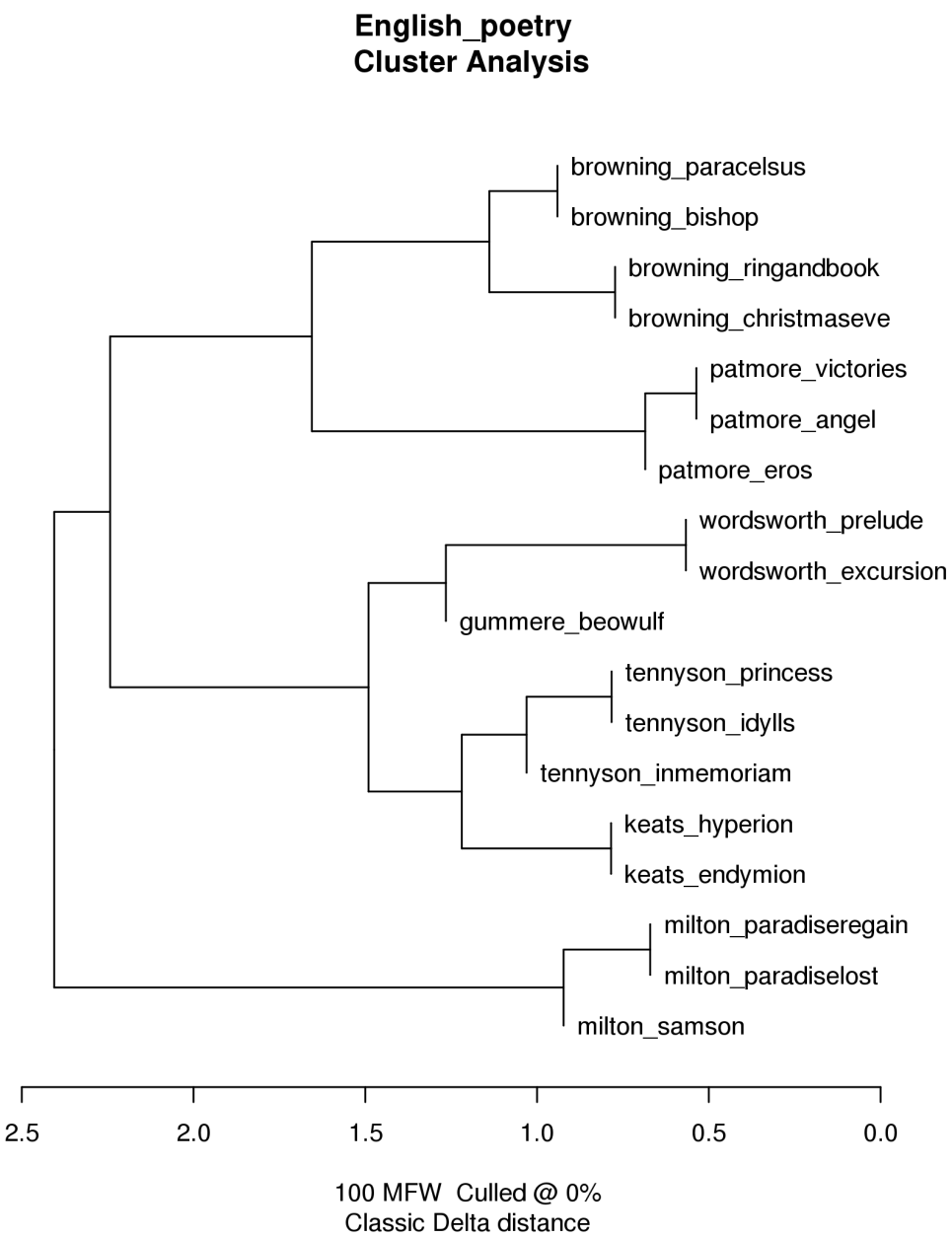
	ABronte *Agnes*	Austen *Emma*	CBronte *Jane*	Conrad *Lord*	Dickens *Bleak*	...
"-th"	1.4901	1.3523	1.5002	1.8030	1.5110	...
"the"	1.2484	1.0411	1.1717	1.4986	1.1899	...
"he-"	0.9835	1.0204	1.1500	1.4710	1.0677	...
"-an"	0.8949	0.7362	0.8166	0.6324	0.8086	...
"nd-"	0.8195	0.6858	0.8380	0.6497	0.8069	...
"er-"	0.6849	0.7428	0.6683	0.4968	0.6282	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Niebanalne pytanie: Ile jednostek językowych powinniśmy brać pod uwagę?

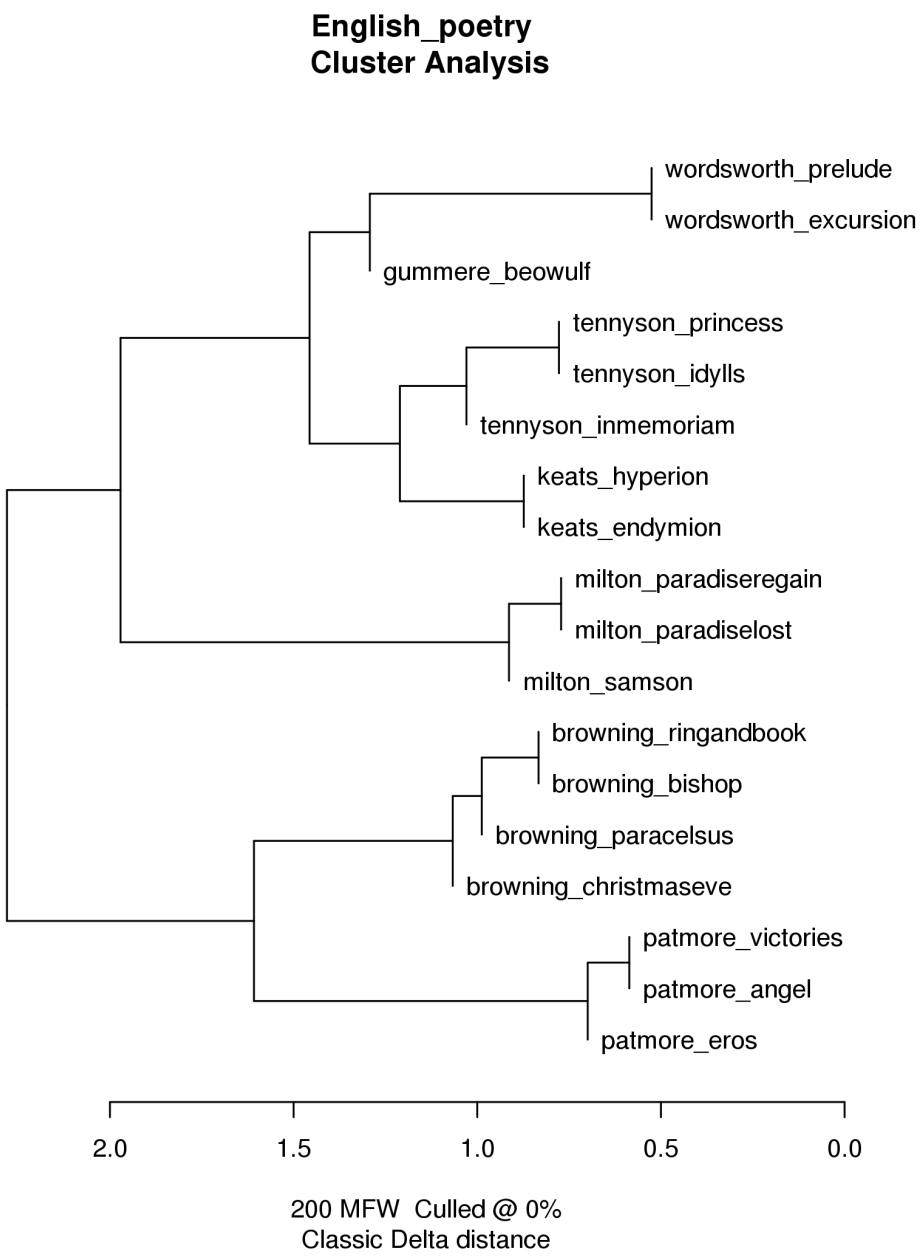
Ile znaczników stylu (np. MFW) powinno się analizować?

- mała liczba słów synsemantycznych (Mosteller & Wallace 1964)
- 50 najczęstszych słów (MFW) lub innych znaczników (Juola 2008)
- 100 najczęstszych słów (Burrows 2002)
- 300 MFW (Smith & Aldridge 2011)
- 500 MFW (Craig & Kinney 2009)
- >1000 MFW (Hoover 2004)
- cała lista frekwencyjna (Juola 2008)
- słowa, które występują u wszystkich autorów (Jockers & Witten 2010)

