



UiO : University of Oslo

Exploring ways of distinguishing prize-winning novels from non-prize-winning ones



Jarle Ebeling
University of Oslo
ICAME 44, 2023, South Africa

Outline

- Aim
- Previous research
- Material
- Method
- Findings
- Further research

Aim

Investigate if/how prize-winning novels can be distinguished from non-prize-winning ones on linguistic grounds

A **genre** is a category of literature, such as mystery, suspense, science fiction or horror. Each genre has its own conventions. Romance, for example, focuses on romantic love between two people and often ends positively. Generally, genre fiction tends to place value on entertainment and, as a result, it tends to be more popular with mass audiences.

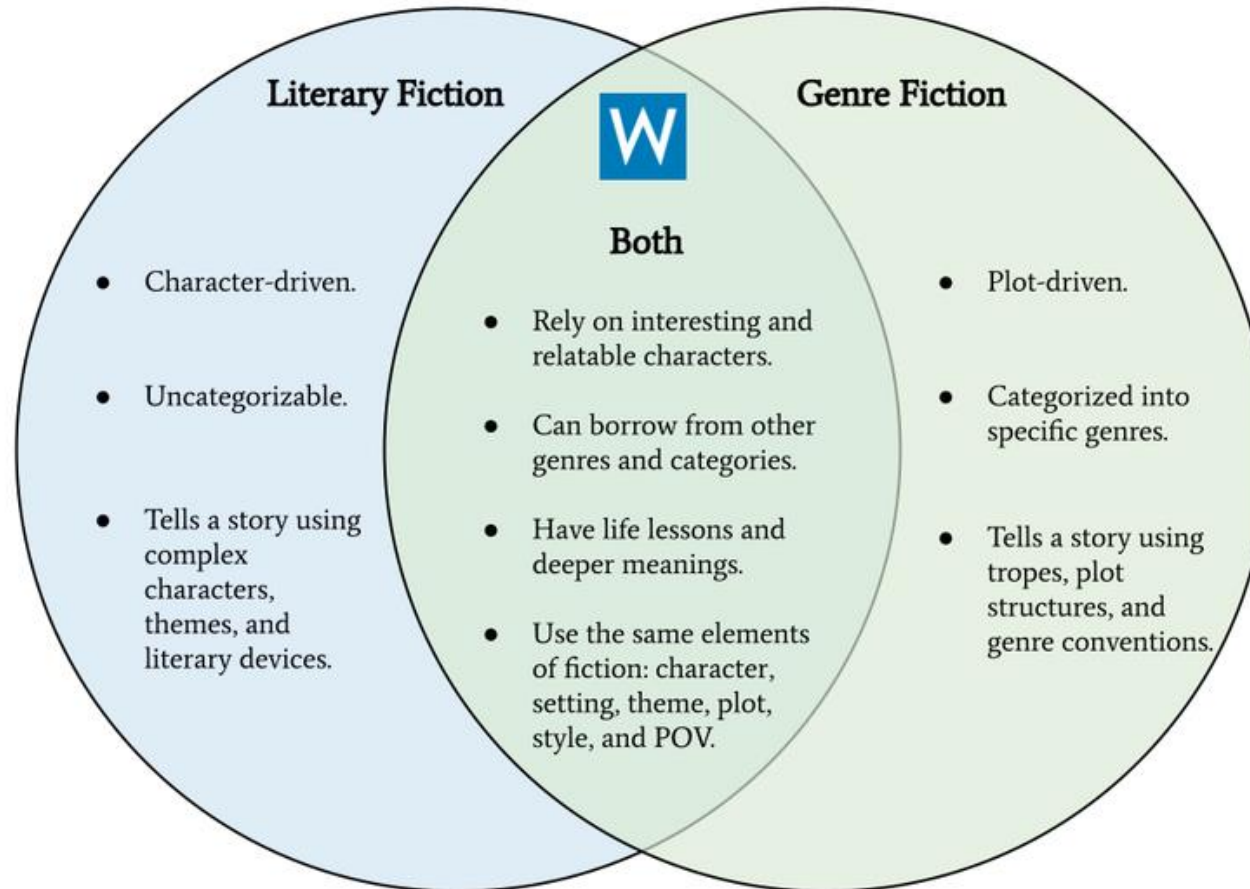
Literary fiction, on the other hand, is a bit trickier to define. In general, it emphasizes meaning over entertainment. Literary fiction also aspires toward art. Of course, that abstract of “art” is where things get most tricky. What is art? In fiction it can be defined as interesting and deep manifestations of the elements of craft: dimensional characters, a pleasing arc of tension, evocative language and thematic purpose.

