Variation in pronoun frequencies in early English letters

Gender-based or relationship-based?
Gender differences in pronoun use

- Women use more pronouns than men in LME and EModE letters (Säily et al. 2011)
  - Similar observations for PDE (Rayson et al. 1997; Argamon et al. 2003)
- Differences explained in terms of gendered styles
  - A high frequency of 1st and 2nd person pronouns indicates an involved style (Biber 1988, 1995; Biber & Burges 2000)
- Women’s style is more involved, whereas men’s style is more informational
Relationship between sender and recipient?

• Säily et al. (2011) mostly ignored the influence of register, audience design, and corpus balance

• We explore the effect of audience design on pronoun use in 17th and 18th century letters

To what extent are pronoun frequencies affected by the relationship between the sender and the recipient of the letter?

• Focus on 1st and 2nd person pronouns I and you, two markers of involved style
Sample from the Corpora of Early English Correspondence

- Letters from 1600–1800
  - 3.85 million words (0.94 million by women, 2.91 million by men)
- Categories based on the sender-recipient relationship
  - Letters written by or addressed to
    1. Members of the nuclear family (FN)
    2. Other family members (FO)
    3. Family servants (FS)
    4. Close friends (TC)
    5. All other acquaintances (T)
### Relationship categories in CEEC

Number of words in different relationship categories

<table>
<thead>
<tr>
<th>Relationship Category</th>
<th>Women</th>
<th>Men</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>445,678</td>
<td>739,724</td>
<td>1,185,402</td>
</tr>
<tr>
<td>FO</td>
<td>164,303</td>
<td>332,818</td>
<td>497,121</td>
</tr>
<tr>
<td>FS</td>
<td>3,486</td>
<td>25,069</td>
<td>28,555</td>
</tr>
<tr>
<td>TC</td>
<td>179,378</td>
<td>709,580</td>
<td>888,958</td>
</tr>
<tr>
<td>T</td>
<td>150,341</td>
<td>1,103,555</td>
<td>1,253,896</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>943,186</td>
<td>2,910,746</td>
<td>3,853,932</td>
</tr>
</tbody>
</table>
## Material

### Family letters studied

<table>
<thead>
<tr>
<th>Relationship</th>
<th>1600–1679</th>
<th>1680–1800</th>
<th>Entire period</th>
</tr>
</thead>
<tbody>
<tr>
<td>husband &gt; wife</td>
<td>101,015</td>
<td>15,716</td>
<td>116,731</td>
</tr>
<tr>
<td>wife &gt; husband</td>
<td>31,502</td>
<td>51,910</td>
<td>83,412</td>
</tr>
<tr>
<td>father &gt; son</td>
<td>26,756</td>
<td>39,560</td>
<td>66,316</td>
</tr>
<tr>
<td>son &gt; father</td>
<td>29,090</td>
<td>94,931</td>
<td>124,021</td>
</tr>
<tr>
<td>mother &gt; son</td>
<td>41,734</td>
<td>18,902</td>
<td>60,636</td>
</tr>
<tr>
<td>son &gt; mother</td>
<td>25,039</td>
<td>15,381</td>
<td>40,420</td>
</tr>
<tr>
<td>Total</td>
<td>255,136</td>
<td>236,400</td>
<td>491,536</td>
</tr>
</tbody>
</table>
Outliers and corpus balance

- Aggregate normalised frequencies (e.g. M, F) problematic due to outliers
  - Use smaller samples
    - Sample: person’s letters from a 20-year period addressed to a specific relationship category
- Compare categories based on median frequency of samples
  - Visualisation: beanplots (Kampstra 2008) with median frequencies
Visualisation: beanplots

Method

Frequency of *she*+*her*

Median freq (F)

Median freq (M)

Median

freq (M)

Median

freq (F)

Frequency

of *she*+*her*

0%

1%

2%

3%

4%

5%

6%
Statistical significance

- **Wilcoxon rank-sum test** (Wilcoxon 1945; Mann & Whitney 1947)
  - Order samples based on normalised frequency, see how surprising the result is
- 1000s of hypotheses tested → **false discovery rate control** (Benjamini & Hochberg 1995) at FDR = 0.1
  - Significance threshold \( p < 0.0035 \)
- Thanks to Jukka Suomela for methodological assistance!
I am (I slept, God) restored to a perfect degree of 
Health; at least of perfect as at this Time of Day, I can ever 
expect to bee, & (I slept God) it is a good One.