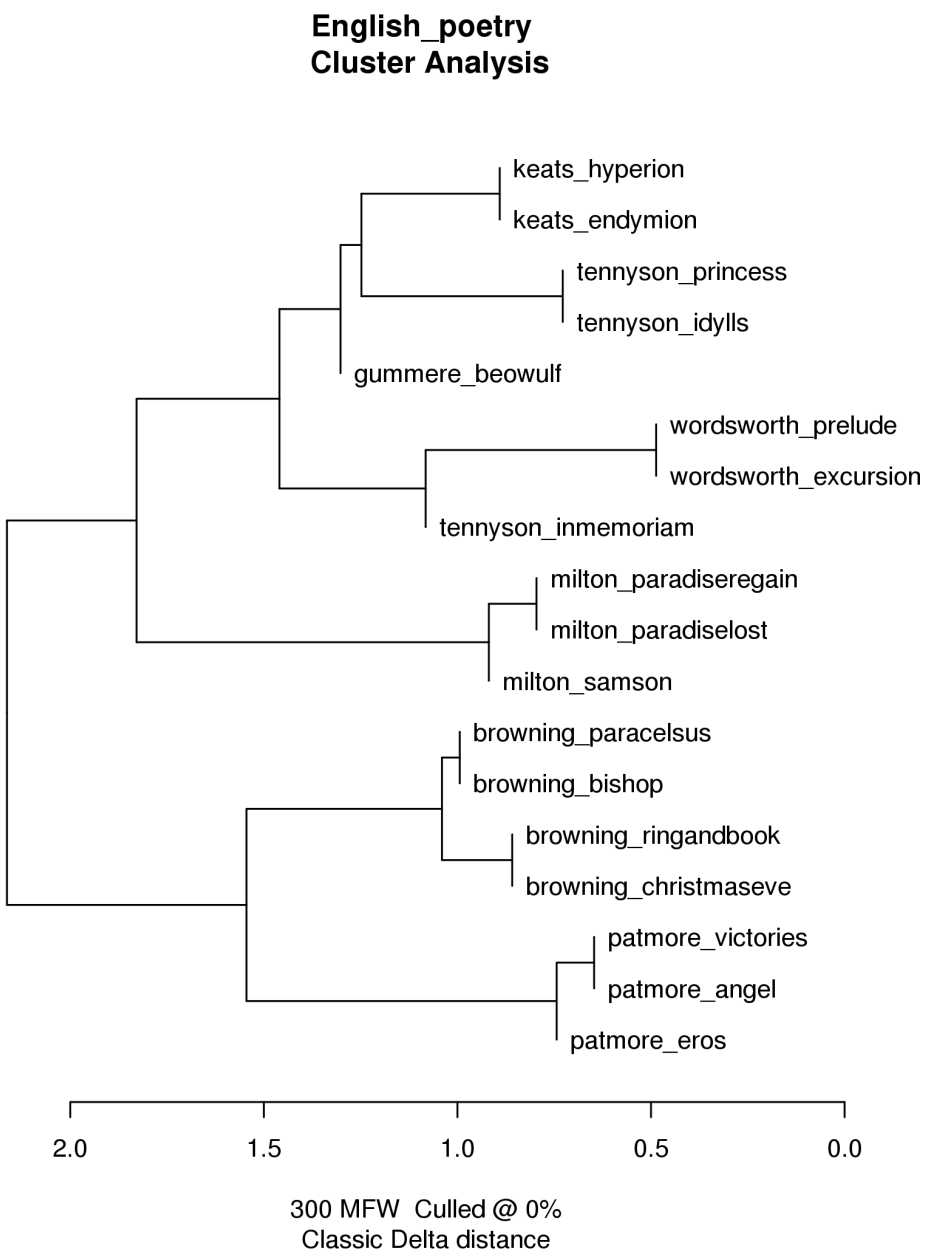
Analiza skupień: 100 MFWs



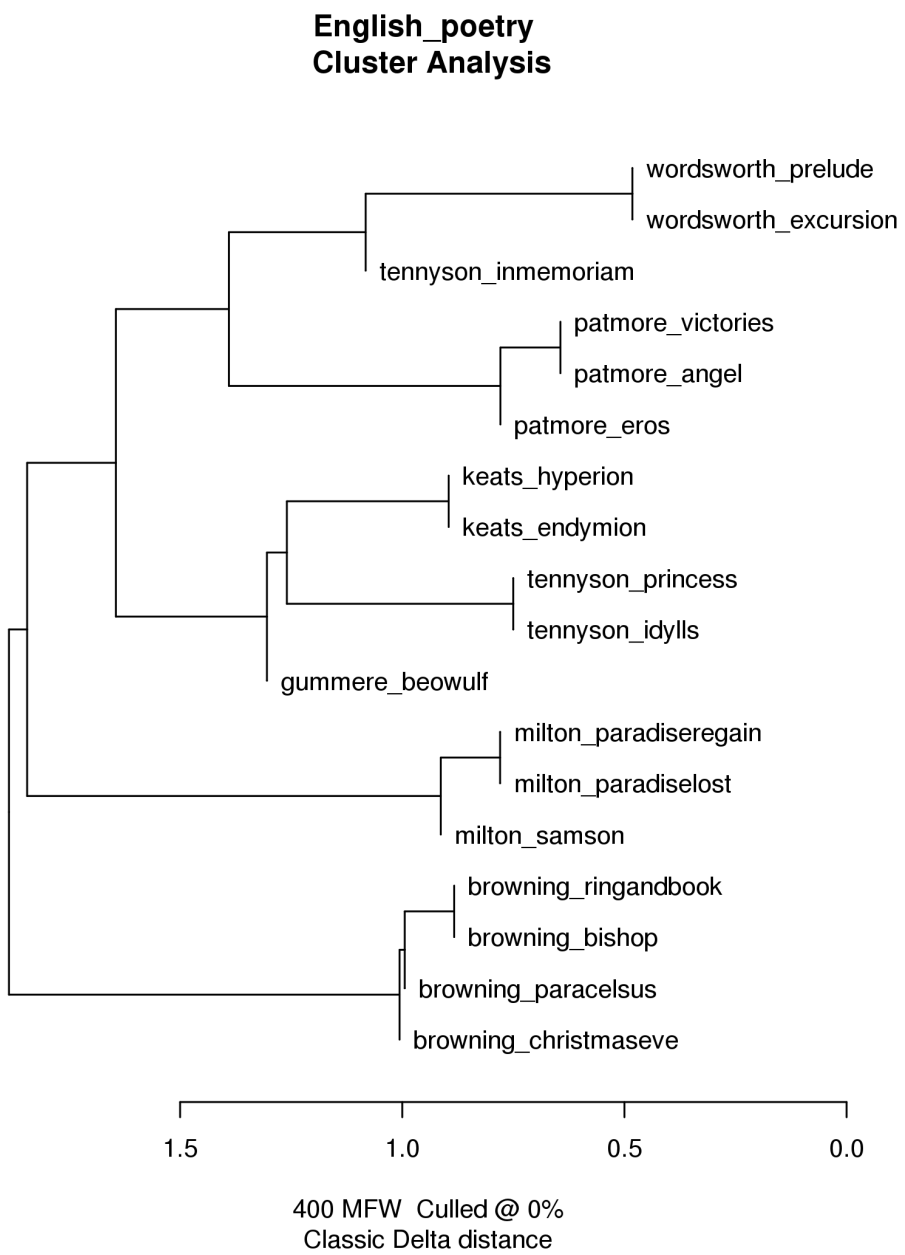
Analiza skupień: 200 MFWs



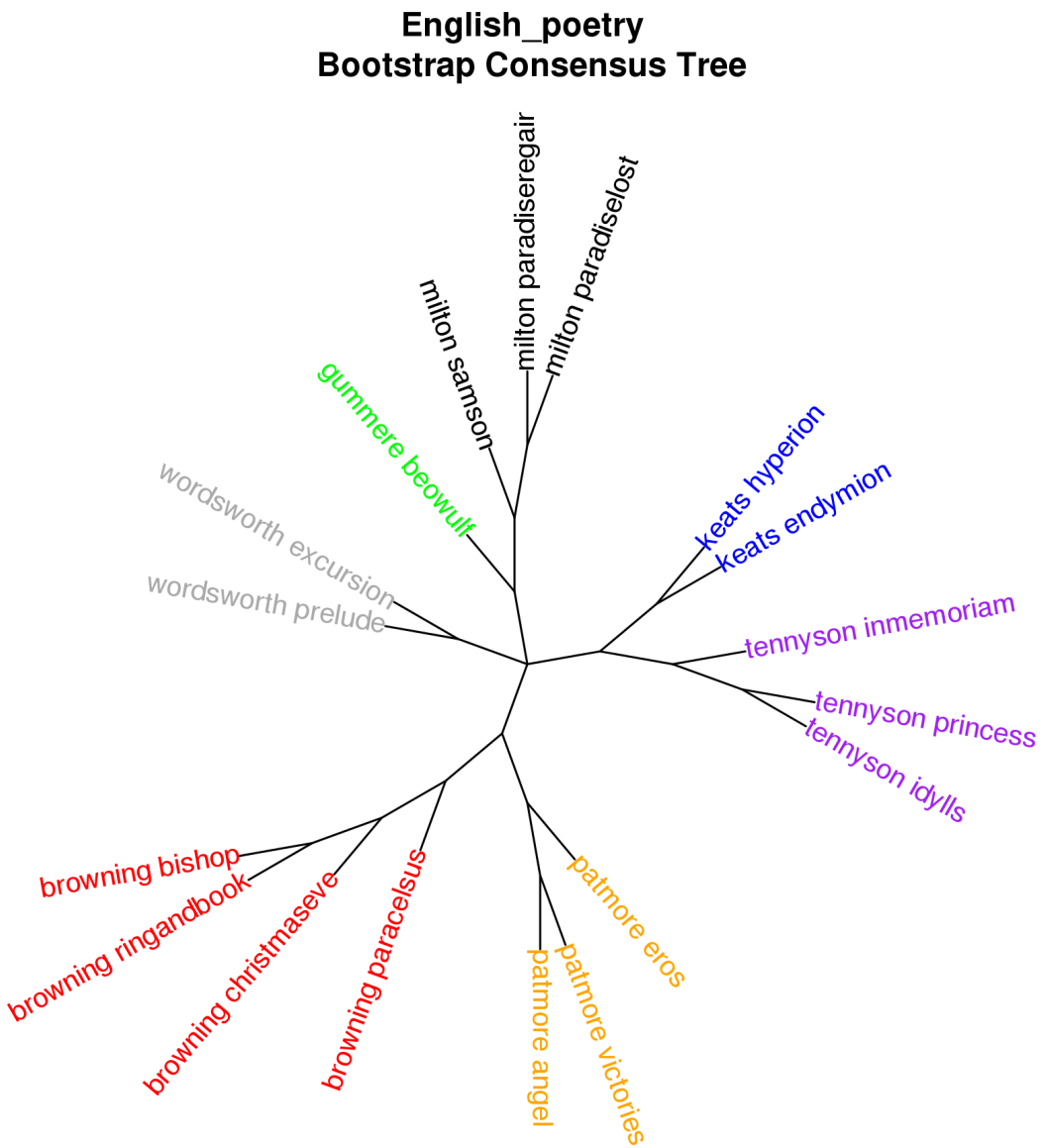
Analiza skupień: 300 MFWs



Analiza skupień: 400 MFWs



Consensus tree: synteza wielu pojedynczych analiz



Akt II

Ajschylos czy łże-Ajschylos?
(problem *Prometeusza skowanego*)

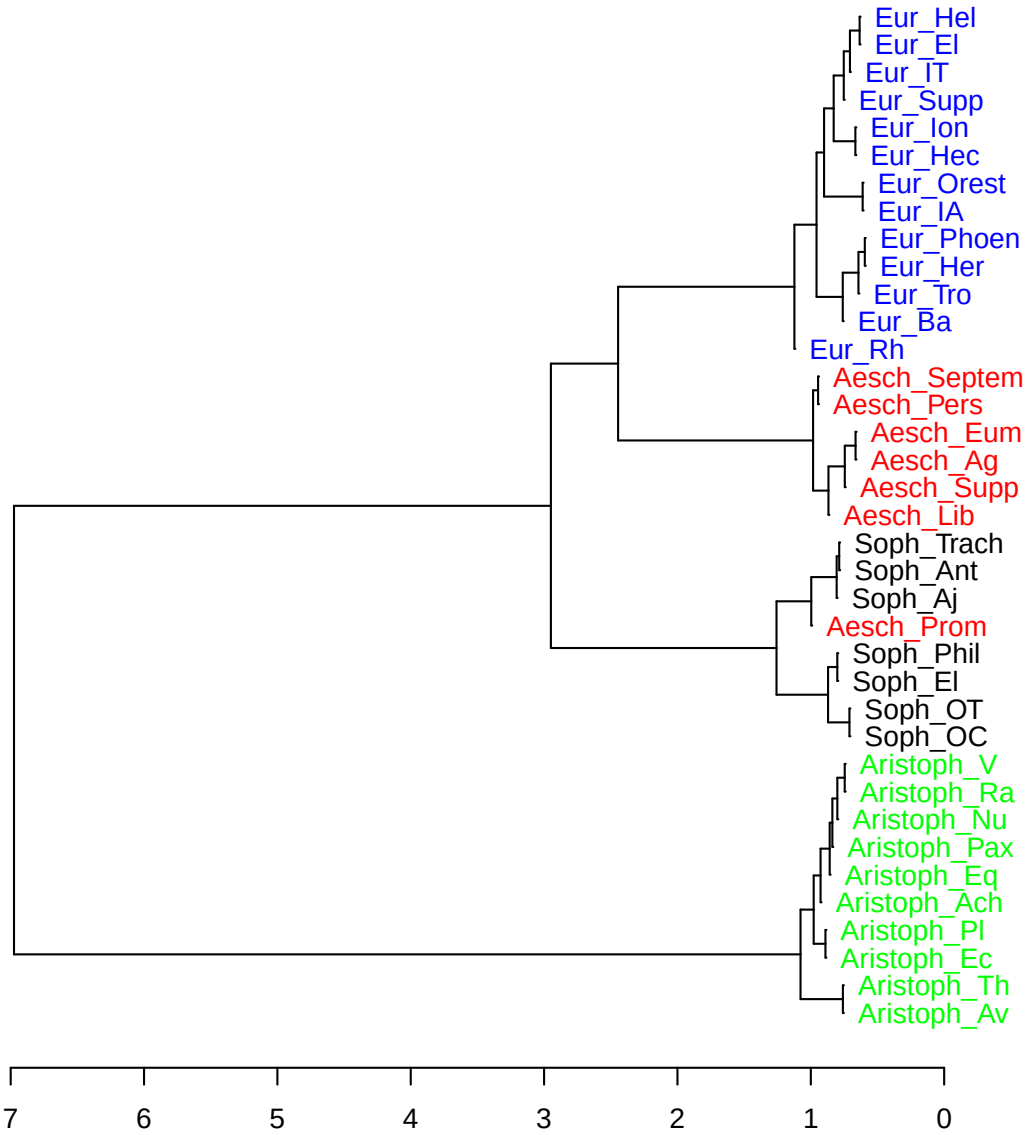
Problemy z autorstwem *Prometeusza skowanego*

- Tak zwany 'problem Zeusa':
  - postawa nabożnej czci we wszystkich tragediach Ajschylosa
  - Zeus odmalowany jako okrutny tyran w *Prometeuszu skowanym*
- Różnice w wersyfikacji między *Prometeuszem* i innymi tragediami
- Trzeci aktor wprowadzony w jednej ze scen

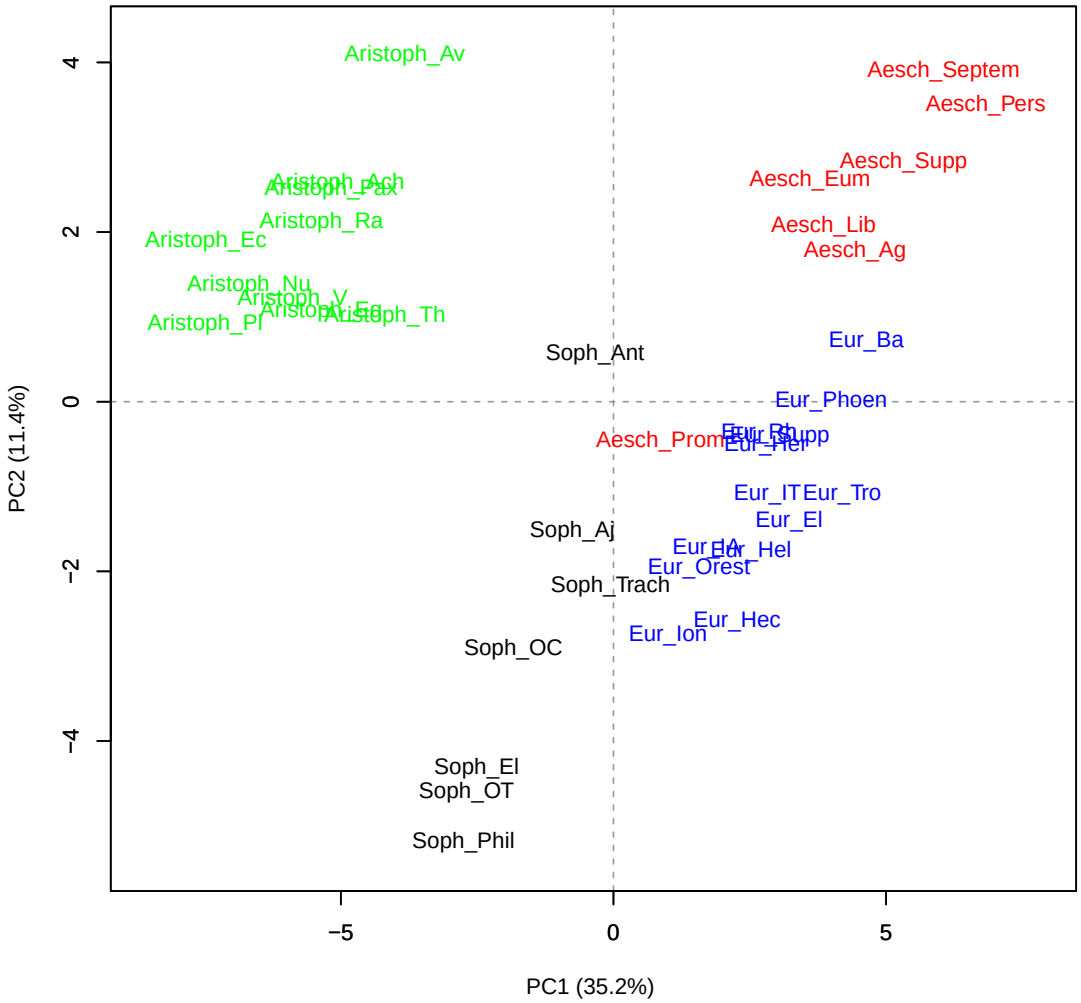
A może jednak Ajschylos?

- Nikt ze starożytnych nie kwestionuje autentyczności
- Aluzje do *Prometeusza* u Arystofanesa

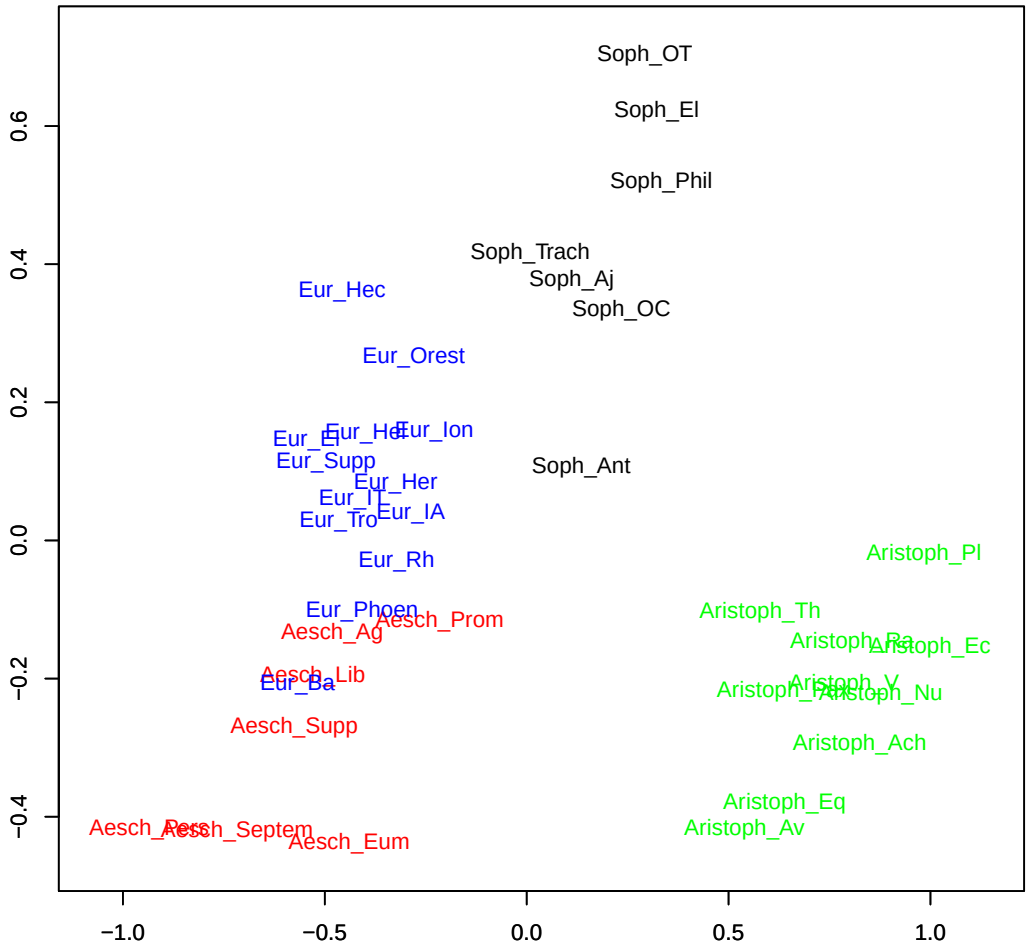
Tragedie i komedie razem



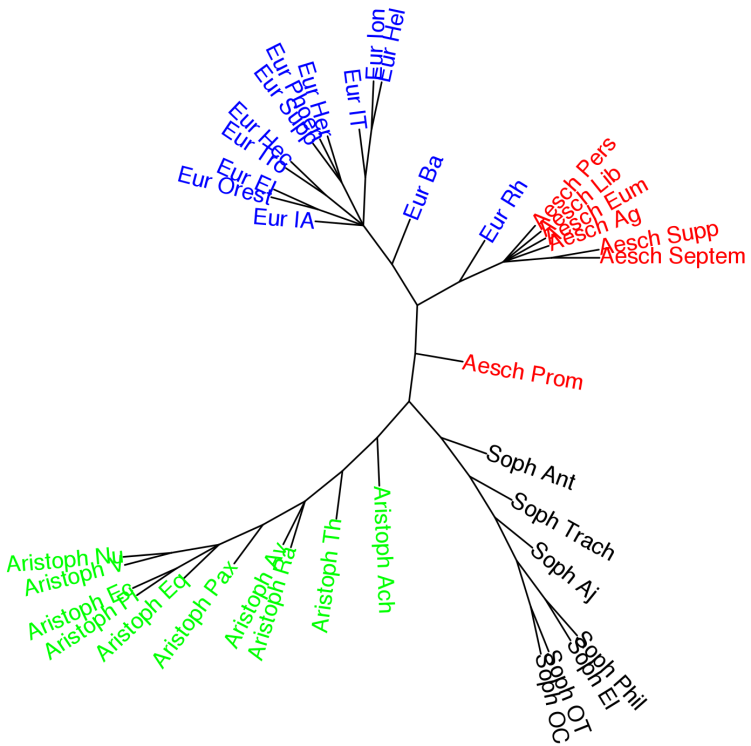
Prometeusz skowany nie jest Ajschylosa!



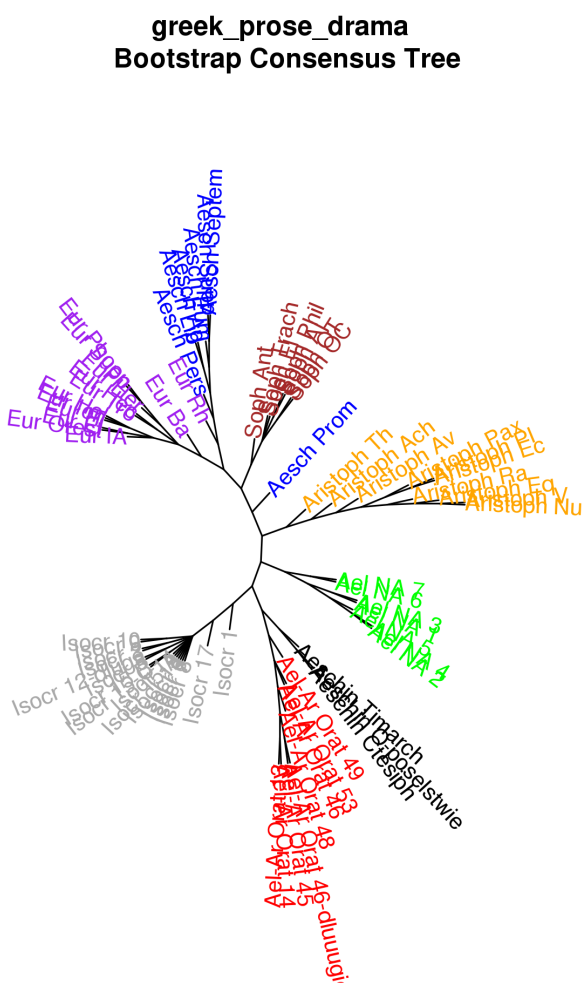
... a może jednak to Ajschylos jest autorem?



