



UiO : **University of Oslo**

What can a diachronic corpus of modern
British fiction tell us about changes in
(British) society?



WiP, ICAME39, Tampere 2018

Jarle Ebeling

Signe Oksefjell Ebeling

Aims

- First explorations of a new resource
 1. In what ways does a keyword analysis reflect changes in (British) society?
 2. In what ways does a general and ambiguous keyword (*poor*) reflect changes in (British) society?

Introducing the



Content

- (Extracts of) Novels and short story collections published between 1900 and 2018 by writers born and/or educated in the United Kingdom
- No children's/juvenile fiction, science fiction, fantasy literature or drama
- Texts published between 1900-1940 mainly harvested from Project Gutenberg (.org, .ca, .net.au) and Fadedpage
- More recent texts are from the BNC and other corpora/sources
- March 2018: 510 texts (190 female / 320 male), approx. 37 mill. words and 321 different writers

Method

- Periods
 - P1: 1900-1939
 - P2: 1940-1979
 - P3: 1980-2018
- Generate keyword lists (WordSmith Tools 6 (Scott 2012))
 - P1 vs. ref. corpus (P2+P3)
 - P2 vs. ref. corpus (P1+P3)
 - P3 vs. ref. corpus (P1+P2)
- Classification of keywords into broad semantic categories
- Focused study of the meanings and use of the word *poor*, which is key in P1

Inspired by previous studies using similar techniques, e.g. Leech & Fallon (1992); Scott (2010); McEnery & Baker (2017)

Keyword analysis / Semantic categories

P1	P2	P3
Family relations	War/Military	Technology/inventions
Forms of address	Food & drink	Leisure
Religion	Technology	Indulgence

Examples:

- P1: divine, religion, cousin, aunt, marry, servant(s), captain, lady, lord, honour, bowed
- P2: enemy/ies, smoke, war, uniform, cigarette(s), eating, whisky, telephone(d), receiver
- P3: screen, camera, plastic, TV, pub, sex(ual), shopping, weekend, coffee, wine,

Changes in keywords reflect changes in society and involve a shift from

- a family-centered class society occupied with religion and the divine
to
- a war-ridden society followed by post-war optimism
to
- a technology-centered, pleasure-seeking, affluent society

... because of [...] multiple meanings, it is sometimes impossible, looking at the comparative frequency lists, to judge to what extent a contrast is due to a particular meaning.
(Leech & Fallon, 1992: 34)

The case of *poor*

- Keyword in P1 (1910-1939), but frequent in the CBF overall
- Sharp decrease in the use of *poor* over the century:
 - P1: 393 pmw
 - P2: 292 pmw
 - P3: 169 pmw
- Is there any particular use or meaning that can account for the decrease, and if so, can it be explained from a socio-cultural perspective?

Poor: meaning and use

- Attributive and predicative uses
 - lacking sufficient money – ... because they're poor ...
 - low quality – ... made a poor witness ...
 - deserving pity/sympathy (attributive only): My poor girl ...
- Nominalized adjective
 - lacking sufficient money: ... the poor get poorer...

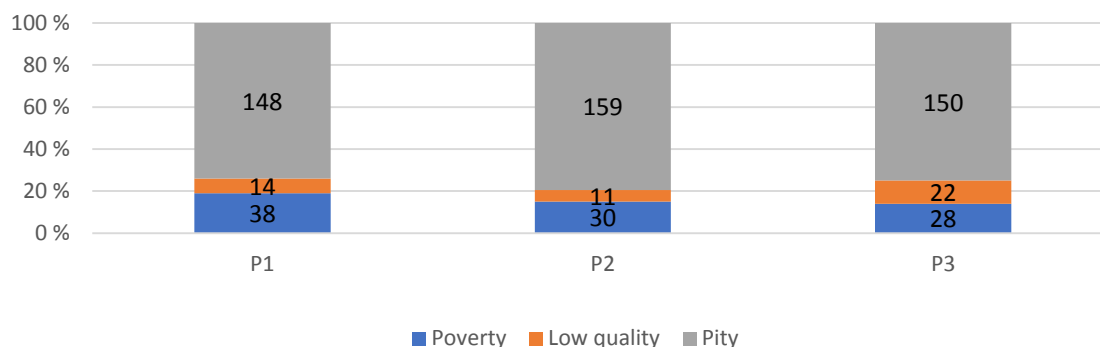
Hypothesis: there will be a marked drop in the use of the «lacking sufficient money»-meaning, which may account for the decrease from P1 to P3.

Random sample of 200 concordance lines per period

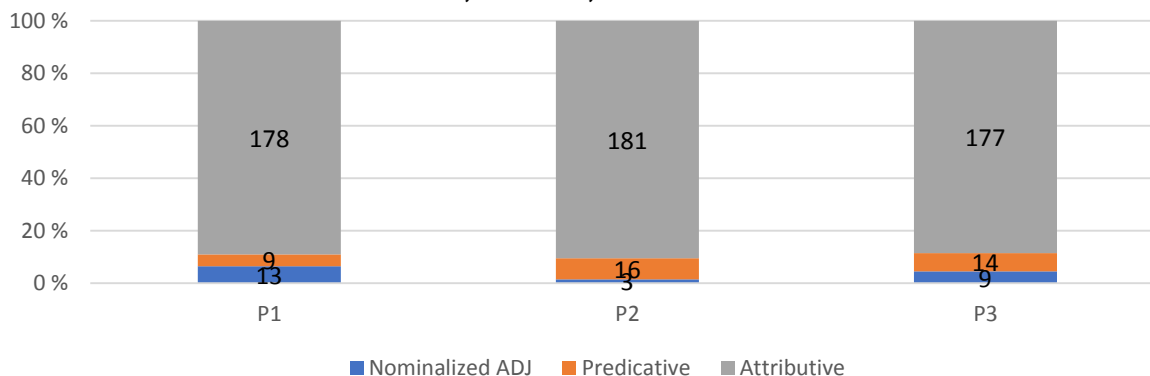
Proportions of meanings and uses in the sample

