The interaction between frequency and morphological productivity

## 1. ABSTRACT

## 2. INTRODUCTION

### 2.1. Background

INTRO

Recent studies () have investigated morphpoloigcal productivity in English. X points out how x happened; Y says this.

One aspect of morphological innovation is the spread of morphological innovation over a longer time period. In this paper, two aspects of this spread will be studied: the development of new forms by morphological means as well as the later spread — or failure to spread — of these new words. The study goes on to explore whether these two trends are somehow interrelated. That is to say, to what extent lexical frequency interacts with morphological productivity (Bauer). Previous papers show that ment does X and Y (Bauer from book, new paper).

### 2.2. Statement of purpose

RESEATCH QUESTION

The morphological process in this paper is nominalization of lexical items by adding the suffix –ment, as in to enjoy 🡪 enjoyment. Plag says

### 2.3. Hypotheses

## 3. METHOD

### 3.1. Dataset

DATA, METHOD

This paper is better in that it 1) focuses on true innovations. 2) long term view with a relatively large dataset.

Texts from OTA (so many, so many words). Use Python. Use R. Remove duplicates. Add genres. Add dialect, remove anyone non-British. Metadata include publication date, author age. Consult Michigan library if pubdate is missing.

From 1700 to 1800 cause convenient with copyright

Selection of ment words

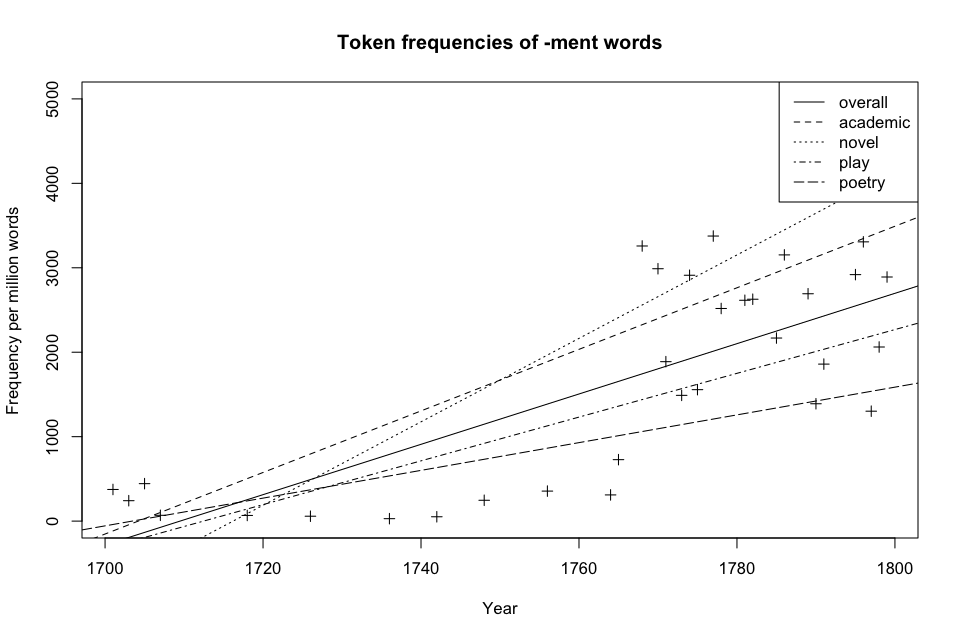
### 3.3. Procedure

### 3.4. Statistical procedures

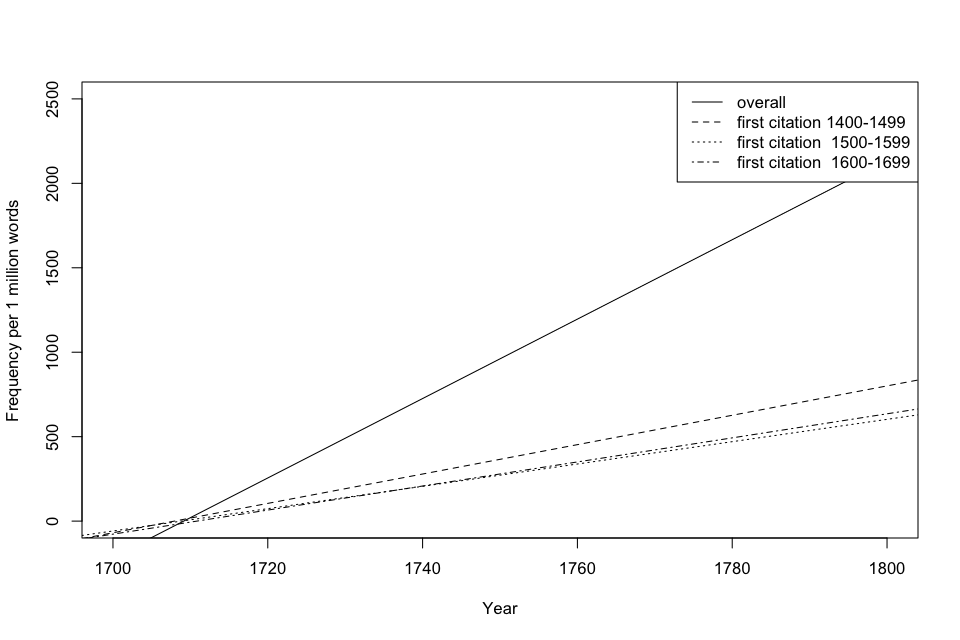
## 4. RESULTS

The development of these 1206 words is plotted below. It is important to note that this tracks the development of words coined or loaned before 1700 only; the addition of new words is not reflected in this graph. The graph shows a marked increase in the use of –ment words. Early-century per million word frequencies are in the low hundreds. In the late 1700s, we see frequencies higher than 3,000 per million words. The regression lines show that this development holds true across genres, but is least pronounced in poetic texts. It is also not limited to individual words; all words in the sample that showed any kind of frequency change did so towards increased use.

Genre cause they have different styles. (Biber)



The plot below illustrates that this development is independent of a word’s first citation.



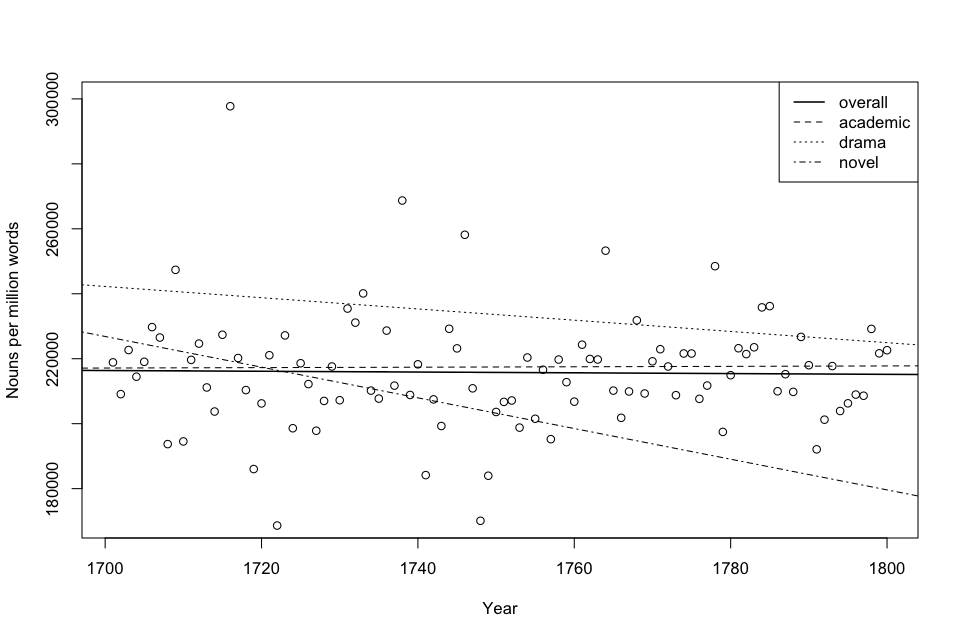
While innovations from the time period 1400 to 1499 dominate with a slope of 8.7, i.e. a statistical increase of 8.7 items per million words each year, the linear regressions have positive slopes for all time periods (6.6 for 1500 to 1599, 7.1 for 1600 to 1799, all statistically significant at p < 0.01).