Drzewko konsensusu

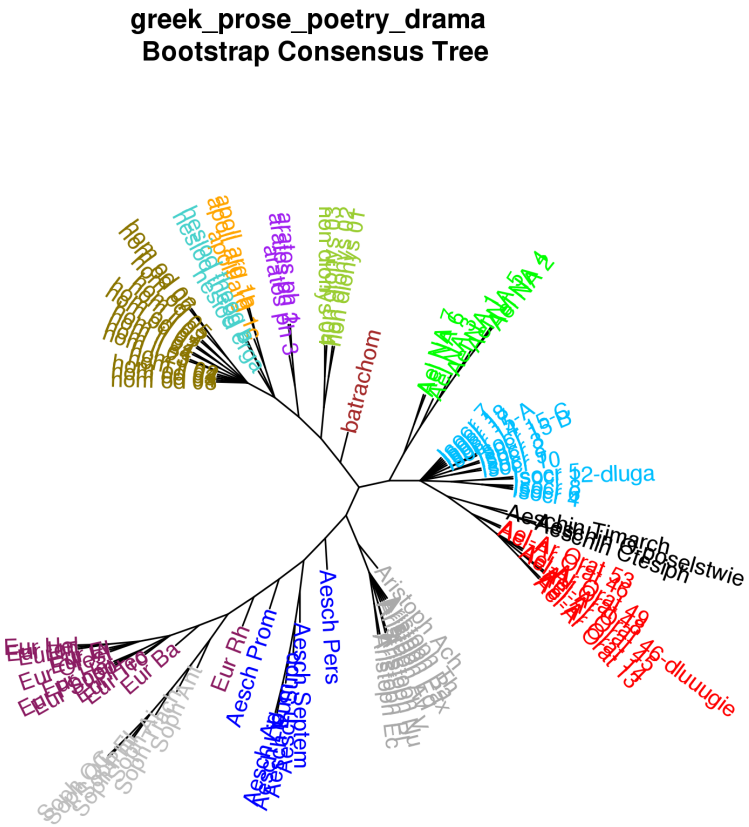


Wprowadzenie dystraktora



43-5000 MFW Culled @ 0-100%
Classic Delta distance Consensus 0.5

Wprowadzenie dystraktora



Metody uczenia maszynowego: znakomite, ale nudne...

```
verg_ecl_1-5 --> verg hor stat (delta score: 1.0988)
verg_ecl_6-10 --> verg stat ov (delta score: 1.0847)
verg_georg_1 --> verg stat hor (delta score: 1.1303)
verg_georg_2 --> stat hor sil (delta score: 1.1204)

500 MFW , culled @ 0%, 18 of 22 (81.8%)

⋮

General attributive success: 143 of 176 (81.2%)
MFWs from 300 to 1000 @ increment 100
Culling from 0 to 0 @ increment 20
Pronouns deleted: FALSE; Classic Delta distance
```

Prometeusz skowany nie chce być Ajschylosa

```
Aesch_Prom --> Eur Soph Eur (delta score: 1.0151)
   100 MFW , culled @ 0%, 27 of 27 (100%)
Aesch_Prom --> Eur Soph Eur (delta score: 1.0151)
   100 MFW , culled @ 0%, 27 of 27 (100%)
Aesch_Prom --> Eur Eur Soph (delta score: 1.0143)
   200 MFW , culled @ 0%, 26 of 27 (96.3%)
Aesch_Prom --> Eur Soph Eur (delta score: 1.0087)
   300 MFW , culled @ 0%, 26 of 27 (96.3%)
Aesch_Prom --> Eur Soph Eur (delta score: 1.0215)
   400 MFW , culled @ 0%, 27 of 27 (100%)
Aesch_Prom --> Eur Soph Eur (delta score: 1.0317)
   500 MFW , culled @ 0%, 27 of 27 (100%)
Aesch_Prom --> Soph Eur Eur (delta score: 1.0379)
   600 MFW , culled @ 0%, 26 of 27 (96.3%)
Aesch_Prom --> Soph Eur Aesch (delta score: 1.018)
```

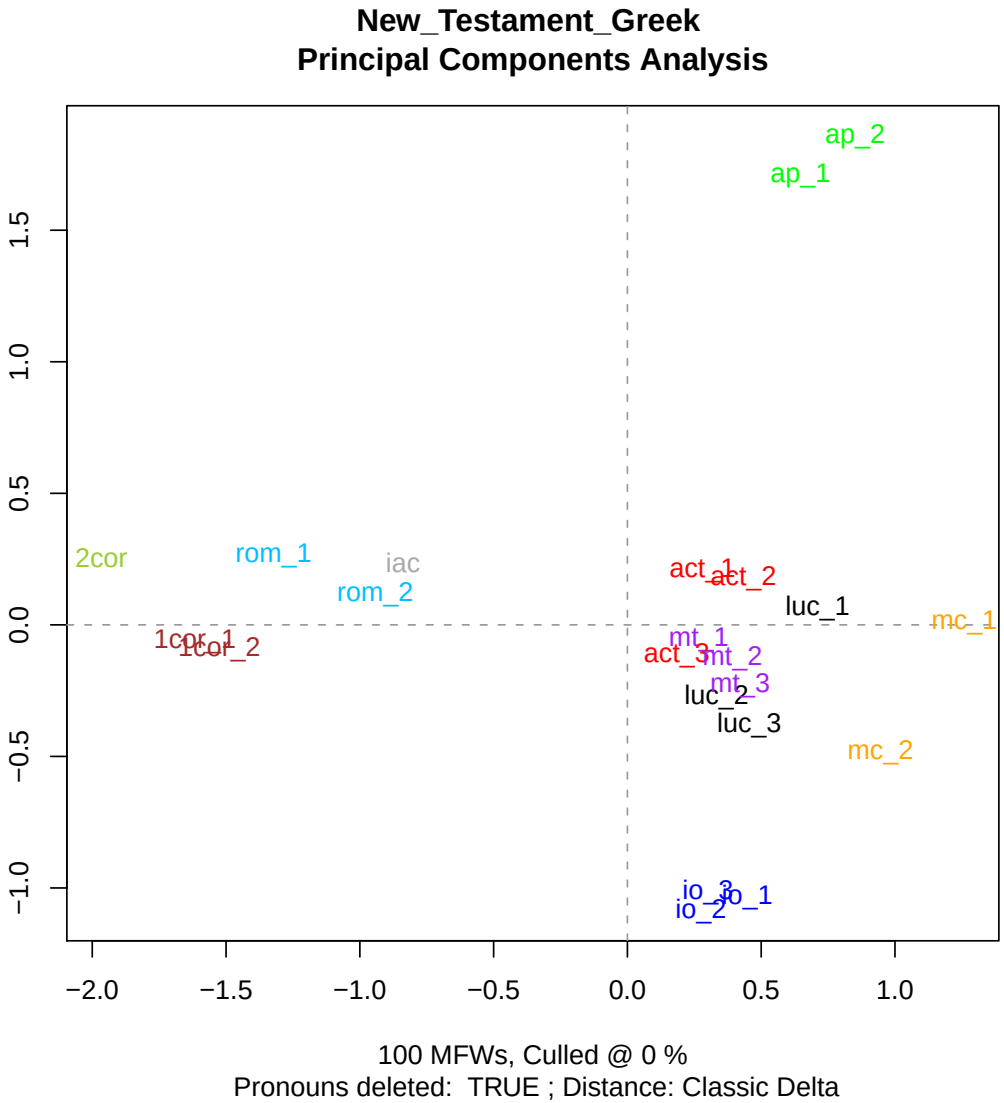
Interludium

Nadzorowane metody statystyczne

Przykład: 20 tekstów pięciu autorów, 1 tekst anonimowy

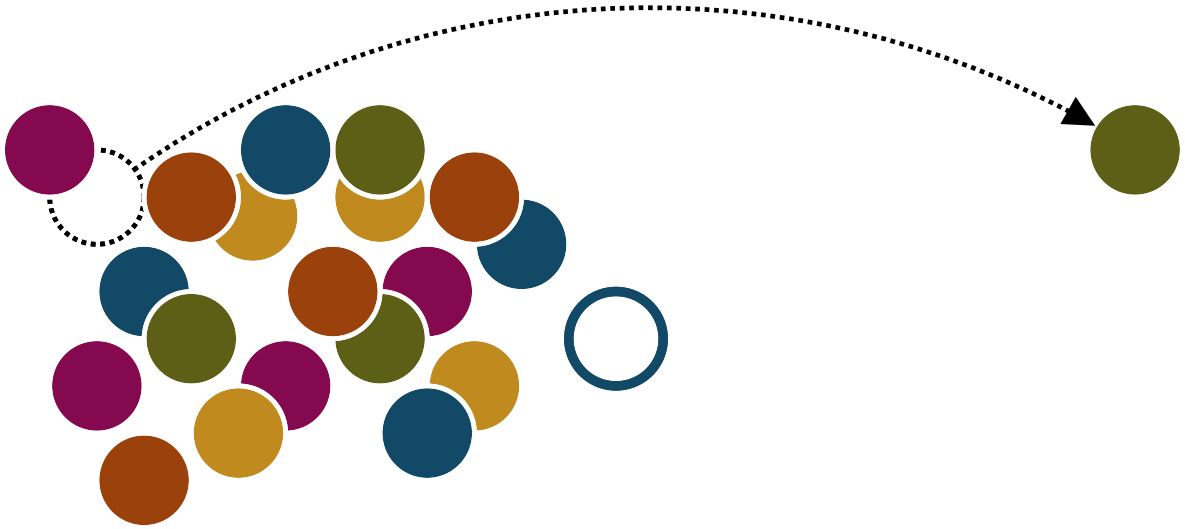
Nienadzorowane metody najbliższych sąsiadów

Metoda nienadzorowana w praktyce

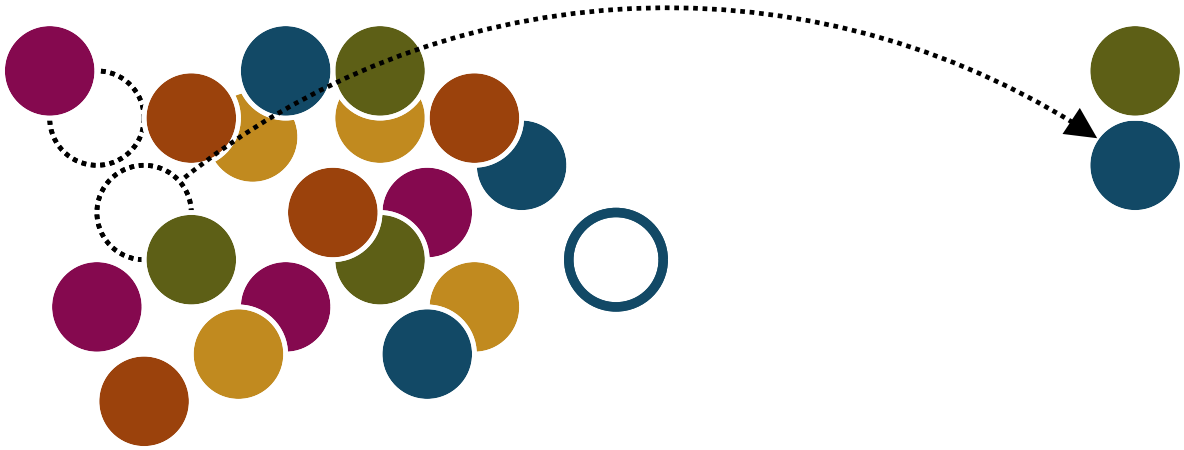


Nadzorowana klasyfikacja najbliższych sąsiadów

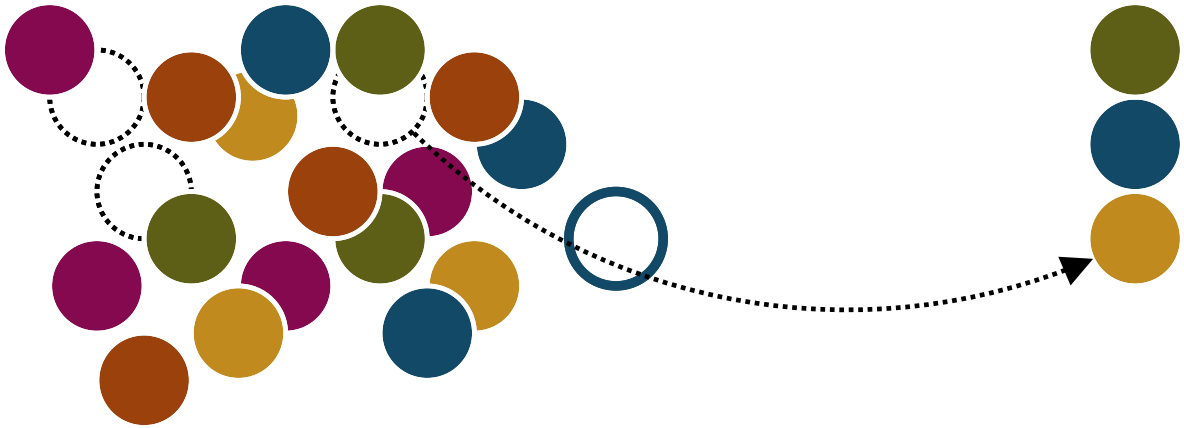
Konstruowanie klasyfikatora (I)



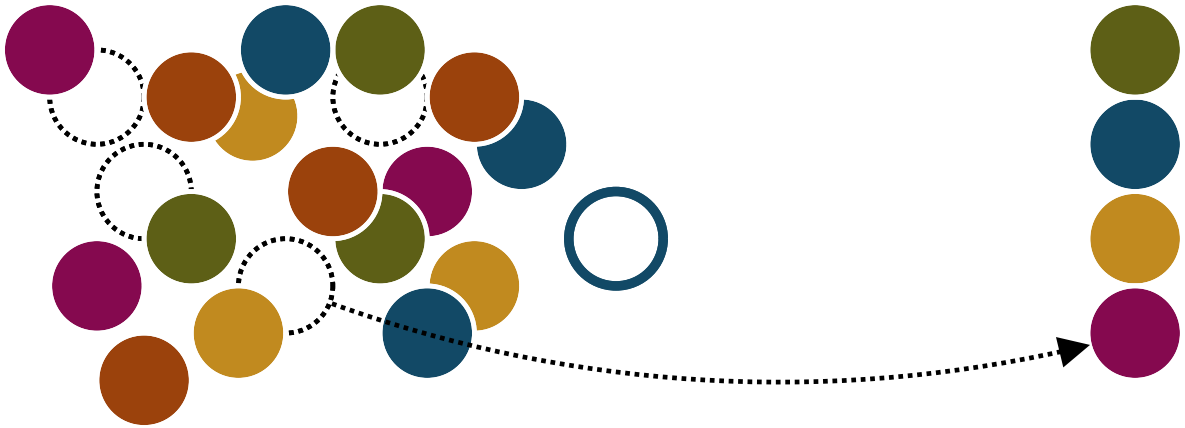
Konstruowanie klasyfikatora (II)



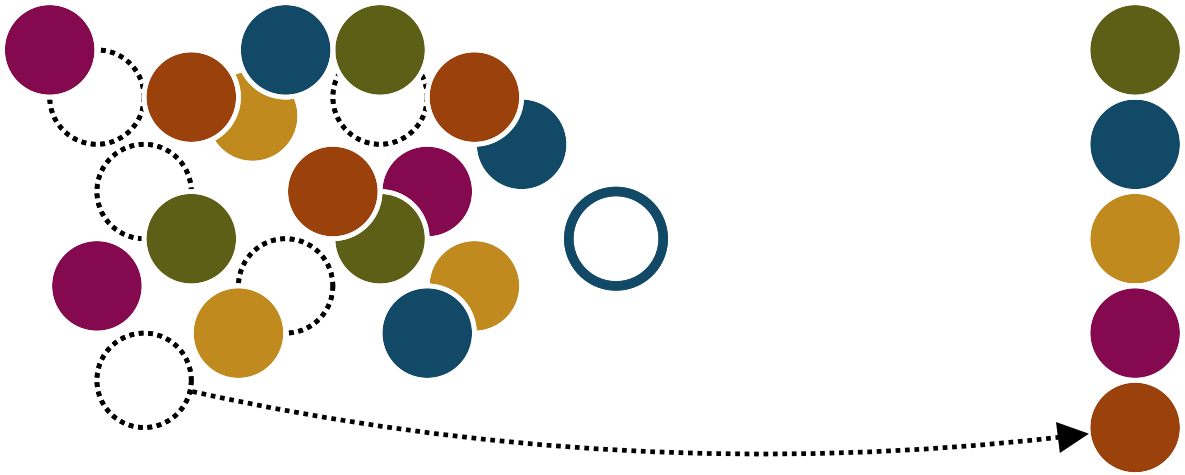
Konstruowanie klasyfikatora (III)