<https://www.writingclasses.com/toolbox/ask-writer/what-is-the-difference-between-genre-fiction-and-literary-fiction-how-do-i-know>

General Fiction, sometimes called contemporary fiction, focuses on the everyday experiences and conflicts of a protagonist, usually an adult, with detailed characterization and background. General Fiction is intended for older readers and has more mature themes.

<https://support.wattpad.com/hc/en-us/articles/360020325871-Content-Categories>

Literary Fiction vs. Genre Fiction Venn Diagram



<https://writers.com/literary-fiction-vs-genre-fiction>

Ashok et al (2013)

Mores/Less successful = download counts from Project Gutenberg

Categories:

- Adventure
- Detective/Mystery
- Fiction
- Historical fiction
- Love Stories
- Poetry
- Science fiction
- Short stories

Min. 2 x 50 books from each category (i.e. successful ones and not successful ones)

More successful books use more:
prepositions, determiners, nouns, coordinating conjunctions, numbers, pronouns, adjectives

Less successful books use more:
adverbs, verbs, foreign words, symbols, interjections

Observation: More successful books closer to informative writing

Piper & Portelance (2016)

1,200 modern (post 2000) novels

Categories:

Prize-winners (the National Book Award (US), PEN/Faulkner Award (US), Governor General's Award (Canada), Giller Proze (Canada), Man Booker Prize (UK))

Bestsellers (10-200 weeks on the New York bestseller list)

NYTimes (reviewed in New York Times)

SciFi (ranking on Amazon)

Mystery (ranking on Amazon)

Romance (ranking on Amazon)

LIWC (Linguistic Inquiry and Word Count software)

Prizewinners, significantly more use of:

Nature: *village, birds, dust, sea, horses, winter, forest, river, grass, earth, trees, mountain*

Childhood: *childhood, boy, father, child, children, dream, imagine, mother, become, remember*

Time/Period: *winter, youth, afterwards, whenever, sometime, old, born, spring, young, older, once, summer*

Nature

Childhood

Time



Nostalgia

Material

Booker Prize Winners (abbr.: PW-<author>) (literary fiction)

Non-prize Winners (abbr.: NPW-<author>) («general» fiction/bestsellers)

25 novels in each group, all by British authors published
between 1969 and 2018 (2.2 (NPW) – 3 (PW) mill. tokens per group)

Tagged with CLAWS

TextCode	Author	Title	Year of Publication	Gender
AlaHol3	Alan Hollinghurst	The Line of Beauty	2004	male
AniBro2	Anita Brookner	Hotel du Lac	1984	female
AnnBur1	Anna Burns	Milkman	2018	female
AntBya3	A. S. Byatt	Possession: A Romance	1990	female
BarUns5	Barry Unsworth	Sacred Hunger	1992	male
BerRub1	Bernice Rubens	The Elected Member	1969	female
DavSto2	David Storey	Saville	1976	male
GraSwi1	Graham Swift	Last Orders	1996	male
HilMan2	Hilary Mantel	Bring Up the Bodies	2012	female
HowJac1	Howard Jacobson	The Finkler Question	2010	male
IanMce1	Ian McEwan	Amsterdam	1998	male
IriMur4	Iris Murdoch	The Sea, the Sea	1978	female
JamFar1	J. G. Farrell	The Siege of Krishnapur	1973	male
JamKel1	James Kelman	How late it was, how late	1994	male
JohBer1	John Berger	G.	1974	male
JulBar1	Julian Barnes	The Sense of an Ending	2011	male
KazIsh1	Kazuo Ishiguro	The remains of the day	1989	male
KinAmi3	Kingsley Amis	The Old Devils	1986	male
PauSco1	Paul Scott	Staying On - A novel	1977	male
PenLiv3	Penelope Lively	Moon Tiger	1987	female
PerNew1	P. H. Newby	Something to Answer For	1968	male
SalRus1	Salman Rushdie	Midnight's Children	1981	male
SopMac1	Sophie Mackintosh	The Water Cure (long-listed)	2018	female
StaMid1	Stanley Middleton	Holiday	1974	male
WilGol4	William Golding	Rites of Passage	1980	male