This supports hypothesis 1 that the use of nouns with the ending –ment increased in English during this time. Even older, well-established words increased in frequency.

A simple explanation for this were if nouns overall increased in frequency during that time period. Writing style can change over time, in recent years toward a more nouny (Biber) lexical choice. If something similar were at work here, the increase in –ment nouns would merely be a reflection of the larger trend towards nouns in general rather this type of nouns.

To investigate, the whole corpus was tagged for Parts of Speech using the Stanford POS Tagger(). Note of caution:

Poetry was excluded from this part of the analysis as the tagger really can’t be expected to do well here.



H1 confirmed: More –ment based nouns over time, when controlling for genre and overall noun occurrence.

H2

Has, this, however, led to increased formation of NEW nouns with the suffix ment? To this end, all genuine native word-formations with ment were extracted from the original list of 1000-odd words.

Only nouns that were pretty clearly identifiable (as far as this is possible, see Bauer ())) as English formations were included. The criteria were 1) identified as “English-internal” in the OED (such as); 2) derived an Anglo-Saxon stem (e.g. settlement); 3) for a Romance origin form, no French parallel form recorded (e.g. enjoyment).

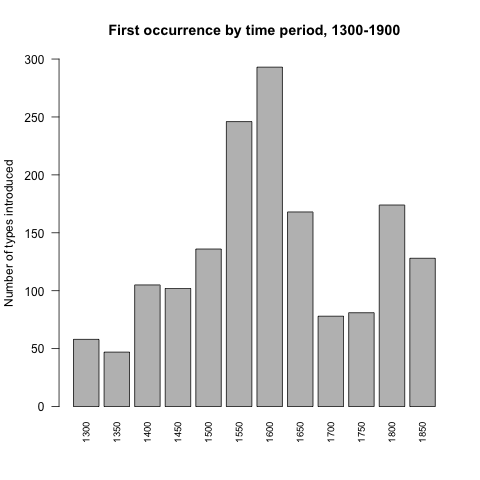
The resulting list of xzx nouns

Neologisms are based on Gadd’s list (1910) of xzy “native formations”. Out of the xzx items, xzx occurred at least once in the dataset. These were retained and double-checked against the Oxford English Dictionary. Only forms that were identified as “derived in within English” or with an Anglo-Saxon stem were kept. “Formed within English, by derivation”

TYPES

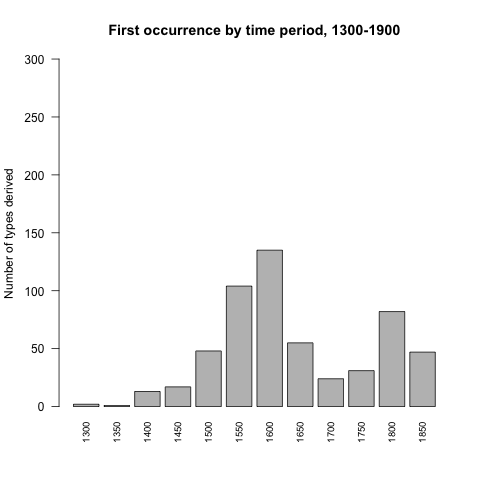
All nouns from the OED. Only ambiguous words like comment and compunds such as act of Parliament removed.

ALL NOUNS



|  |  |  |
| --- | --- | --- |
| Overall introduced | "int" | "freq" |
| "2" | "1" | 58 |
| "3" | "2" | 47 |
| "4" | "3" | 105 |
| "5" | "4" | 102 |
| "6" | "5" | 136 |
| "7" | "6" | 246 |
| "8" | "7" | 293 |
| "9" | "8" | 168 |
| "10" | "9" | 78 |
| "11" | "10" | 81 |
| "12" | "11" | 174 |
| "13" | "12" | 128 |