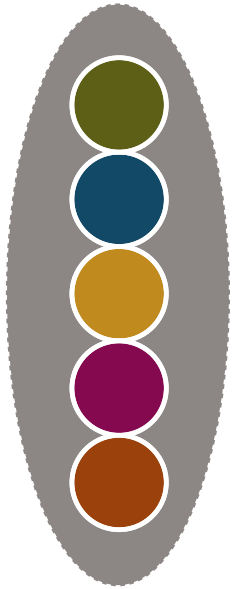
Konstruowanie klasyfikatora (IV)



Konstruowanie klasyfikatora (V)



Zbiór trenujący (klasyfikator) oraz zbiór testujący



Odgadnięte:

Nadzorowany test atrybucyjny (etap I)

Odgadnięte: ✔

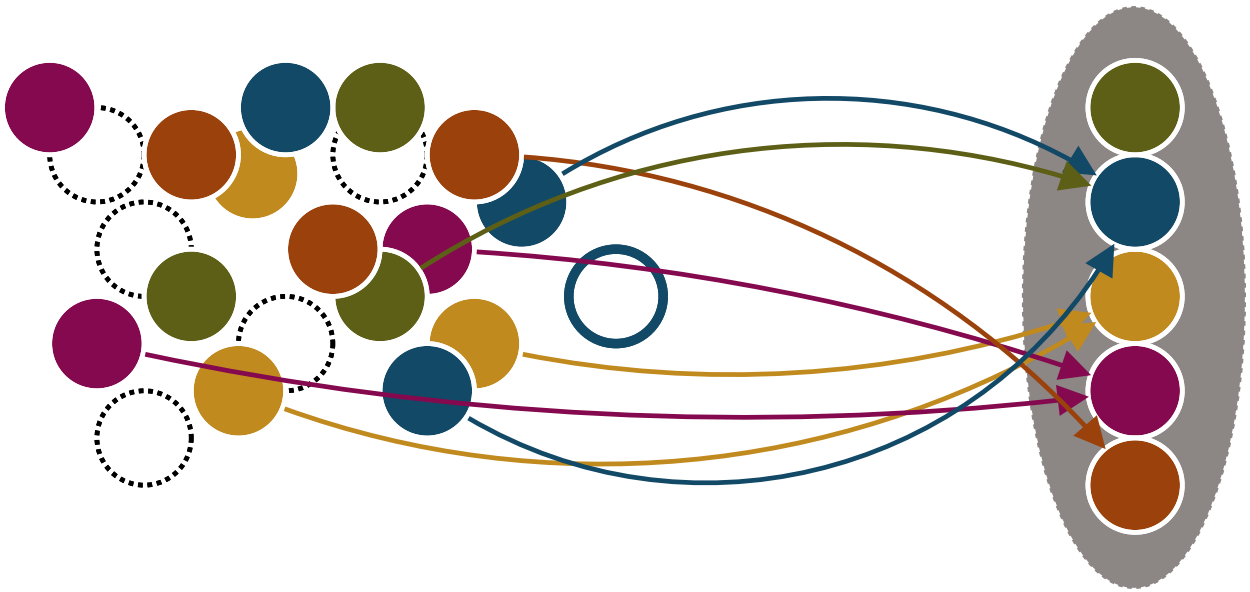
Nadzorowany test atrybucyjny (etap I)

Odgadnięte: ✔ ✔

Nadzorowany test atrybucyjny (etap I)

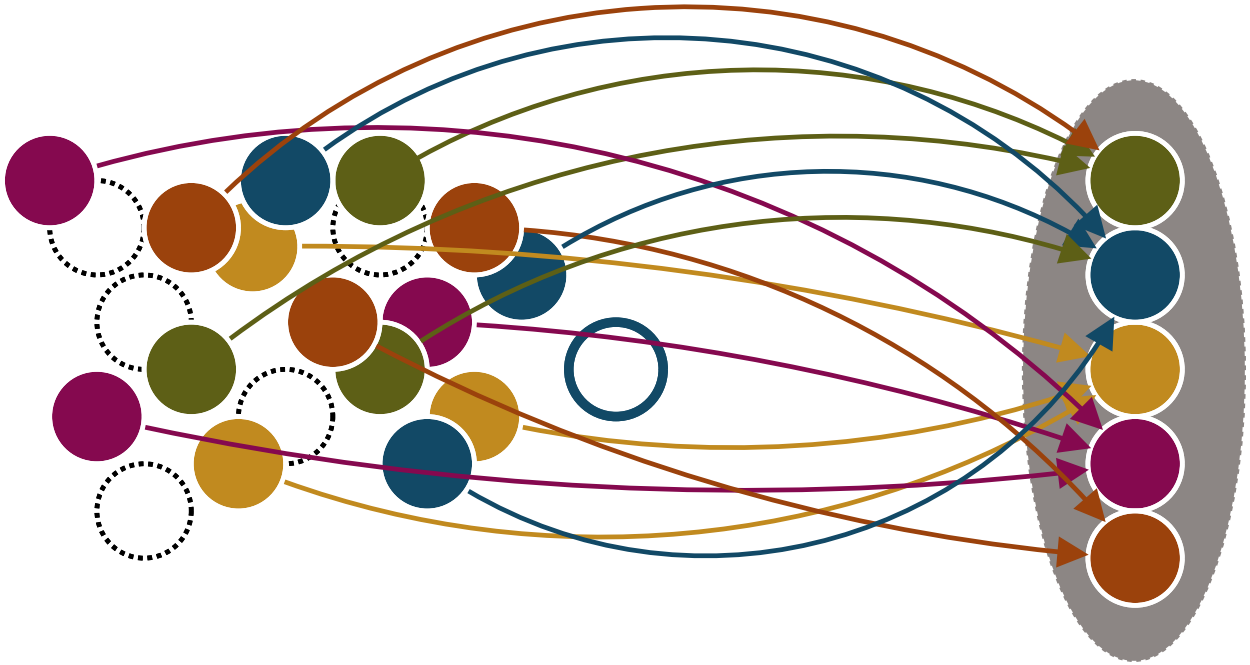
Odgadnięte: ✔ ✔ ✔

Nadzorowany test atrybucyjny (etap I)



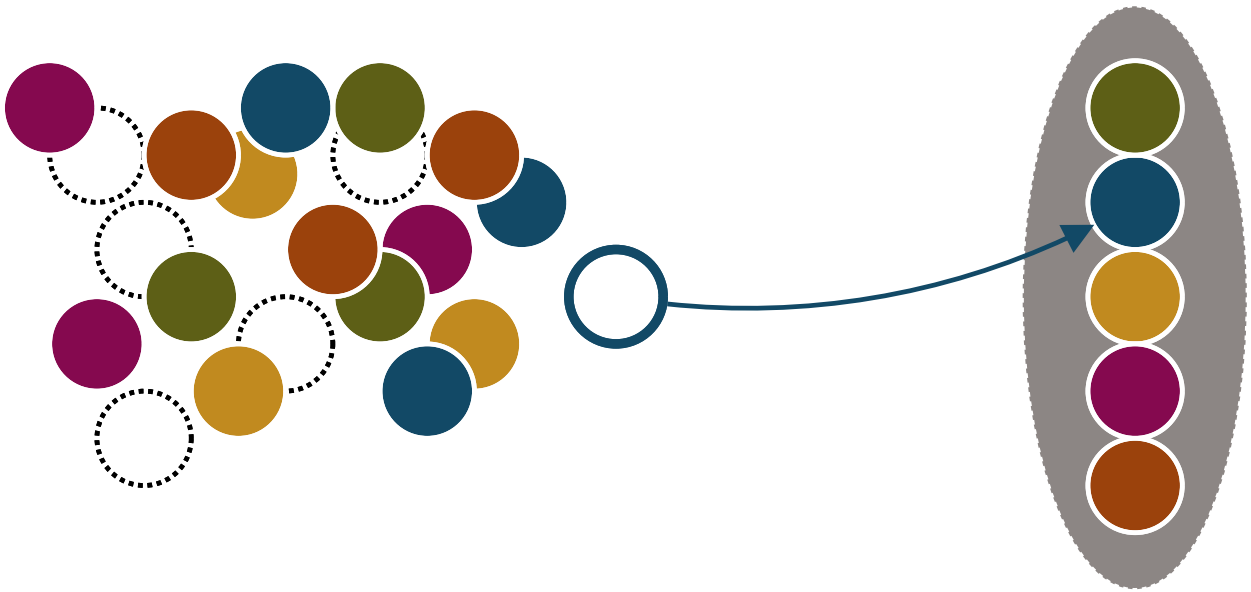
Odgadnięte: ✔ ✔ ✔ ✔ ✔ ✘ ✔ ✔

Nadzorowany test atrybucyjny (etap I)



Odgadnięte: ✔ ✔ ✔ ✔ ✔ ✘ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✘ ✔ (13 z 15)

Nadzorowany test atrybucyjny (etap II)



Akt II

Błędy systematyczne i atrybucja autorska

Błędy systematyczne i atrybucja autorska

- Celem stylometrii jest znalezienie porządku w „chaosie”, tzn. wyłuskanie regularności w pozornie chaotycznych ciągach słów
- Im lepsza metoda, tym więcej informacji można wydobyć z analizowanych tekstów
- Każdy dodaktowy szum (np. niestaranny tekst zawierający literówki) może zaciemnić końcowy obraz lub wręcz zafałszować wyniki.
- Pytanie badawcze: skoro nie ma korpusów bezbłędnych, to na jaki poziom nonszalancji można sobie pozwolić, by wyniki były wciąż wiarygodne?

Garbage in, gospel out? (1) Zepsuty korpus

Grecki korpus epicki przekonwertowano, by uniknąć możliwych błędów systematycznych związanych z kodowaniem znaków:

Unicode:

- Μῆνιν ἄειδε, θεὰ Πηληϊάδεω Ἀχιλῆος

Beta-Code:

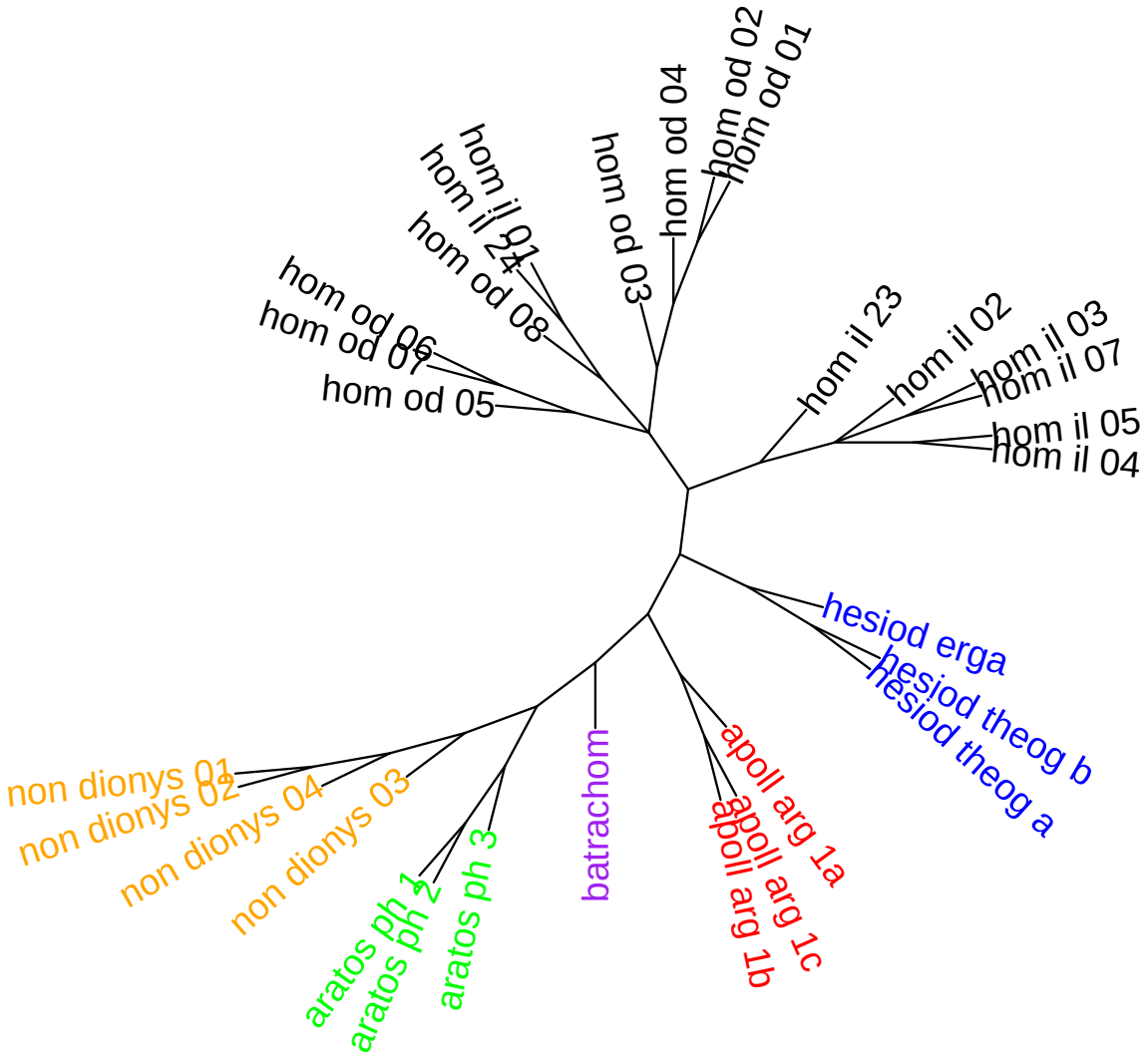
- *mh=nin a)/eide qea\ *phlhi+a/dew *)axilh=os

Rzekomy Beta-Code, a właściwie zepsute fragmenty słów:

- mh nin a eide qea phlhi a dew axilh os

(Homer, *Iliada* I 1)

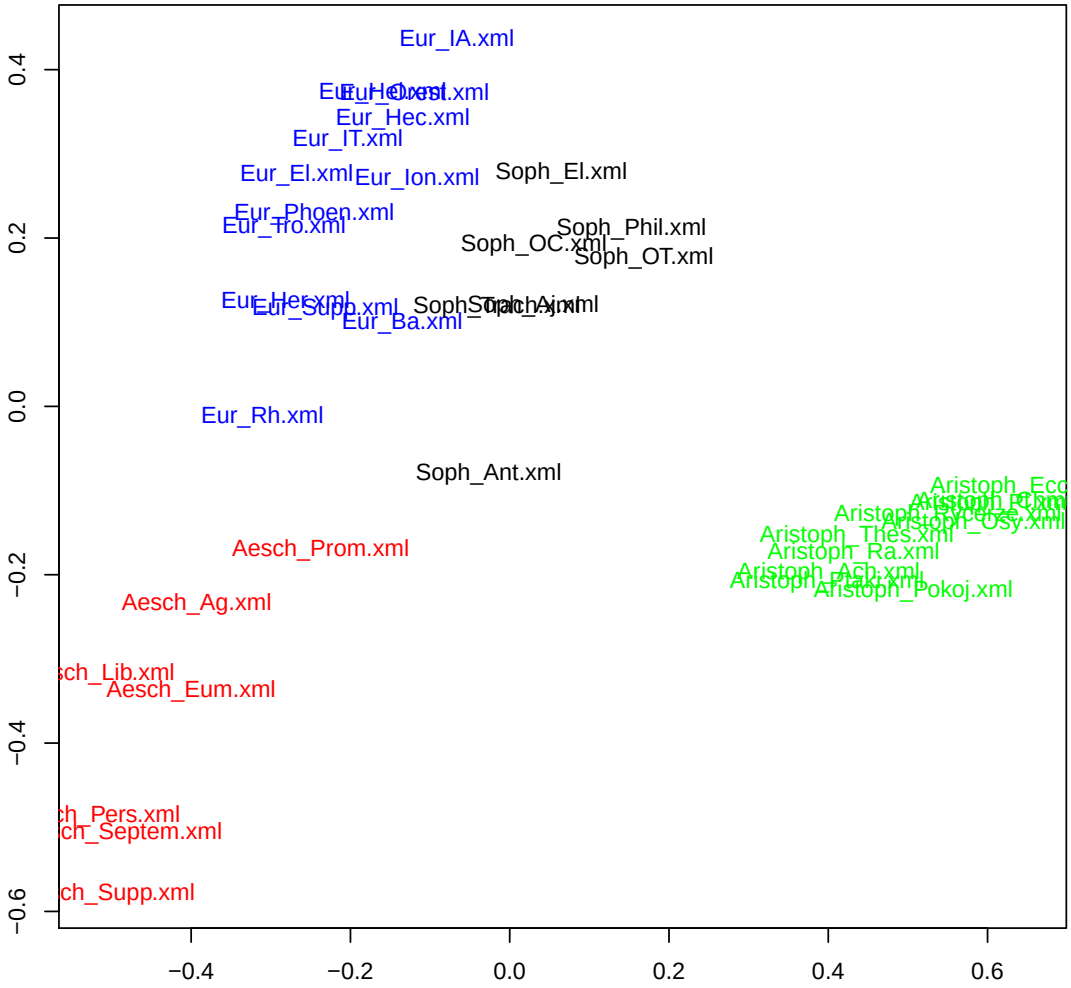
Garbage in, gospel out? (2) Eksperyment



XML + BetaCode (Euripides, *Hecuba* 629–37)

```
</sp>
</div1>
<div1 type="choral" org="uniform" sample="complete">
<div2 type="strophe" org="uniform" sample="complete">
<sp><speaker>*xoro/s</speaker>
<l>e)moi\ xrh=n sumfora/n,</l>
<l n="630">e)moi\ xrh=n phmona\n gene/sqai,</l>
<l>*)idai/an o(/te prw=ton u(/lan</l>
<l>*)ale/candros ei)lati/nan</l>
<l>e)ta/meq', a(/lion e)p' oi)=dma naustolh/swn</l>
<l n="635">*(ele/nas e)pi\ le/ktra, ta\n</l>
<l>kalli/stan o( xrusofah\s</l>
<l>*(/alios au)ga/zei.</l>
<milestone ed="p" n="638" unit="card" /></sp></div2>
<div2 type="antistrophe" org="uniform" sample="complete">
<sp><speaker>*xoro/s</speaker>
```