Distributions of meanings and uses fairly stable across the three periods

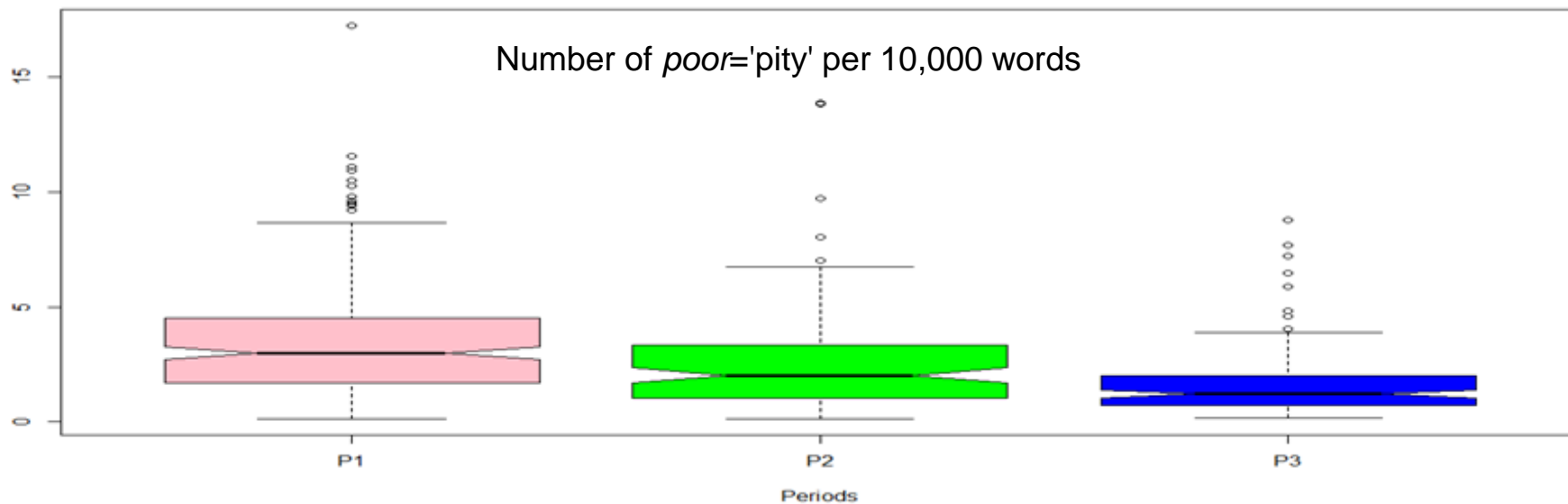
- lacking sufficient money (poverty): P1: **38**, P2: 30, P3: 28
- low quality: P1: 14, P2: 11, P3: 22
- deserving pity/sympathy: P1: 148, P2: 159, P3: 150



- Attributive: P1: 178, P2: 181, P3: 177
- Predicative: P1: 9, P2: 16, P3: 14
- Nominalized ADJ: P1: **13**, P2: 3, P3: 9



Attributive pity/sympathy use



```
> summary(pity_P1$Per)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.130  1.700   2.990   3.499  4.500   17.250

> summary(pity_P2$Per)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.140  1.020   2.030   2.599  3.340   13.890

> summary(pity_P3$Per)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.170  0.740   1.215   1.630  2.005    8.790
```

“Literature reflects society and
society shapes literature.”
-- Oscar Wilde

Concluding remarks

- These first explorations of the CBF show some of the potential of a large diachronic corpus of one broad genre in one variety of English.
- Gives a snapshot of society through literature over a century.
 - A keyword analysis of a corpus of fiction texts seems to mirror society.
 - Even the analysis of the ambiguous keyword *poor* may shed some light on changes in society.
- This WiP report has only scratched the surface of how literature reflects society. Further study is needed to validate the (preliminary) findings presented.

Quiz: exclusive to P1, P2 or P3?

poor

- bastard
- boy
- bugger
- child
- creature
- dear
- devil
- fellow
- girl
- man
- sod
- soul
- thing
- woman

Rank	P1		P2		P3	
1	thing	339	man	94	man	68
2	man	288	<i>devil</i>	64	girl	51
3	<i>dear</i>	246	thing	64	thing	48
4	child	233	girl	52	woman	43
5	<i>devil</i>	205	boy	51	bastard	40
6	<i>fellow</i>	202	<i>fellow</i>	46	sod	33
7	girl	197	<i>dear</i>	41	child	21
8	boy	134	child	40	bugger	19
9	soul	129	woman	34	boy	18
10	woman	124	soul	28	creature	15

Rank 11-14

P1: **chap**, creature, father, **people**

P2: **chap**, **people**, darling, *bastard*

P3: soul, devil, **people**, **chap**



The open part of the corpus (1900-1940):

- <https://nabu.usit.uio.no/hf/ilos/oelc/>

References

Leech, G. & Fallon, R. (1992). Computer corpora – What do they tell us about culture? *ICAME Journal*, 16, 29-50.

McEnery, T. & Baker, H. (2016). *Corpus linguistics and 17th-century prostitution: computational linguistics and history*. London: Bloomsbury Academic.

Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking our trails... In M. Bondi & M. Scott (Eds.), *Keyness in text* (pp. 43-57). Amsterdam: Benjamins.

Scott, M. (2012). WordSmith Tools version 6. Stroud: Lexical Analysis Software.