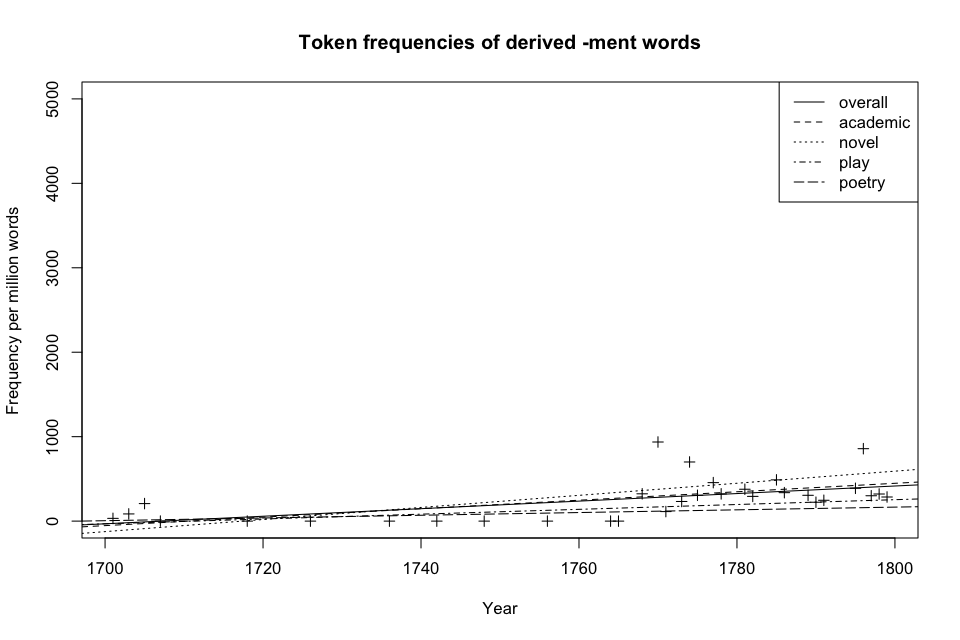
Derived nouns, following Gadd



|  |  |  |
| --- | --- | --- |
| derived | "int" | "freq" |
| "1" | "1" | 2 |
| "2" | "2" | 1 |
| "3" | "3" | 13 |
| "4" | "4" | 17 |
| "5" | "5" | 48 |
| "6" | "6" | 104 |
| "7" | "7" | 135 |
| "8" | "8" | 55 |
| "9" | "9" | 24 |
| "10" | "10" | 31 |
| "11" | "11" | 82 |
| "12" | "12" | 47 |
|  |  |  |

How did these fare tokenwise?

## Macintosh HD:Users:ps22344:Desktop:rplots:allgenres_gaddsyes .pngMacintosh HD:Users:ps22344:Desktop:rplots:allgenres_gaddsyes .png



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| "" | "Group.1" | "x.x" | "x.y" | "freq" |  |
| "1" | 1701 | 2 | 61364 | 32.5923994524477 |  |
| "2" | 1703 | 19 | 223268 | 85.0995216511099 |  |
| "3" | 1705 | 29 | 139756 | 207.50450785655 |  |
| "4" | 1707 | 0 | 72793 | 0 |  |
| "5" | 1718 | 0 | 74468 | 0 | 1725 |
| "6" | 1726 | 0 | 120680 | 0 |  |
| "7" | 1736 | 0 | 138364 | 0 |  |
| "8" | 1742 | 0 | 407671 | 0 |  |
| "9" | 1748 | 0 | 12157 | 0 | 1750 |
| "10" | 1756 | 0 | 8440 | 0 |  |
| "11" | 1764 | 0 | 29009 | 0 |  |
| "12" | 1765 | 0 | 1374 | 0 |  |
| "13" | 1768 | 26 | 80416 | 323.318742538798 |  |
| "14" | 1770 | 21 | 22417 | 936.789044029085 |  |
| "15" | 1771 | 7 | 61966 | 112.965174450505 |  |
| "16" | 1773 | 96 | 412961 | 232.467472715341 |  |
| "17" | 1774 | 19 | 27134 | 700.228495614358 | 1775 |
| "18" | 1775 | 23 | 75158 | 306.021980361372 |  |
| "19" | 1777 | 27 | 59245 | 455.73466115284 |  |
| "20" | 1778 | 211 | 653722 | 322.76717014266 |  |
| "21" | 1781 | 14 | 37091 | 377.450055269472 |  |
| "22" | 1782 | 35 | 119504 | 292.877225866917 |  |
| "23" | 1785 | 57 | 116697 | 488.444433018844 |  |
| "24" | 1786 | 28 | 83108 | 336.911007363912 |  |
| "25" | 1789 | 57 | 186448 | 305.71526645499 |  |
| "26" | 1790 | 32 | 143192 | 223.476171853176 |  |
| "27" | 1791 | 28 | 113485 | 246.728642551879 |  |
| "28" | 1795 | 51 | 131210 | 388.689886441582 |  |
| "29" | 1796 | 180 | 209854 | 857.739190103596 |  |
| "30" | 1797 | 3 | 9987 | 300.390507659958 |  |
| "31" | 1798 | 92 | 286639 | 320.96120904692 |  |
| "32" | 1799 | 8 | 28020 | 285.510349750178 |  |