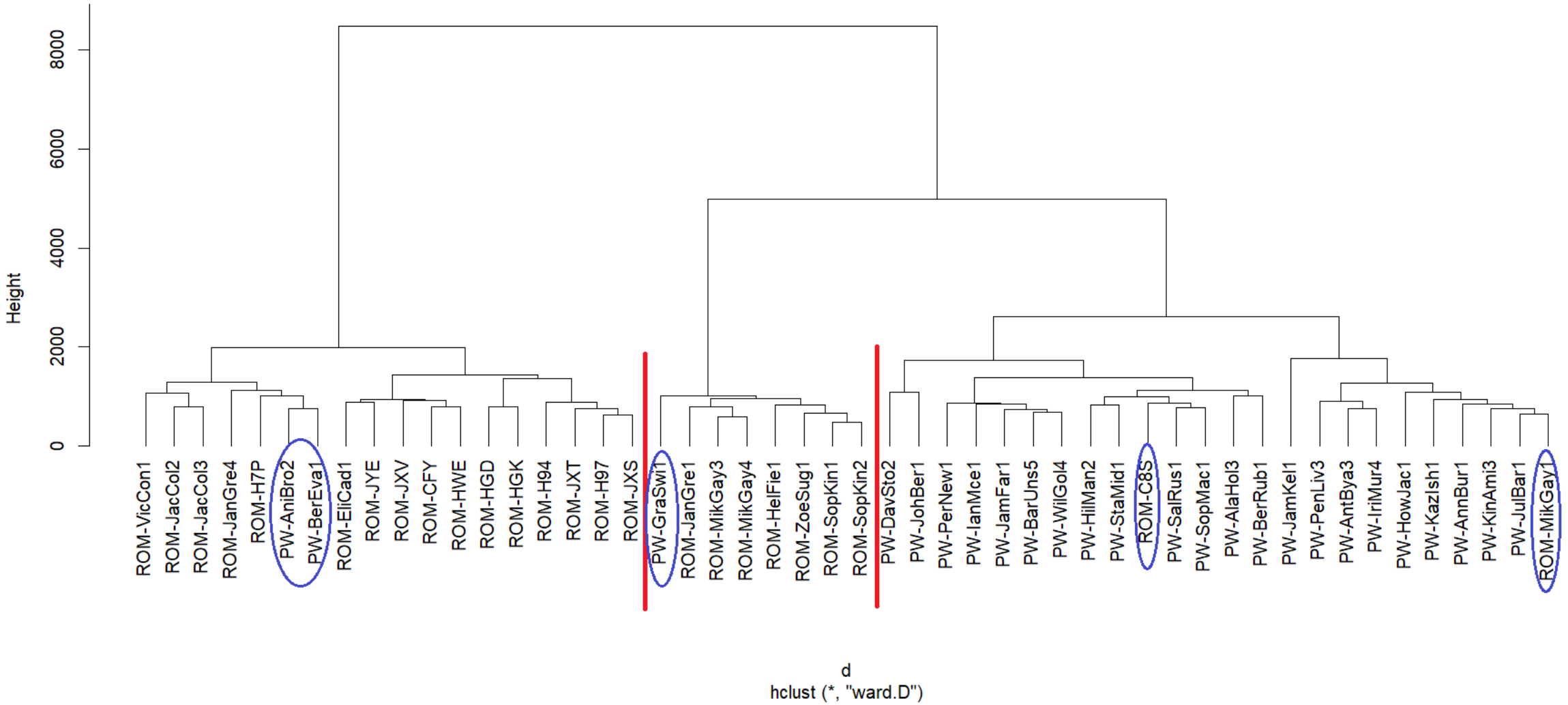
TextCode	Author	Title	Year of Publication	Gender
AlaSil2	Alan Sillitoe	Life Goes On	1985	male
AliSmi4	Ali Smith	Autumn	2016	female
AndGre1	Andrew Greig	When They Lay Bare	1999	male
AngCar3	Angela Carter	Love	1971	female
AntPow3	Anthony Powell	Books Do Furnish a Room	1971	male
BarCom3	Barbara Comyns	The Juniper Tree	1985	female
BarPym1	Barbara Pym	Quartet in Autumn	1977	female
BenElt4	Ben Elton	Time and Time Again	2014	male
BerBai1	Beryl Bainbridge	The Bottle Factory Outing	1974	female
DavLod1	David Lodge	Changing Places	1975	male
EdwDoc1	Edward Docx	The Calligrapher	2003	male
EliBuc2	Elizabeth Buchan	Separate Beds	2010	female
JamBal1	J. G. Ballard	Crash	1973	male
JefArc2	Jeffrey Archer	First Among Equals	1984	male
JohKin2	John King	England Away	1999	male
JonCoe2	Jonathan Coe	The Accidental Woman	1987	male
MarWes2	Mary Wesley	Not That Sort of Girl	1987	female
MeeSya1	Meera Syal	Anita and Me	1996	female
MilGib1	Miles Gibson	The Sandman	1984	male
MonDic2	Monica Dickens	The Room Upstairs	1964	female
NicHar1	Nick Harkaway	Angelmaker	2012	male
NicHor1	Nick Hornby	Funny Girl	2014	male
SimRav2	Simon Raven	The Judas Boy	1968	male
SteFry1	Stephen Fry	The Liar	1991	male
SusHil3	Susan Hill	Mrs de Winter	1993	female