Skalowanie wielowymiarowe i surowy kod XML



Garbage in, gospel out? (3) Wnioski

- Całkiem sporo informacji autorskiej kryje się w zepsutym korpusie
- Metody nienadzorowane (np. kolorowe wykresy) nie są najlepszym sposobem na testowanie rzeczywistego wpływu błędów systemowych w stylometrii

Niektóre źródła szumu

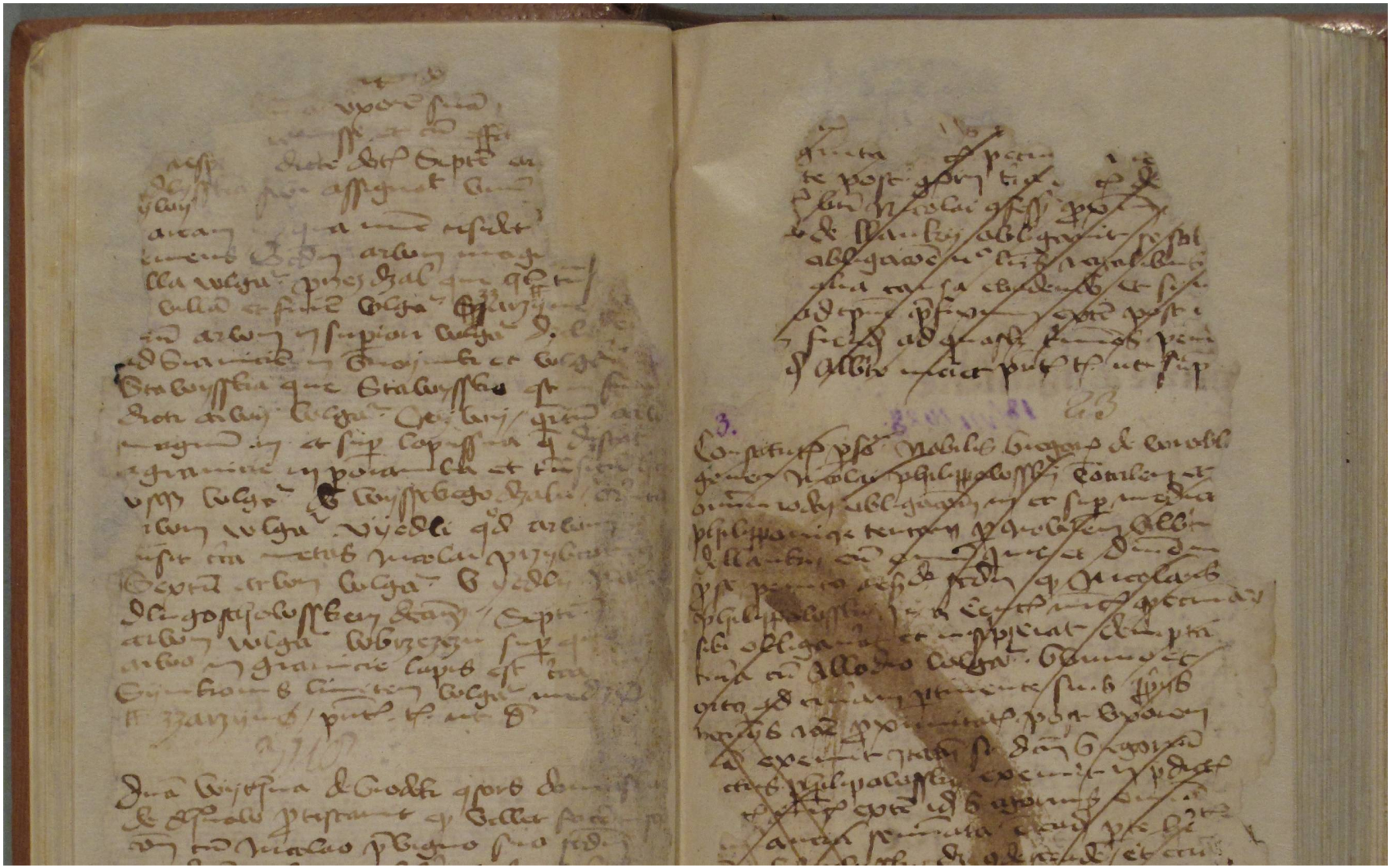
Niestarannie przygotowany korpus:

- lichy OCR
- pomyłka w stronach kodowych
- źle usunięte tagi XML
- nieusunięcie dodaktów pozatekstowych: przedmowy, przypisów, komentarzy edytorskich, noty od wydawcy, etc.

Cechy tekstu literackiego, szczególnie tekstów dawnych:

- warianty wprowadzone przez kopistów (*variae lectiones*)
- opustki w tekście (*lacunae*), interpolacje
- decyzje edytorskie o ujednoliceniu ortografii, modernizowaniu interpunkcji
- cytaty z innych tekstów literackich, ukryty plagiat, etc.

Acta Terrestria Czchoviensia, vol. III, 40–41 (r. 1428)



Dwa eksperymenty, cztery języki...

Dwa eksperymenty:

- wpływ literówek (np. marnego OCR) na wyniki
- wpływ „intertekstualności” (np. licznych cytatów) na wyniki

Cztery języki:

- 63 powieści angielskie (Dickens, Austen, James,...)
- 66 powieści niemieckich (Mann, Kafka, May,...)
- 92 powieści polskich (Orzeszkowa, Żeromski, Sienkiewicz,...)
- 94 łacińskie teksty prozaiczne (Cyceron, Seneka, Liwiusz, Pliniusz,...)

Rozprawa o metodzie

Kontrolowany eksperyment atrybucyjny:

1. korpus tekstów, których autorstwo jest znane i pewne
2. seria ślepych testów atrybucyjnych (komputer „nie zna” autorów)
3. im więcej próbek zostało prawidłowo „odgadniętych” przez komputer, tym bardziej dokładna jest testowana metoda

Eksperyment 1: Symulacja niestarannego OCR

- 100 iteracji
- W każdej kolejnej iteracji coraz większa liczba liter (bez spacji) zostanie losowo zepsuta, tj. podmieniona na inne, losowo wybrane litery
- W konsekwencji uzyskana zostanie symulacja omyłek literowych:
  - litery przekłamane przy zapisie pliku
  - literówki spowodowane niestarannym przepisywaniem na klawiaturze
  - licha jakość dokumentów przeznaczonych do skanowania
  - używanie nie trenowanego oprogramowania OCR
  - ...

Losowo zepsute litery: 1%

It is a truth univfrsally acknowledged that a single man in possession of a good fortune must be in want of a wife However little known the feelings or views of such a man may be on his fiwst eptering a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered the rightful property of some one or other of their daughters My dear Mr Bennet said his lady to him one day have you heard that Netherfield Park is let at last Mr Bennet replied that he had not But it is returned she for Mrs Long has just been here and she told me all about it Mr Bennet made no answer Do you not want to know who has taken it cried his wife impamiently You want to tell me and I have no objection to hearing it This was invitation enough Why my dear you must know Mrs Long says that Netherfield is taken by a young man of large fortune from the north of England that he came down on Monday in a chaise and four to see the place and was so much delighted with it that he agreed with Mr Morris immediately that he is to take possession before Michaelmas and some of his

Losowo zepsute litery: 2%

It is a truth universally acknowledsed that a single mal in poosession of a good fortune must be in want of a kife However little known the feelings or views of such a man may be on his first entering a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered the rightful property of some one or other of their daughters My dear Mr Bennet said his lady to him one day have you heard that Netherfield Park is let at last Mr bennet replied ihat he had not But it is returned she for Mrs Long has just been here and she told me all about it Mr Bennet made no answer Do ybu not want to know who mas taken it cried his wife impatiently You wynt to tell me and I have no objxction to haring it This was invigation enough Why my dear you must hnow Mrs Long says that Netherfield is taken by a young man of large fortune from the north of England that he came down on Mrnday in a chaise and four to see the place and was so much delighted with it that he agreed with Mr Morris immediately thal he is to take possession before Michaelmas and some of his

Losowo zepsute litery: 3%

It im a truth universally acknowledged that a single mao in possession of a good fortuue must be in tant of a wife However little known the feelings or views of such a man may be on his first entering a neighbourhood this truth is so well fixed in the minds of the wjrroudding families thav he is considered the rightful property of some one or other of their daughters My dear Mr Bennet said hfs lady to him one day have you heard that Nedherfield Park is let at last Mr Bednet replied thmt he had not But it is returned she for Mrs Long has juwt been here and she told me all about it Mr Bennet made no wnswer Do you not want to know who has taken it cried his wife impatiently You want to tell me and I have no objectlon to hearing it This was invitation enough Why my dear you yust know Mrs Long says uhat Netherfielo is taken by a young man of large fortune from the north of England that he came down on Monday in a chaise and four to see the place and was so mbch delighted with it ahat he agierd with wr Morris immediately mhkt he is to take possession bsfore Michaelmas and some nf his

Losowo zepsute litery: 4%

It is a truth universally gjknowledgep that a single man in posseusion of a good fhrtune must be in want of a wife However little known the feelings or views of such a man may be on his first entering o neighbourhood twis truth is so well fixed in the mands of the surrounding g famivies that he is considered the rightful property of some one or othhr of their daughters My dear Mr Benret said his lady to him one day have you heard that Netherfield Park is let at last Mr Bennet replied tgad he hjd njt But it is returned she for Mrs Long has just been uere and she told me all about it Mr Bennet made no answer Do you not want to know who hss taken it cried his wihe impatiently You want to tell me and I have no sbjection to hearing it This was invitation enough Why my deyr you must know Mrs Long says thbt Netherfield is taken by a young man of large footune from the north of Englcnd that he came down on Monday in a chaise and four to see the vlace and was so much delighttd iith it that he agreed jith Mr Morris immediately that he is go take possession before dichaelmas and some og his

Losowo zepsute litery: 5%

It is a rruth universaoly acknowledged that n single man in possession of a good fortune must be in want of a wifu However little known the fealings or views of such a man may be on mis first enterini a neighbourhood this truth is so well fixed in the minds of the surrounding families thjt he it considered the uightful broperty of come ony or xther of their daughters My dear Mr Bennet said his ladm to him one day have you heard that Netherfield Park is let at last Mr Bennet replied that he had not But it is returned shv fmr Mrs Long has just been here ang she told me all about it cr Bennet made no answer Dj you not waat to know who has taken it cried his wife impvtiently You want to tell me and I have no objection to heariog it This was invipation enough Why my dyar yci must know Mrs Long sayt that Netherfield is taken by a ykung man of lwrge fortune from the north of England that he came down on Monday in a chxise and foug to see thx wljce and was so much delighted wyth iv that he agreed with sr Morris gmmediately that he ik to take possfssion before Michaelmas and some of his

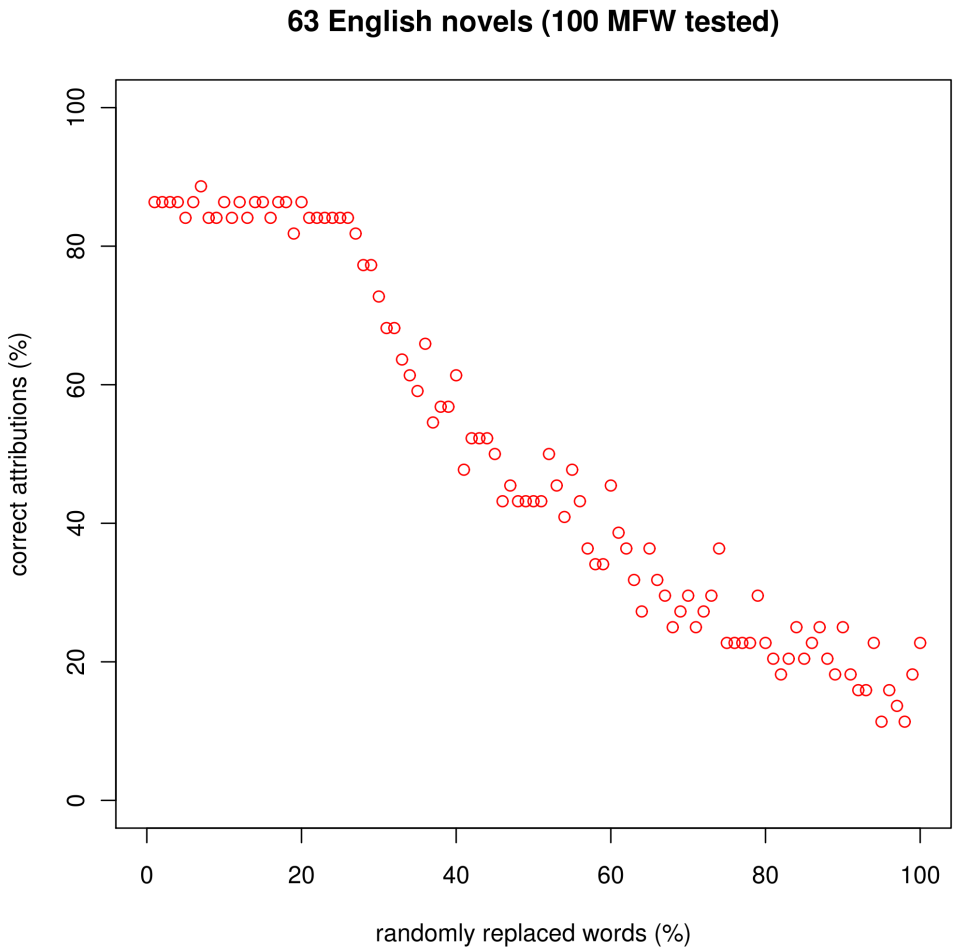
Losowo zepsute litery: 6%

It is a truth universally acknowledged that a svngle man in possession of a good aortune must be in want bf i wifl Howevew tnttle fxown the feelings or views of such a man may be on hws first entering a neighbourhood thia truth is so well fixed in the minds of the surrounding families that he xs consideref the rightfux property of some one or other of theiw daughters My dear Mr Bennet said his lady to him une day havt you hehrd thaj Netherfield Park is let at last Mr Bennet replied that he had nvx Bum it is returned she for Mrs Long has just been here and she told me all about it Mr Bennet maae no answer Do you not want to know who has taken it cried his wife impatiently You want to tell me and I have no objection to hearing it This was invitation enough Why my dean you must know Mrs Loeg xays tsat Netherfield is taken by a yxuag man of large fortune from the noyth bf England that he cumb dopg on Monday in d chaise and four to see the placn and waf so much delighted with it that he agreed witb sr Moruis immediately that he is to take possession