|  |  |
| --- | --- |
| 1700-1725 | 87.4662598 / mio words |
| 1725-1750 | 0 |
| 1750-1775 | 262.5377301 |
| 1775-1800 | 327.3890043 |

Call:

lm(formula = merger$freq ~ merger$Group.1)

Residuals:

Min 1Q Median 3Q Max

-259.14 -128.93 -45.56 53.04 655.27

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -7641.227 2129.426 -3.588 0.001167 \*\*

merger$Group.1 4.476 1.207 3.708 0.000845 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 206.8 on 30 degrees of freedom

Multiple R-squared: 0.3143, Adjusted R-squared: 0.2915

F-statistic: 13.75 on 1 and 30 DF, p-value: 0.0008452

-Make these pre-1700 innovations only DONE

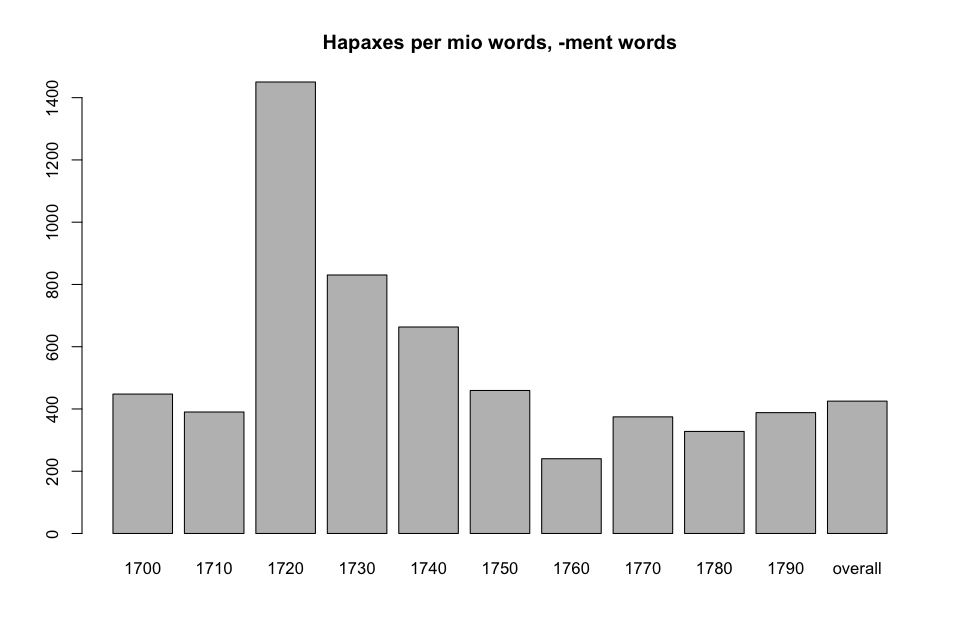
-normalized per word:

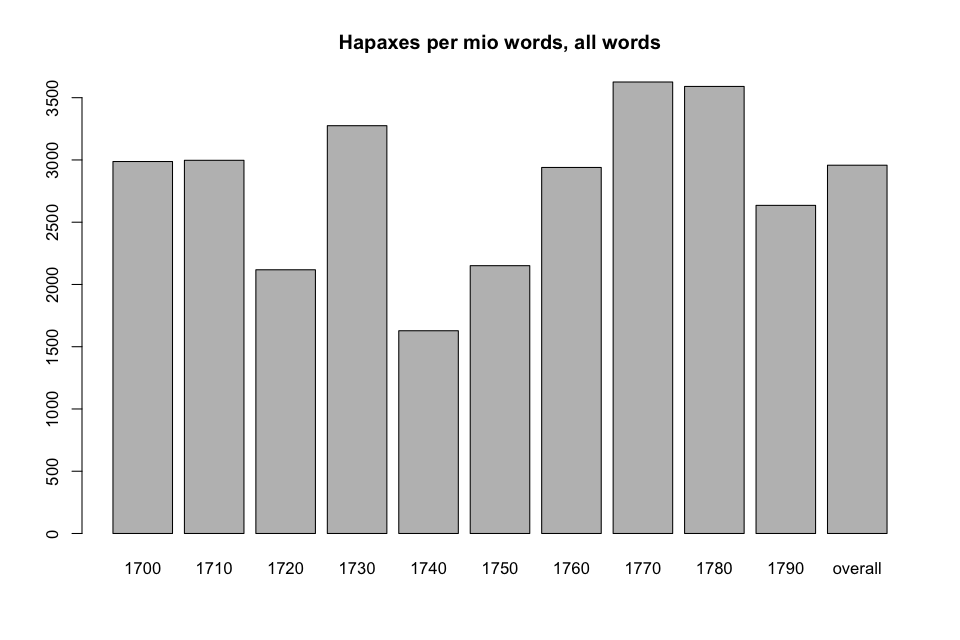
Gadds is run on 428, out of which 119 actually occur

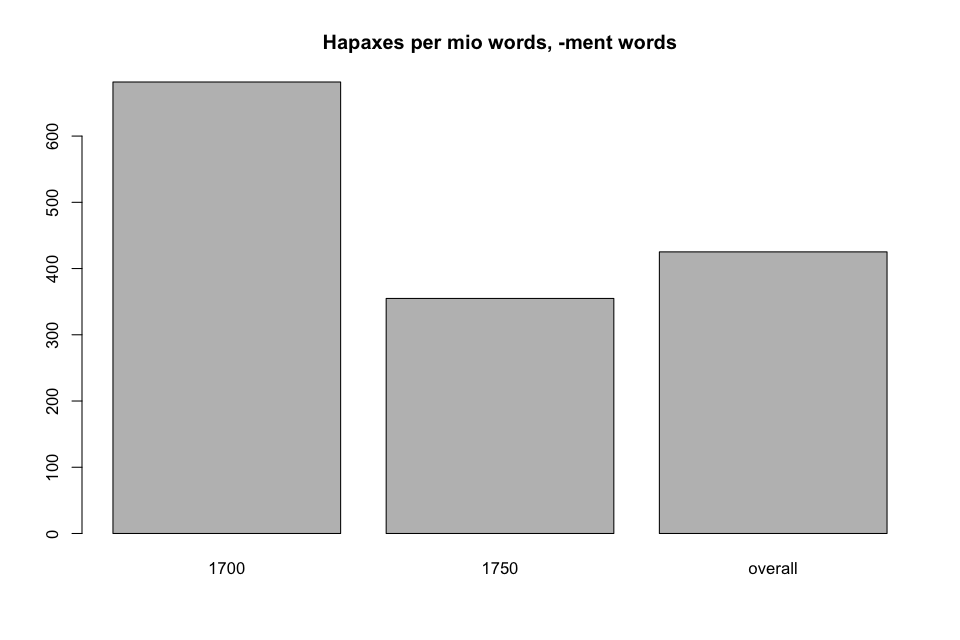
Alllist on 1175, out of which ??? actually occur