Method

- Unigrams, bigrams and trigrams
 - tokens, lemmas, parts-of-speech
- Standardised type-token ratio
- Average token and sentence length
- Most-common-words approaches

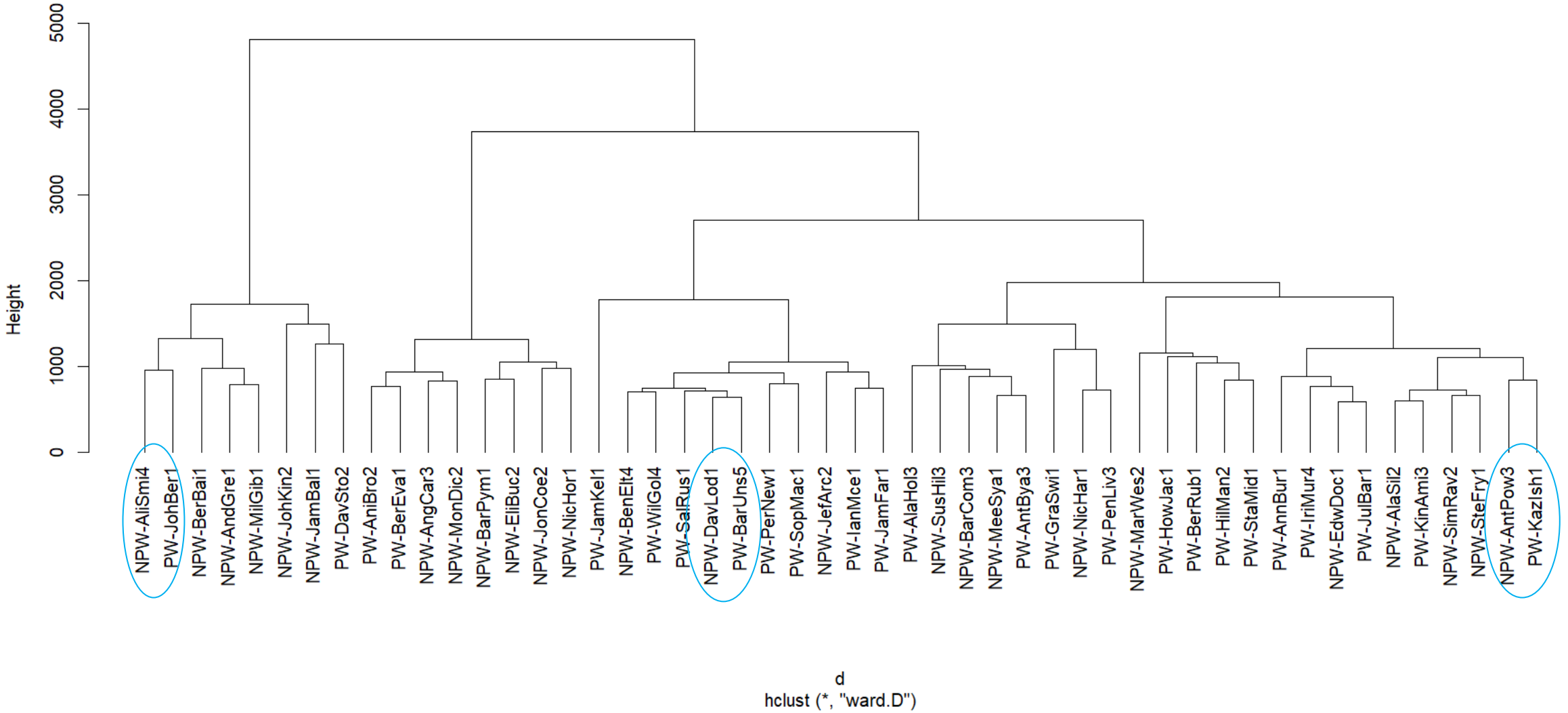
Cluster Dendrogram

PW and ROM: tokens, except numbers



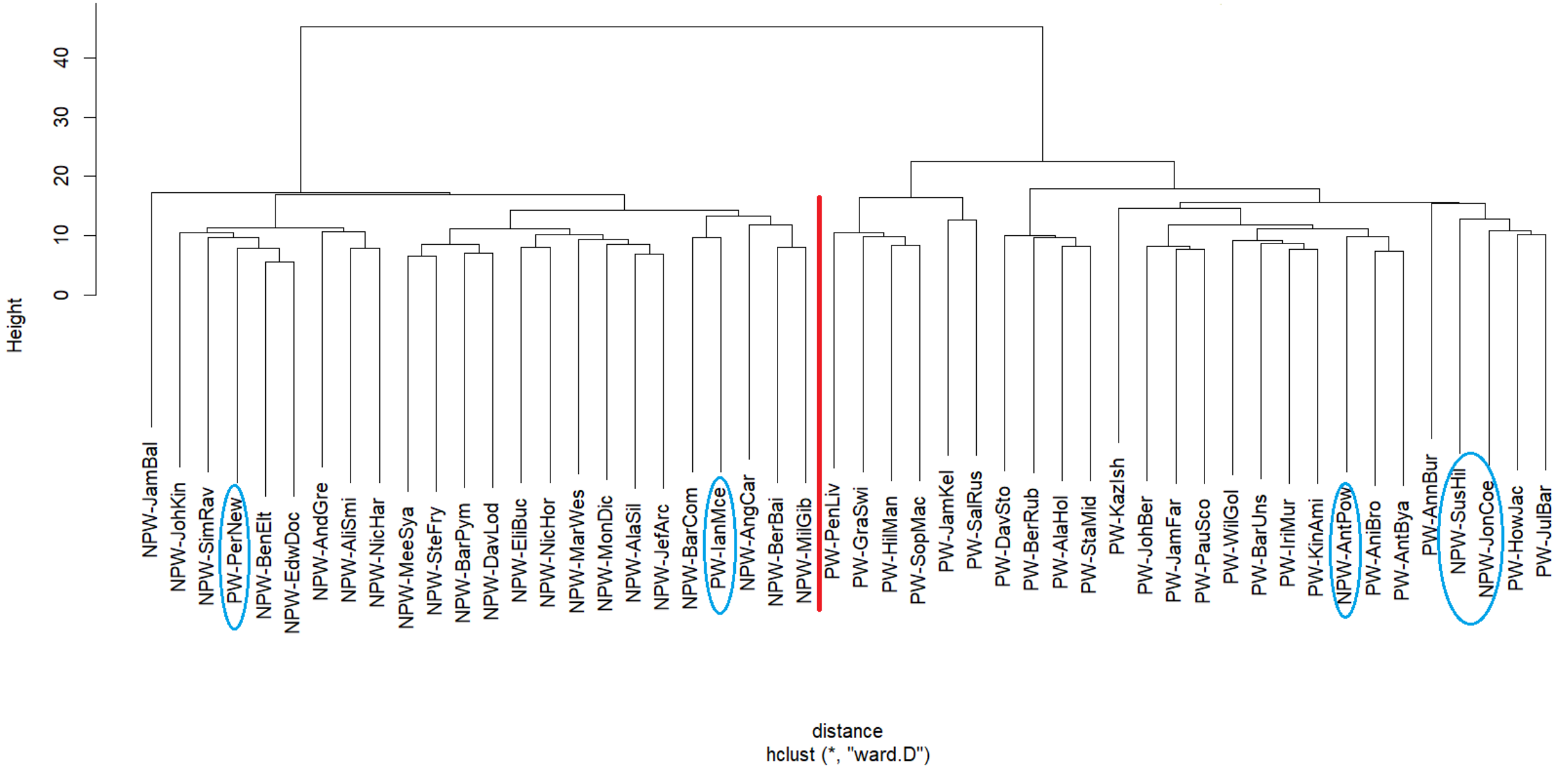
Cluster Dendrogram

PW and NPW: tokens, except numbers



Cluster Dendrogram

PW vs. NPW, parts-of-speech



List of POS bigrams (CLAWS tags, incl. punctuation)

comma DD	comma IW	VVG RT	qmark VVD
IF DDQ	RT II	JJR NN2	NN1 NN2
IO DDQ	IO DD2	RRQ TO	VVG IW
VVN RT	RT IF	CC RL	CC VVD
RT IW	CC IO	DDQ RR	VVI VVD
VVI semicolon	DD2 VBDR	CS DD	qmark VBDZ
CCB II	CC IW	CSA VBDZ	DDQ TO
DDQ VV0	NN2 semicolon	comma RRQ	CCB VDD
IO comma	DA NN2	comma TO	VVD RR
comma VV0	DD1 NN1	II DD1	VVGK TO
DDQ comma	IO VVG	<hr style="border: 2px solid red;"/>	NNT1 VVD
RR NN1	DB NN1	VVI RG	NN2 IF
RR semicolon	NN2 RT	NN IO	VVG AT1
CS XX	CC DDQ	II AT1	RG DA2
comma IO	VV0 VVI	MD IO	NN1 NN1
NN1 semicolon	comma II	VVD CC	IF NNT2
IO DD1	comma RGR	VVD AT1	