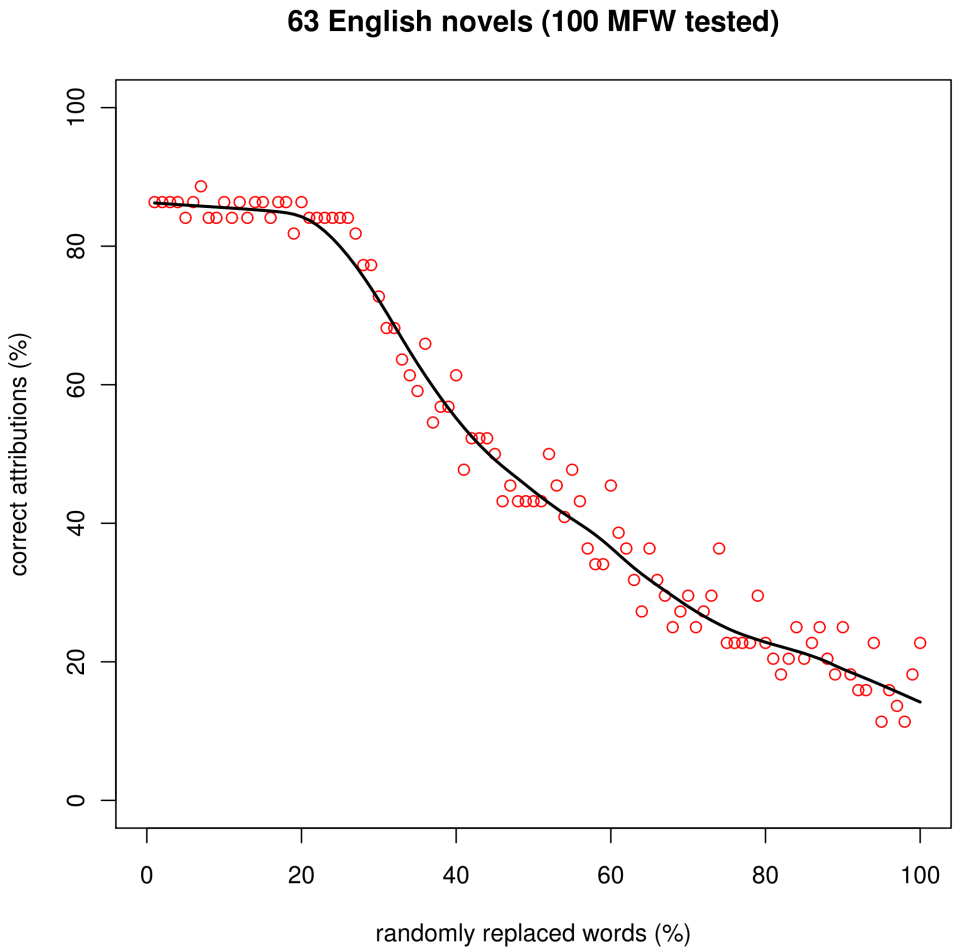
Losowo zepsute litery: 7%

It is c trutd universalty acknywledged that a single man in possession of a good hortune must be in want of a wife However little knowv the feelings or views of suth a man may be on his firjt enrering a seighbourhood this truth is ho well fixed in tge minds of the surrounming bamilies that he is considered the rigktfun property of some pne or othed of theia daugyeers My dear nr Bennet said hil lady to him onv day have you heard that Netherfielc Park is let ai last Me Bemnet repyied that he had mot But it is returned she for Mrb Long aas just xeen here and she told me all about it Mm Bennet made no answex Do you not want tf know who hes talen it cried his wife impatientcy You want to teol me and I have no objectijn to hearing it Thes was insitatapn enougw Why my dear you muvt know Mos Long smys that Netherfield ps taken by a youdg man of large iohtune from thw norbh of Engrand that hu came down on Monday in a chaise and four to see the place and was so much dwlighted hith it thht he agrerd wiuh Mr Morris imufdiately that he is fo dake possesbion

Pierwsze wyniki: powieści angielskie, 100 MFW



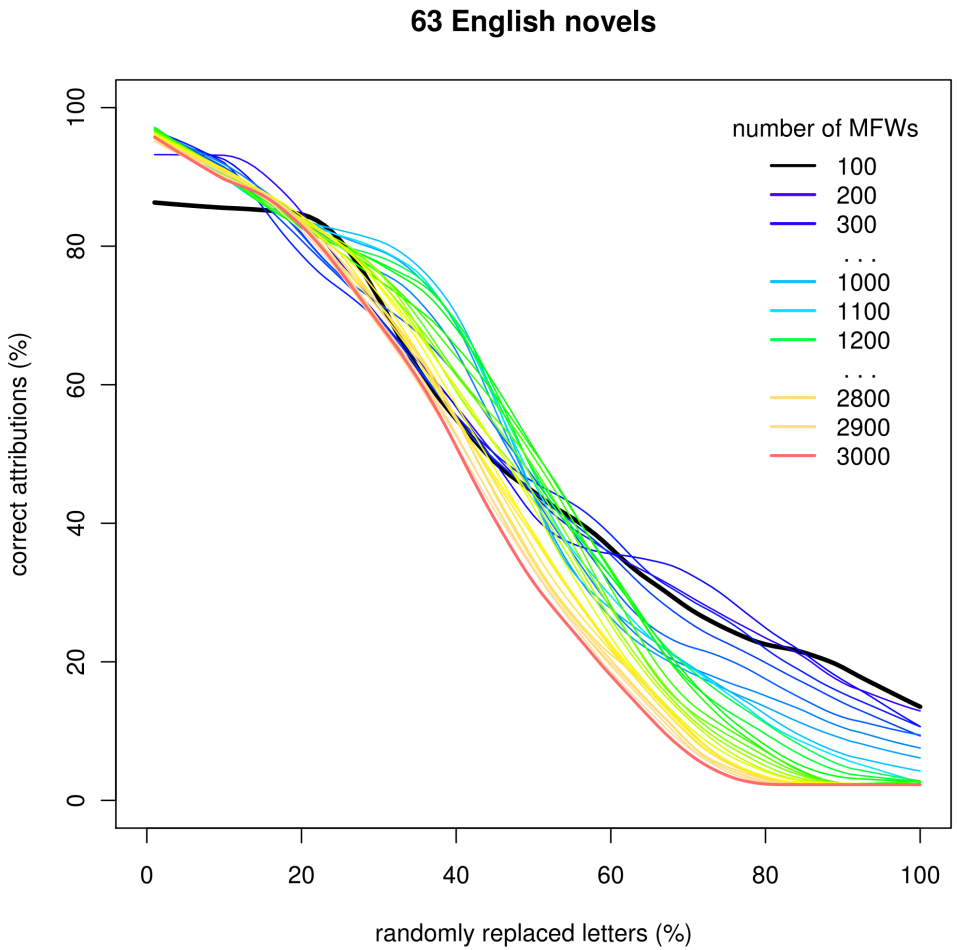
Pierwsze wyniki: powieści angielskie, 100 MFW



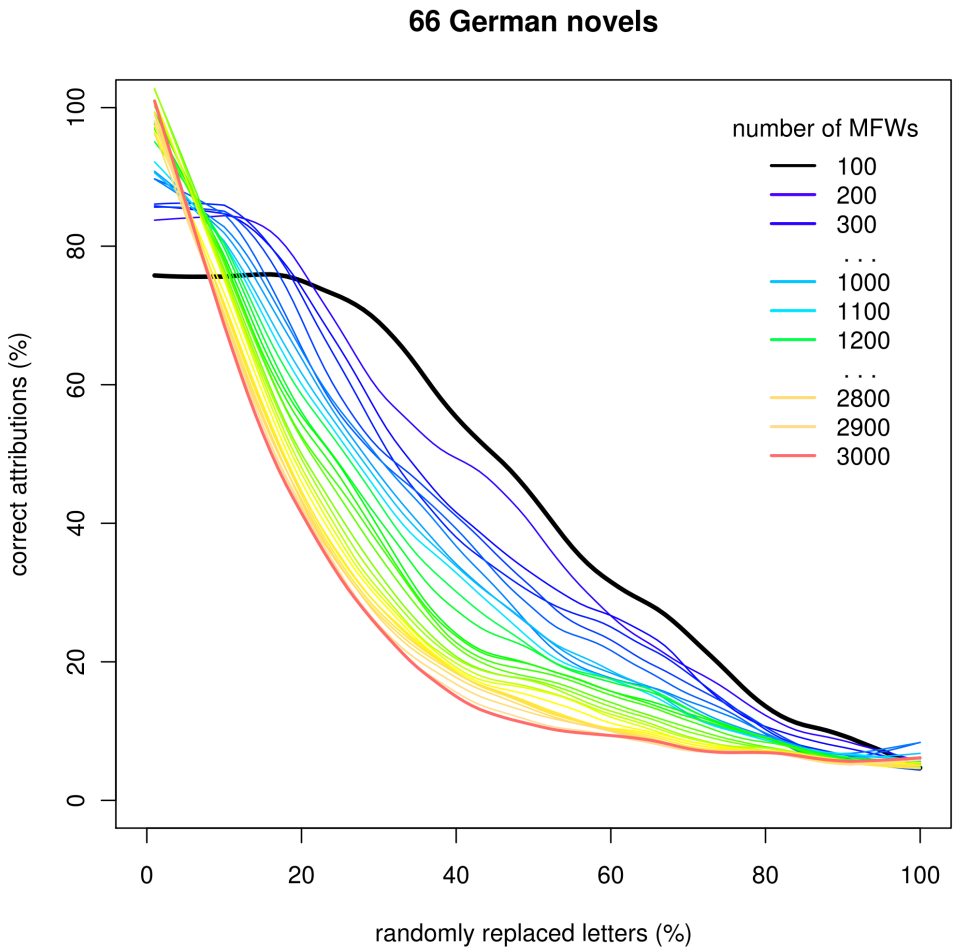
Ile znaczników stylu (np. MFW) powinno się analizować?

- mała liczba słów synsemantycznych (Mosteller & Wallace 1964)
- 50 najczęstszych słów (MFW) lub innych znaczników (Juola 2008)
- 100 najczęstszych słów (Burrows 2002)
- 300 MFW (Smith & Aldridge 2011)
- 500 MFW (Craig & Kinney 2009)
- >1000 MFW (Hoover 2004)
- cała lista frekwencyjna (Juola 2008)
- słowa, które występują u wszystkich autorów (Jockers & Witten 2010)

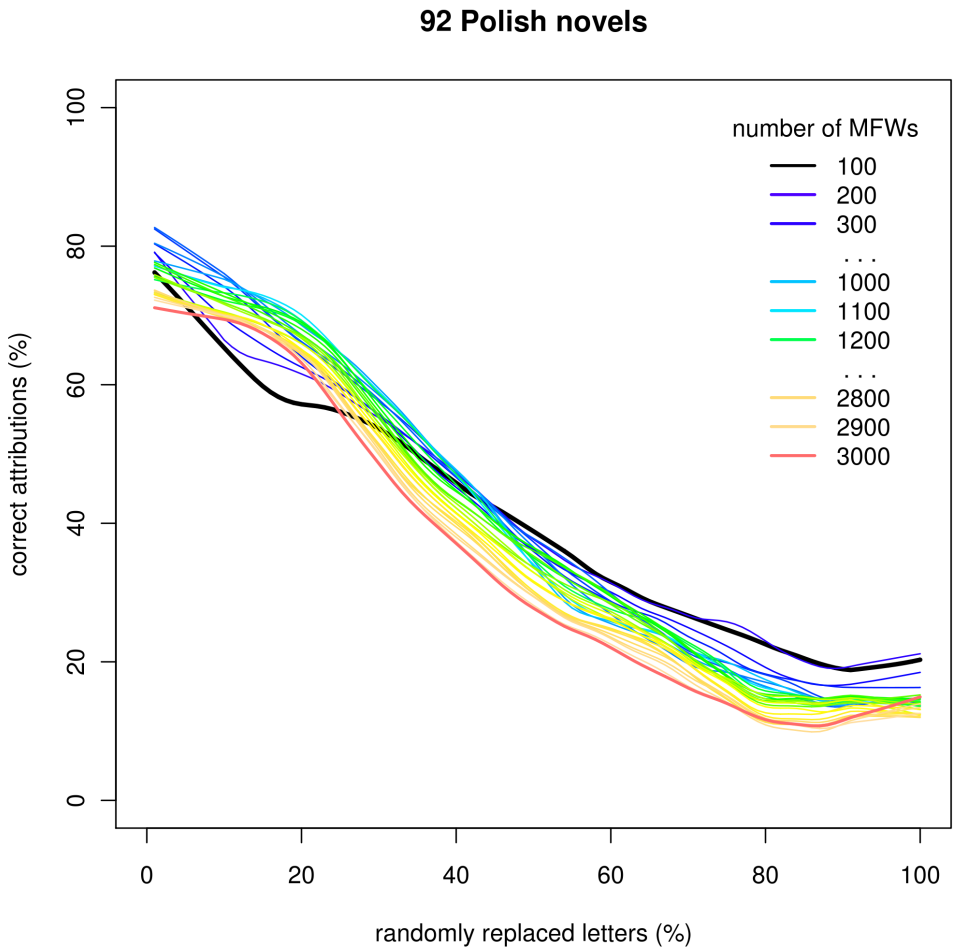
„Garbage in” i przedwczesny optymizm...



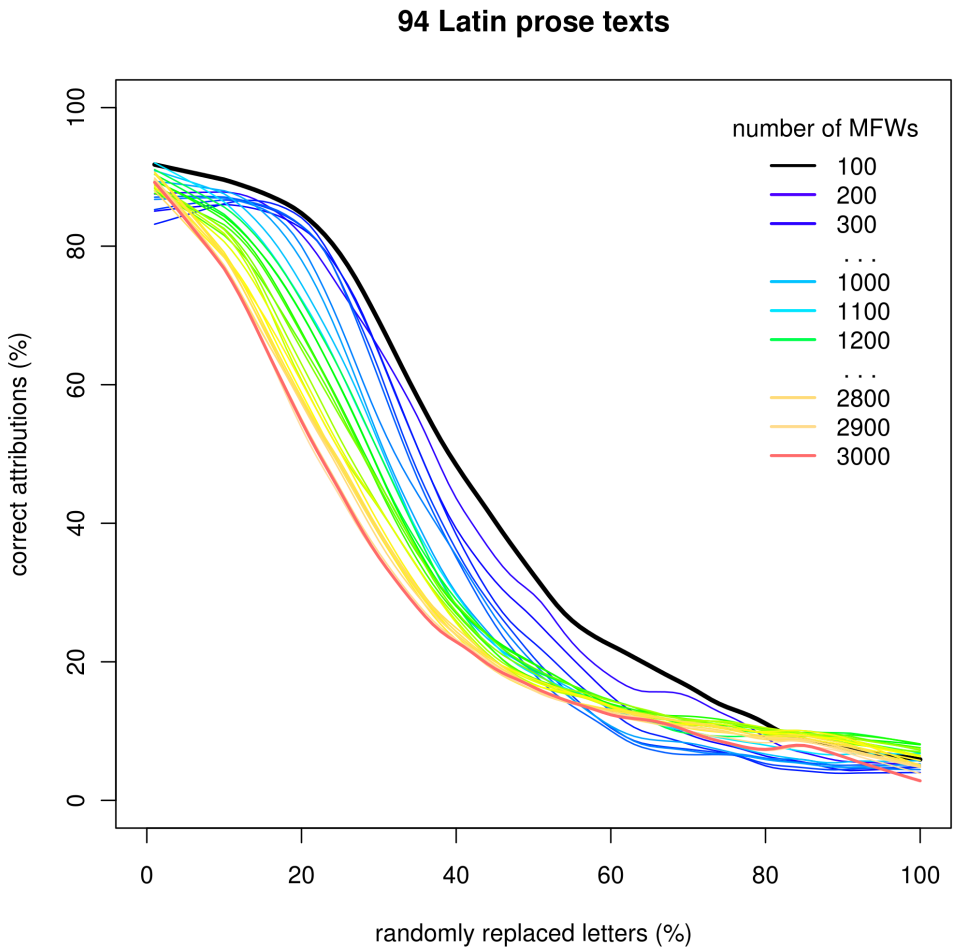
Losowo zepsute litery: powieści niemieckie



Losowo zepsute litery: powieści polskie



Losowo zepsute litery: proza łacińska



Eksperyment 1: Wnioski

- mała liczba badanych MFW jest odporna na zakłócenia (literówki)
- nawyższą skuteczność w atrybucji zapewniają jednak długie wektory MFW, które z kolei są *bardzo* wrażliwe na szum w korpusie
- zatem: lichy OCR powinien być *zawsze* unikany

Eksperyment 2: Wpływ “tradycji literackiej”

- Milczące założenie, że wszystkie teksty w korpusie są całkowicie „indywidualne”,...
- ...tak jakby jakikolwiek tekst mógł powstać bez jakiegokolwiek odwołania się do dzieł poprzedników
- możliwe rodzaje zależności pomiędzy tekstami:
  - plagiat
  - imitacja
  - jawne/ukryte cytaty
  - nieświadome inspiracje literackie
  - przynależność do konkretnej literackiej tracycji/epoki/prądu
  - ...

Libertatem apud omnes prudentes
in magno semper habitam fuisse pretio

CAPUT I

[1] Cum maxima sint commoda vitae generi humano a natura tributa, nescio tamen an longe omnium rerum humanarum pulcherrimum sit libertas. Nam etsi praeclaras animi et corporis homini natura inseverit dotes, in quibus vera felicitatis ratio consistere videtur, tamen nisi adsit libertas, neque hae munus suum rite peragere, neque homini solidam felicitatem afferre possunt. Nullus enim fere virtuti exercendae in servitute datur locus, nulla valetudinis bonae tuendae et ad usus vitae utiles conferendae conceditur ratio, cum non quod tibi, sed quod improborum libeat libidini, id facere cogaris semper.