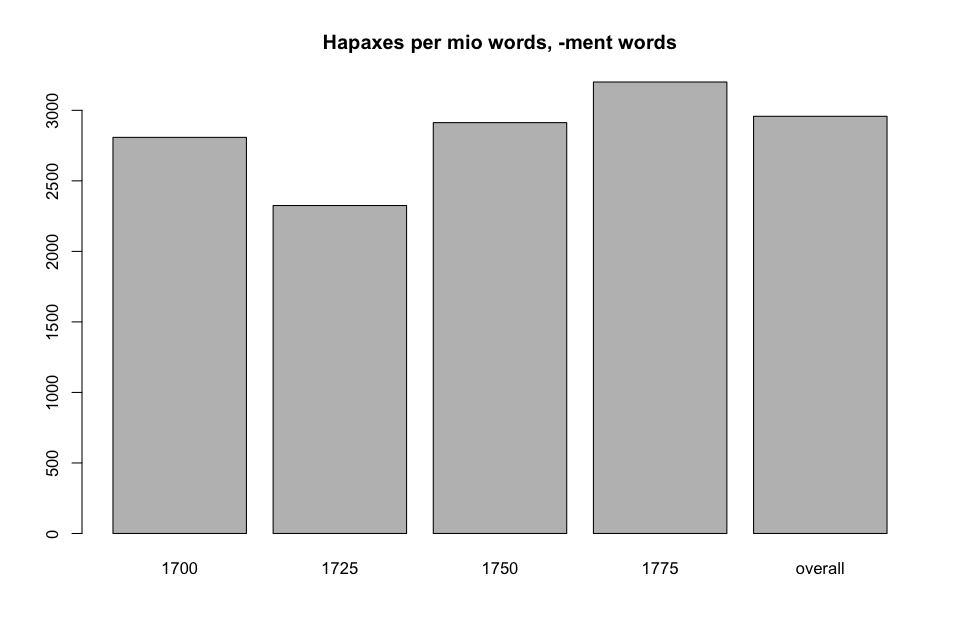
HAPAXES

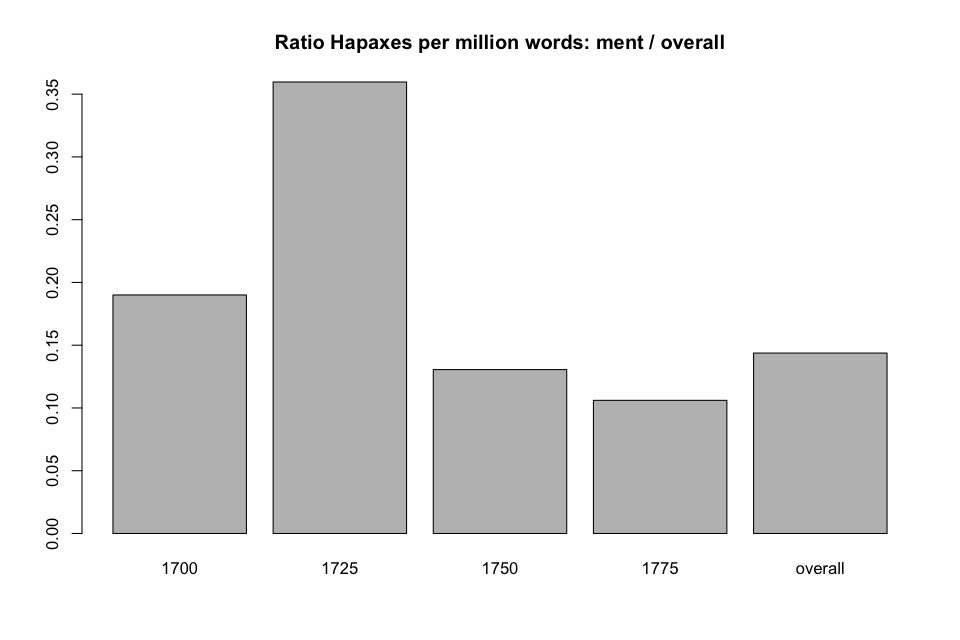
We extract one off words.











## 5. DISCUSSION

## 6. REFERENCES

FINDINGS

OVERALL

ONLY NEOLOGISMS

The results plotted in graph 1 suggest that the frequency of these neologisms increased quite drastically over the time period 1700 – 1800. We see a marked increase after around the year 1750; this trend holds true for all genres, most pronounced in firction writing and least in poetry.