VV0 IO	JJR comma	NN2 RRQ	

comma VV0

1. However , **let** us return for the moment to this dog kennel. (RR comma VV0 PPIO2 VVI IF AT NN1 II DD1 NN1 NN1.)
2. And of course she does, in answer to my deliberately cool missive , **protest** too much. (CC RR RR PPHS1 VDZ comma II II II APPGE RR JJ NN1 comma VV0 RG DA1.)
3. There are those who, even if valued , **remain** sinister witnesses from the past. (EX VBR DD2 PNQS comma CS CS VVN comma VV0 JJ NN2 II AT NN1.)
4. I must now, in mentioning my mother , **speak** of Aunt Estelle. (PPIS1 VM RT comma II VVG APPGE NN1 comma VV0 IO NNB NP1.)

Further research

- More data
- Other prizes / prize categories
- Include authors from other countries/cultures
- Include other measures, e.g. amount of (free in)direct speech
- Use more sophisticated statistical analyses/models
- Investigate the bigrams in more detail
- Compare publishers/genders

The End

Bibliography

- Ashok, Vikas G., Song Feng & Yejin Choi. 2013. Success with style: using writing style to predict the success of novels. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington: Association for Computational Linguistics, 1753–1764.
- Biber, Douglas & Susan Conrad. 2009. *Register, Genre and Style*. Cambridge: CUP.
- Bondi, Marina and Mike Scott. 2010. *Keyness in Texts*. Amsterdam: John Benjamins.
<<https://doi.org/10.1075/scl.41>>
- Hoover, David L., Jonathan Culpeper & Kieran O'Halloran. 2014. *Digital Literary Studies. Corpus Approaches to Poetry, Prose and Drama*. London: Routledge.
- Pennebaker, James W., Ryan J. Booth, Roger L. Boyd & Martha E. Francis. 2015. *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).
- Piper, Andrew. 2017. Fictionality. *Journal of Cultural Analytics*, 2(2). <https://doi.org/10.22148/16.011>
- Piper, Andrew & Eva Portelance. 2016. How Cultural Capital Works: Prizewinning Novels, Bestsellers, and the Time of Reading. *Journal of Cultural Analytics/Post45*.
- Stubbs, Michael & Isabel Barth. 2003. Using recurrent phrases as text-type discriminators. A quantitative method and some findings. *Functions of Language* 10(1): 61–104.