[2] Non immerito ergo apud omnes sanae mentis homines et in omnibus rebuspublicis bene constitutis summa semper et diligens libertatis habita fuit ratio. Animadverterunt nimirum viri prudentes omne illud, quod in hac civili societate bonum est, aut sola servitute labefactari aut in universum tolli. Hinc pulchre illud et sapienter a Cicerone dictum esse constat, servitutem videlicet postremum omnium malorum esse et non modo vita, sed morte etiam repellendum. Nec vero tam dictis, quam factis, quale sit malum servitus, ab optimis quibusque semper fuit declaratum.

[3] Athenienses certe quamdiu duro servitutis iugo pressi fuerunt a tyrannis, nulla insigni laude bellica prae aliis vicinis populis excellebant. At quamprimum libertatis compotes facti sunt, quasi alii prorsus redditi, facile palmam virtutis bellicae aliis

Wolność u wszystkich mądrych ludzi
zawsze w wielkim poważeniu była miana

ROZDZIAŁ I

[1] Gdyż barzo wielkie pożytki rodzaju ludzkiemu są z natury dane, jednak nie wiem, jeśli ze wszytkich rzeczy ludzkich napiękniejsza jest wolność. Bo zacne tak na umyśle, jako i na ciele natura człowiekowi wpoiła dary, w których prawdziwy szczęścia sposób ludzie upatrują, wszakoż gdzieby nie było wolności, ani te dary zaczęte szczyrze urzędu swego wyprawować, ani człowiekowi gruntownego szczęścia przynieść mogą. Bo w niewoli żadne miejsce nie zostaje ku wprawowaniu się w cnotę, żaden sposób ku zachowaniu zdrowia dobrego i ku pożytecznym potrzebom żywota ludzkiego nie bywa pozwolony, gdyż nie to, co się tobie, ale co się złych chciwości podoba, to zawsze czynić musisz.

[2] Nie darmo tedy u wszytkich zdrowego umysłu ludzi i we wszytkich rzeczachpospolitych dobrze postanowionych wielkie zawsze i pilne ku zachowaniu wolności było staranie. Uważali to bowiem ludzie mądrzy, że to wszytko, cokolwiek się dobre w tym spółku a stowarzyszeniu ludzkim nazywa, albo samą niewolą to się targa, albo do końca zniszczono bywa. Stąd pięknie ono od Cycerona i mądrze jest powiedziano, że niewola prawdziwie nawiętsze jest złe miedzy wszytkiemi złemi rzeczami, a nie tylko żywotem, ale <i> śmiercią ma być odpędzone. A nie tylko słowy, ale i uczynkami, jakowa jest rzecz zła niewola, od nazacniejszych i cnotliwych ludzi znacznie się pokazywało.

[3] Atenieńczykowie zaprawdę jako długo srogim jarzmem niewolej ciśnieni byli od tyranów, żadną osobliwą chwałą wojenną nad insze sąsiady nie przodkowali; ale skoro wolności dostali, jakoby się prawie inszemi stawszy, snadnie górę dzielności wojennej przed wszytkimi innymi otrzymali. To zaiste, jako świadczy

Wolan cytuje Cycerona, Herodota, ...

Libertatem apud omnes prudentes
in magno semper habitam fuisse pretio

CAPUT I

[1] Cum maxima sint commoda vitae generi humano a natura tributa, nescio tamen an longe omnium rerum humanarum pulcherrimum sit libertas. Nam etsi praeclaras animi et corporis homini natura inseverit dotes, in quibus vera felicitatis ratio consistere videtur, tamen nisi adsit libertas, neque hae munus suum rite peragere, neque homini solidam felicitatem afferre possunt. Nullus enim fere virtuti exercendae in servitute datur locus, nulla valetudinis bonae tuendae et ad usus vitae utiles conferendae conceditur ratio, cum non quod tibi, sed quod improborum libeat libidini, id facere cogaris semper.

[2] Non immerito ergo apud omnes sanae mentis homines et in omnibus rebuspublicis bene constitutis summa semper et diligens libertatis habita fuit ratio. Animadverterunt nimirum viri prudentes omne illud, quod in hac civili societate bonum est, aut sola servitute labefactari aut in universum tolli. Hinc pulchre illud et sapienter a Cicerone dictum esse constat, servitutem videlicet postremum omnium malorum esse et non modo vita, sed morte etiam repellendum. Nec vero tam dictis, quam factis, quale sit malum servitus, ab optimis quibusque semper fuit declaratum.

[3] Athenienses certe quamdiu duro servitutis iugo pressi fuerunt a tyrannis, nulla insigni laude bellica prae aliis vicinis populis excellebant. At quamprimum libertatis compotes facti sunt, quasi alii prorsus redditi, facile palmam virtutis bellicae aliis

Wolność u wszystkich mądrych ludzi
zawsze w wielkim poważeniu była miana

ROZDZIAŁ I

[1] Gdyż barzo wielkie pożytki rodzaju ludzkiemu są z natury dane, jednak nie wiem, jeśli ze wszytkich rzeczy ludzkich napiękniejsza jest wolność. Bo zacne tak na umyśle, jako i na ciele natura człowiekowi wpoiła dary, w których prawdziwy szczęścia sposób ludzie upatrują, wszakoż gdzieby nie było wolności, ani te dary zaczęte szczyrze urzędu swego wyprawować, ani człowiekowi gruntownego szczęścia przynieść mogą. Bo w niewoli żadne miejsce nie zostaje ku wprawowaniu się w cnotę, żaden sposób ku zachowaniu zdrowia dobrego i ku pożytecznym potrzebom żywota ludzkiego nie bywa pozwolony, gdyż nie to, co się tobie, ale co się złych chciwości podoba, to zawsze czynić musisz.

[2] Nie darmo tedy u wszytkich zdrowego umysłu ludzi i we wszytkich rzeczachpospolitych dobrze postanowionych wielkie zawsze i pilne ku zachowaniu wolności było staranie. Uważali to bowiem ludzie mądrzy, że to wszytko, cokolwiek się dobre w tym spółku a stowarzyszeniu ludzkim nazywa, albo samą niewolą to się targa, albo do końca zniszczono bywa. Stąd pięknie ono od Cycerona i mądrze jest powiedziano, że niewola prawdziwie nawiętsze jest złe miedzy wszytkiemi złemi rzeczami, a nie tylko żywotem, ale <i> śmiercią ma być odpędzone. A nie tylko słowy, ale i uczynkami, jakowa jest rzecz zła niewola, od nazacniejszych i cnotliwych ludzi znacznie się pokazywało.

[3] Atenieńczykowie zaprawdę jako długo srogim jarzmem niewolej ciśnieni byli od tyranów, żadną osobliwą chwałą wojenną nad insze sąsiady nie przodkowali; ale skoro wolności dostali, jakoby się prawie inszemi stawszy, snadnie górę dzielności wojennej przed wszytkimi innymi otrzymali. To zaiste, jako świadczy

omnibus praeripuerunt. Hoc nimirum, Herodoto teste, demonstrare volebant, ut quia non sibi, sed dominis rem acquirerent, ideo de industria peccarent, at vero iuris sui facti, tum demum summam operam reipublicae, utpote rebus iam suis consulentes, impenderent. Nec vero philosophiae tam insigne apud eos efflo-resceret unquam studium, non ea legum honestissimarum nata esset disciplina, nisi libertas, bonarum rerum omnium magistra, sedem illic suam fixisset. Barbari enim, qui tyrannorum imperio parere didicerunt, adeo ab omni civilium morum et omnis honestatis notitia alieni semper fuerunt, ut non magnum sui discrimen a reliquis animantibus brutis ostenderent.

[4] Sed si qua urbs alia praeclaros libertatis protulit fructus, Roma certe fuit, in qua maiorane belli decora an civilium artium ornamenta fuerint, haud facile discernere possis. Quamdiu enim fidelem et strenuam libertatis communis custodem sese gessit, et plurimos imperatores bellica laude illustres et sapientissimos quosque senatores, velut ex insigni quodam virtutis domicilio, proferebat semper. Tantus autem Romanis libertatis retinendae inerat amor, ut non solum ipsi pro libertate sua fortissime dimicaverint, sed alienam oppugnari etiam minime passi sint.

[5] Testatur id bellum illud duce Quinctio contra Philippum regem Macedonum gestum, cuius quidem magna impensa, labore ac periculo feliciter peracti non aliam mercedem sibi proposuerant Romani quam solam civitatum Graeciae libertatem – cum summa quidem Graecorum laetitia et virtutis Romanae admiratione. Videbatur id Graecis divinum esse quiddam: existere aliquam in terris gentem, quae nec periculo, nec labore ullo deterrita tantum alienae libertati faveret, ut potentissimum etiam hostem ultro lacessere et imperium in gentem liberam illi eripere non dubitarit. Tanta autem ex fructu libertatis recuperatae tunc Graecis laetitia fuit oborta, ut et gestu, et vocibus declaraverint nihil prorsus in universis mortalium rebus esse, quod cum bono

Herodotus, pokazać chcieli, iż ponieważ nie sobie, ale panom swym majętności nabywali, przeto umyślnie z kresu występowali, ale gdy już sami sobą władali, tedy dopiero wielkie staranie o rzeczpospolitą czynili, jakoby już nikomu innemu, ale rzeczom swym dogadzali. I nigdy by nauka filozofijej tak znacznie u nich nie była zakwitnęła, i nigdy by poczciwych praw u nich szkoła się nie była otworzyła, gdzieby wolność, dobrych wszytkich rzeczy mistrzynia, stolice swojej tam była nie zasadziła. Abowiem grubi ludzie, którzy pod rozkazowanie tyranów podlegali, od wszech zawsze obyczajów wypolerowanych i od znajomości wszelakiej uczciwości dalecy byli, że niewielką w sobie różność od inszych źwierząt niemych pokazowali.

[4] Ale jeśliże które miasto insze znaczne owoce z siebie wolności wydawało, Rzym ten zaprawdę był, w którym jeśliżeby ozdoby były więtsze wojenne alboli też nauki pokojowi służące, niesnadnie rozeznać możesz. Bo jako długo wiernym i ochotnym stróżem pospolitej wolności być się okazował, i wiele hetmanów wojenną sławą ozdobnych, i barzo też mądrych senatorów, jakoby z ozdobnego niejakiego cnót mieszkania zawsze z siebie wydawał. A tak wielka u Rzymianów ku zachowaniu wolności była miłość, iż nie tylko sami dla swojej wolności barzo mężnie wojowali, ale też i innych wolności namniejszego uszczerbku cierpieć nie dopuścili.

[5] Poświadcza tego ona wojna przez hetmana Kwinkcyjusza przeciwko Filipowi królowi macedońskiemu wiedziona, której z wielką pracą i niebezpieczeństwem szczęśliwie dokończywszy, nie insze zapłaty Rzymianie przed oczyma mieli, tylko samę wolność greckich miast, z wielką zaprawdę radością Greków i z podziwieniem dzielności rzymskiej. Zdało się to być nieco boskiego Grekom, że taki jest naród na świecie, który ani żadnym niebezpieczeństwem, ani pracą ustraszony, tak się wielkim przyjacielem cudzej wolności okazał, że też nieprzyjaciela barzo potężnego dobrowolnie na się wyzwać i panowanie nad ludem wolnym onemu wyrwać nie wątpili. A tak wielka z pożytku wolności przywróconej natenczas Grekom radość była się stała, że i postawą, i głosy swemi to pokazowali, że zgoła we wszyt-

libertatis comparari possit. Ac propterea multoties id evenisse constat ut mortem potius oppetere certam quam in servitute ducere vitam plurimi homines maluerint.

[6] Nec ea fortitudo aut robur animi in solis animadversa sunt viris, sed mulieres etiam saepius hanc virtutem sunt aemulatae. Celticae quidem feminae, Antonino Caracalla cum viris earum bellum gerente et victoria potito, cum a Romanis captae essent, interrogatae, utrumne vendi an occidi mallent, mori se malle responderunt; at cum nihilominus posthac venditae essent, iugum servitutis amissa libertate imponi sibi iniquo ferentes animo, mortem sibi ipsis consciverunt occisis prius filiis, ne et ii servire cogerentur. Simile quiddam et de Cantabricis mulieribus memoratur, quae itidem prius filiis, ne in hostium manus devenirent, deinde et sibi manus intulerunt.

[7] Verissime itaque dictum a quodam esse constat: dulce nimirum bonum esse libertatem, quae morte emitur, miseram rem servitutem, cui mors anteponitur. Diogenes quoque Cynicus, ut testatur Laertius, quantum contineat boni libertas demonstrare volens, cum a Cratero quodam, Alexandri Magni praefecto, invitatus esset, homine praedivite et opulento, ut ad se commigraret, illi ita respondit: Malo Athenis, inquit, salem lingere quam apud Craterum delicatis cibi frui, sentiens inopiam in libertate omnibus etiam deliciis divitum, quae aliquam redoleant servitutem, esse praeferendam. Ac proinde cum Callisthenem philosophum hoc nomine beatum quidam dicerent, quod apud Alexandrum lautissime ac splendide viveret, Immo ego, inquit, miserum et infelicem puto, quia illi prandendum et cenandum sit, quando Alexandro visum fuerit.

[8] Hanc vitam et mus ille rusticus, ut est in apologo, potius quam muris urbani probavit, quem etsi opiparis semper eduliis vescentem conspiceret, tamen quia perpetua quadam sollicitudine et continuo pavore haec illi voluptas constaret, non potuit

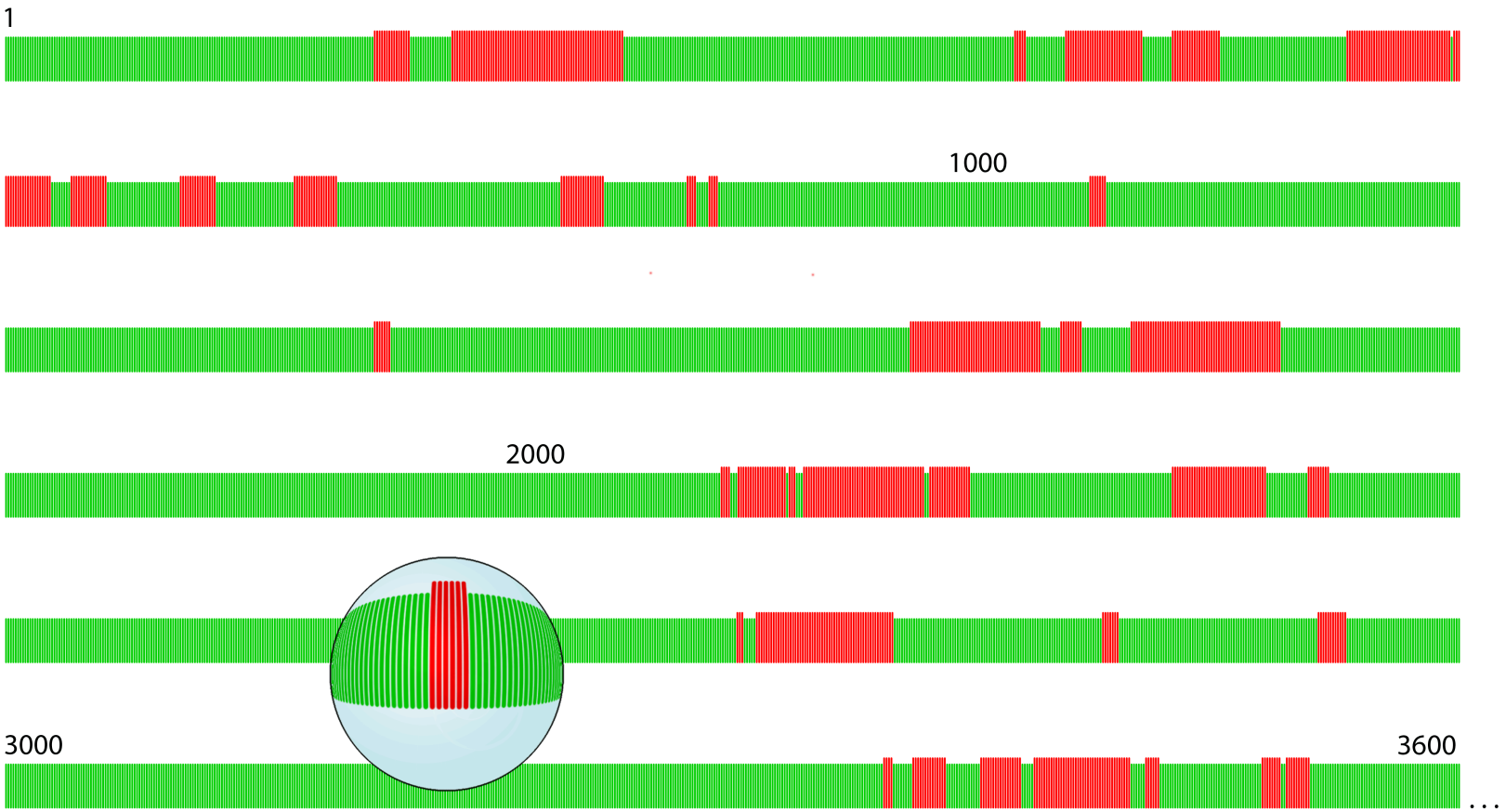
kich rzeczach ludzkich nic takowego niemasz, co by z dobrem wolności porównać się mogło. I dlatego częstokroć to się przytrafiało, że wiele ludzi wolało śmierć pewną podjąć, niżli żywot swój w niewoli prowadzić.

[6] A nie tylko takowa mężność i stałość umysłu w samych mężoch była upatrowana, ale też tej dzielności częstokroć i niewiasty naśladowały. Niewiasty francuskie, gdy Antoninus Karakalla z mężami ich wojnę wiódł i zwycięstwo nad nimi otrzymał, gdy od Rzymian poimane były, będąc pytane, co by wolały, jeśliby albo przedanemi, albo zabitemi być wolały, odpowiedziały, że umrzeć wolały. A jednak przedsię potym były przedane; gniewając się, że jarzmo niewoli zatraceniem wolności na nich było włożone, śmierć sobie same zadały, pobiwszy naprzód syny, aby też w niewoli być nie musieli. Temu też podobny przykład o niewiastach hiszpańskich przypominają, które także pierwej syny, potym same siebie zabiły, aby w ręce nieprzyjacielskie nie przyszły.

[7] A tak prawdziwie jest od niektórego powiedziano, że barzo rzecz jest smaczna wolność, która śmiercią kupiona bywa, a mizerna jest niewola, nad którą śmierć przekładają. Także też Dyjogenes Cynikus, jako świadczy Laercyjusz, chcąc pokazać, co w sobie dobrego zamyka wolność, gdy od Kratera niektórego, Aleksandra Wielkiego przełożonego, był proszony, człowieka barzo bogatego i dostatecznego, aby się do niego przeprowadził, tak mu odpowiedział: „Wolę – powiada – w Atenach sól lizać niżli u Kraterów rozkosznych pokarmów używać", rozumiejąc, że ubóstwo w wolności nad wszytkie rozkoszy bogaczów, które by niejaką niewolą pachnęły, ma być przełożone. I dlatego, gdy Kallistena filozofa przeto szczęśliwym niektórzy być nazywali, iż u Aleksandra barzo hojnie i rozkosznie żył, „i owszem, ja – rzekł – biednego i nieszczęśliwego być rozumiem, gdyż on obiedwać i wieczerzać wtenczas musi, kiedy się Aleksandrowi podobać będzie".

[8] Taki żywot i mysz ona sielska, jako o tym *apologus* świadczy, raczej niż myszy miejskiej pochwaliła, którą chocia hojnemi zawsze pokarmy karmiącą się widziała, wszakoż iż z ustawiczną niejaką trwogą i częstą bojaźnią tej rozkoszy zażywała, nie

3600 pierwszych słów Wolanowego traktatu



Eksperyment 2: Symulacja „intertekstualności”

- 100 iteracji
- W każdej kolejnej iteracji coraz większa liczba słów zostanie losowo podmieniona na taką samą liczbę słów wybranych losowo z całego korpusu
- W konsekwencji uzyskana zostanie symulacja zależności intertekstualnej pomiędzy tekstami w korpusie

Losowo zamienione słowa: 5%

It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife However little known the feelings or views of such a man may be on his first entering a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered the rightful property of some one or other of their daughters My dear Mr Bennet said his lady to him one day have you heard that Netherfield Park is let at last Mr Bennet replied that he had not But it is returned she for Mrs Long has just been here and she told me all about it Mr Bennet made no answer Do you not want to know who has taken it cried his wife impatiently You want to tell me and I have no objection to hearing it This was invitation enough Why my dear you must know Mrs Long says that Netherfield is taken by a young man of large fortune from the north of England that he came down on Monday in a chaise and four to see the place and was so much delighted with it that he agreed with Mr Morris immediately that he is to take possession before Michaelmas and some of his

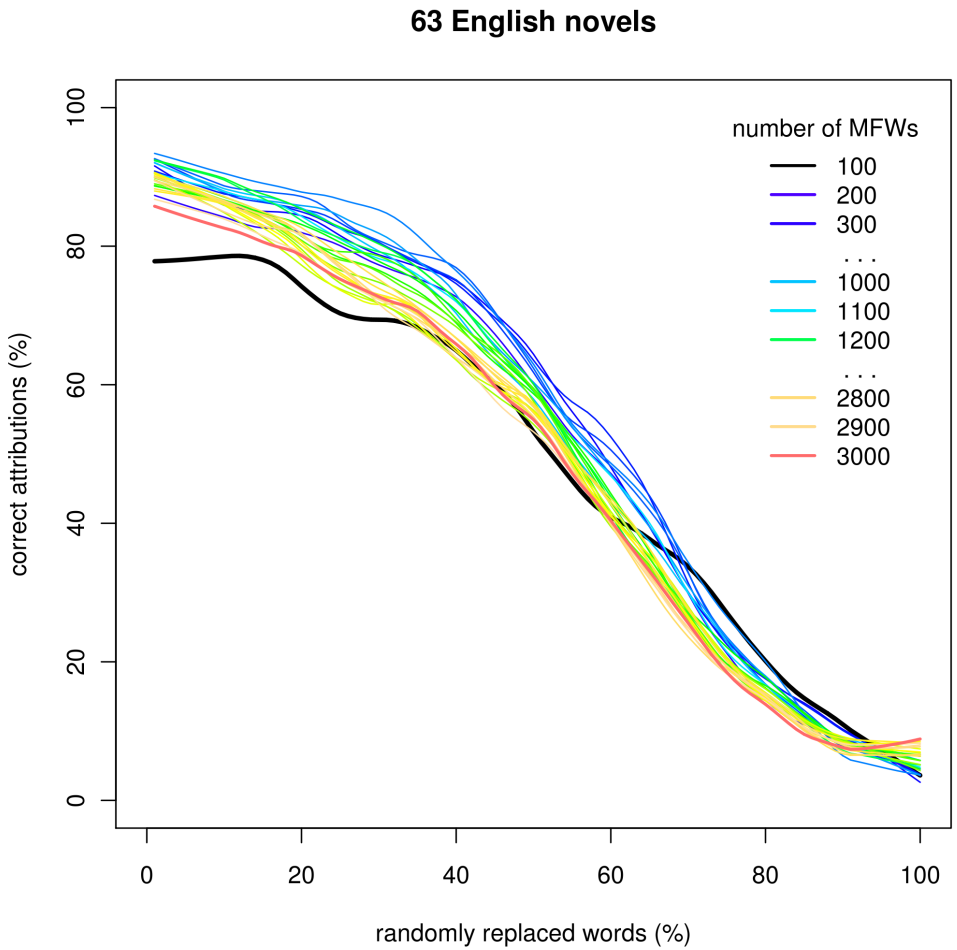
Losowo zamienione słowa: 5%

It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife However little known the feelings or of such a man may be on his first entering a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered the rightful of some one or other of their daughters My dear Mr Bennet said his lady to him one day have you heard that Netherfield Park is let at last Mr Bennet replied that he had not But it is returned for Mrs Long has just been here and she told me all about it Mr Bennet made no answer Do you want to know who taken it cried his wife impatiently You want to tell me and I have no objection to hearing it This was enough Why my dear you must know Mrs Long says that Netherfield is taken by a young man of large from the north of England he came down on Monday in a chaise and four to see the place and was so much delighted with it that he agreed with Mr Morris immediately that he is to possession Michaelmas and some of his

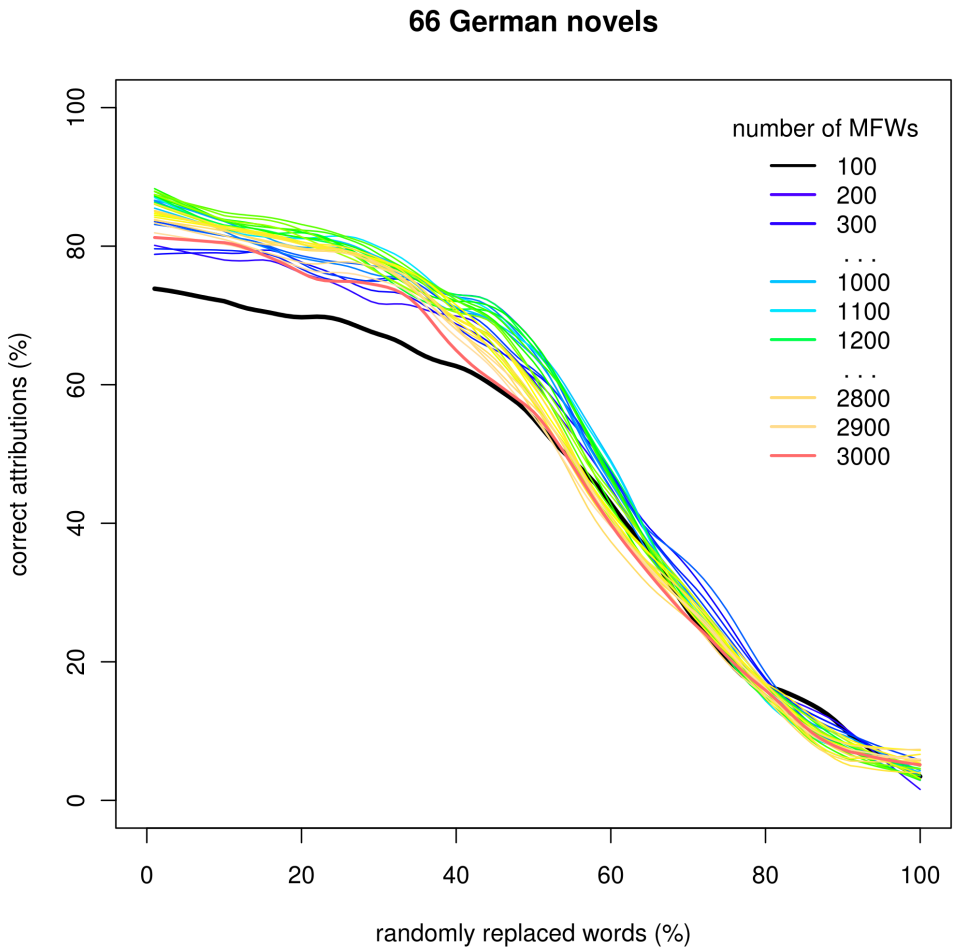
Losowo zamienione słowa: 5%

It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife However little known the feelings or you of such a man may be on his first entering a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered the rightful could of some one or other of their daughters My dear Mr Bennet said his lady to him one day have you heard that Netherfield Park is let at last Mr Bennet replied that he had not But it is returned with for Mrs Long has just been here and she told me all about it Mr Bennet made no answer Do you custom want to know who do taken it cried his wife impatiently You want to tell me and I have no objection to hearing it This was another enough Why my dear you must know Mrs Long says that Netherfield is taken by a young man of large mean from the north of England thinking he came down on Monday in a chaise and four to see the place and was so much delighted with it that he agreed with Mr Morris immediately that he is to so possession box Michaelmas and some of his

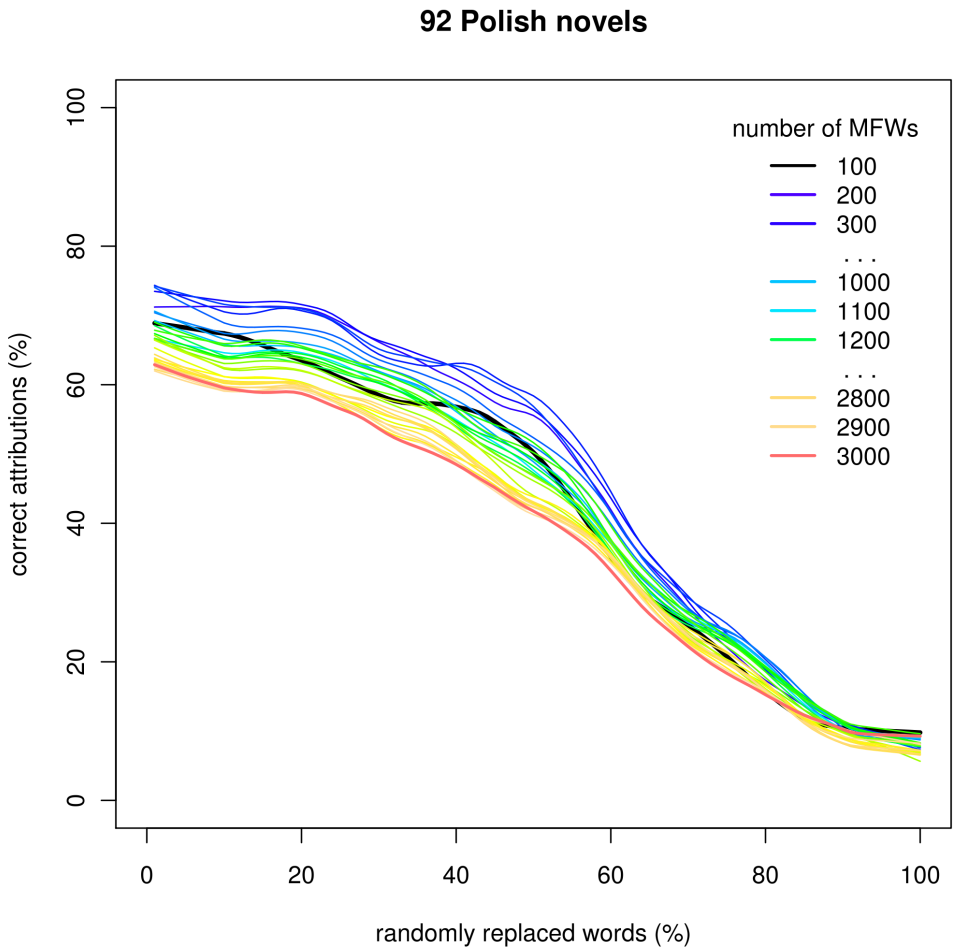
Powieści angielskie: powolny spadek rozpoznawalności



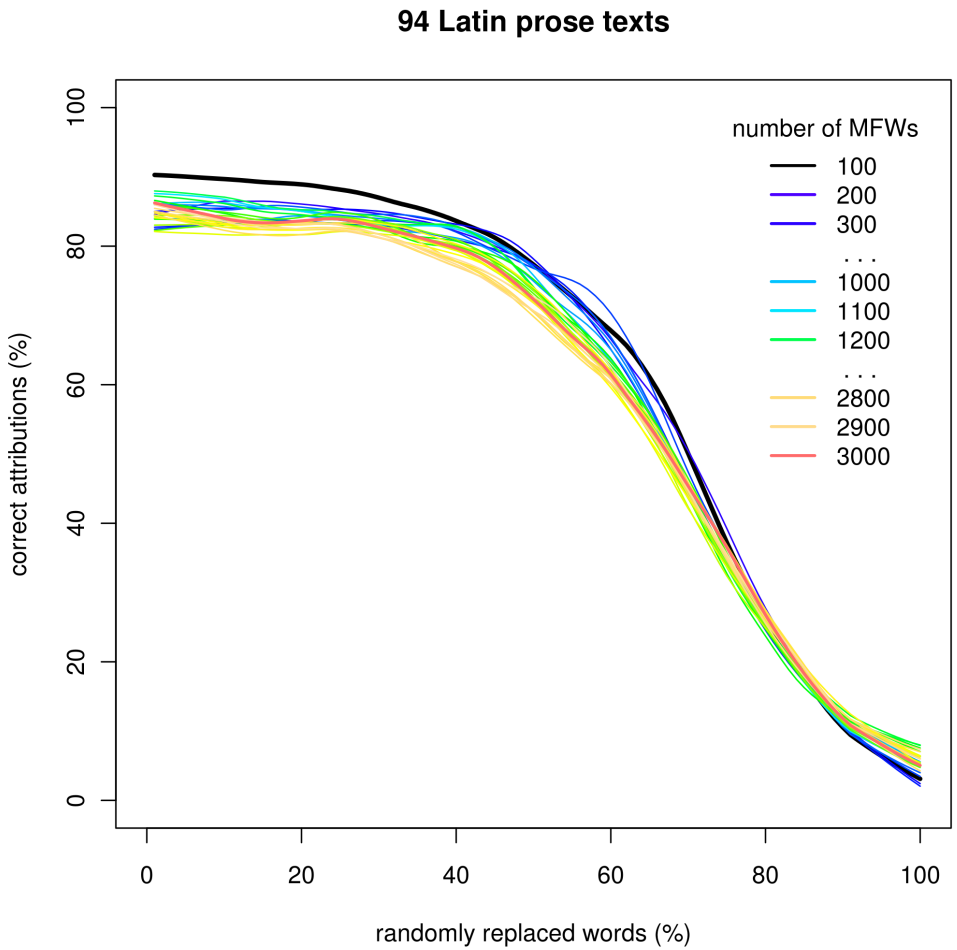
Powieści niemieckie: wyniki podobne do powieści ang.



Powieści polskie: spokojny spadek efektywności



Proza łacińska: niewiarygodna odporność na cudze słowa



Eksperyment 2: Wnioski

- wpływ “intertekstualności” jest wyraźnie widoczny,...
- ... ale tylko przy dużej domieszce cudzych słów
- zdumiewające zachowanie korpusu łacińskiego wymaga dalszych badań

Akt IV

Spojrzenie z oddali,
czyli stylometria wielkoskalowa