My whole Family alsoe, with your whole 
Uncle's & yo. friends taking in my Way, are well, & on all opportunities 
are yo. Lorde Rememberers. We desire your Advice.

Your truly affected Wife.
Results

Overall gender differences: /
Results

Overall gender differences: *you*

![Graph showing overall gender differences for the pronoun "you" across different time periods.]
Results

Overall gender differences

- Gender affects pronoun use
- Women use more *I* and *you* than men
- **BUT**
  - The difference is not significant in every period
  - Variation across time (a levelling tendency in the 18th century)
  - Women: more FN letters, men: more T letters
Results

Pronoun use within the nuclear family

• Letters written by men
  1. Son – father
  2. Father – son
  3. Husband – wife

• Letters written by women
  1. Wife – husband
  2. Mother – son
Case 1: Fathers vs. sons
Results

Father – son vs. son – father: I + you

![Graph showing the frequency of pronouns in early English letters, comparing 'I' and 'you' usage between father-son and son-father relationships over the periods 1600-1679 and 1680-1800.](image)
Results

Father – son vs. son – father: /
Results

Father – son vs. son – father: you

![Graph showing the frequency of pronouns over time]

- Father-son vs. son-father relationships in early English letters
- Analysis of pronoun frequency over the period 1600-1800
Sons use / more than fathers

My Lord, That dutie which inioynes this taske, is not more the same then the matter which I have to write of. In that method of exercises, and studie, which I have heretofore made known unto your Lordship doe I stil goe on: though in the successe, you shall perhaps meet with little that may take you, yet I hope in my desires you will find something at least capable of your pardon. I am drawing nigh to Fortifications, but such as imagination is able to raise, and to demolish: if honor were not so strictly espoused to danger, I could wish your Lordship had at this time no harder a taske in hand, that I might boldly presume of your safety…

(Edward Conway, aged 17, to his father, 1640)
Honest Tom, God blesse & protect thee & mercifully lead you through the wayes of his providence. I am much greived you have such a cold, sharpe & hard introduction, wch addes newe feares unto mee for your health, whereof pray bee carefull & as good an husband as possible, wch will gayne you credit & make you better trusted in all affayres. I am sorry you went unprovided with bookes without which you cannot well spend time in those great shipps. If you have a globe you may easily learne the starres as also by bookes. Waggoner you will not bee without wch will teach the particular coasts, depths of roades & how the land riseth upon severall poyns of the compasse. Observe the variation of the compasse. Blundevill or Moxon will teach you severall things.

(Sir Thomas Browne, aged 60, to his son, 1665)
Case 2: Husbands vs. sons
Results

Husband – wife vs. son – father: I

![Diagram showing variations in pronoun frequencies between husband-wife and son-father relationships over time.]
Results

Husband – wife vs. son – father: you
Results

Husbands use I even more and you even less than sons

My busines with Jones falls out still vntowardly, ther being a compaekte of knavery amongst them to defraud the creditors but I hope to prevaille for my mony notwithstanding in a short time. I haue made inquiry of the land of Starbourg, and haue spoken with Beareblock, which Mr. Seriant Richardson did take very ill, fearing least I had made an agreement with him for his statute, which I would to god I had. I cannot tell yett how we shall agree but as neere as I can I will deale to our best advantage.

(Thomas Knyvett, aged 26, to his wife, 1622)
Results

Husbands, sons and fathers

- Implicational scale for *I*

  Husbands > Sons > Fathers

- Implicational scale for *you*

  Fathers > Sons > Husbands

- Statistical significance only for husbands vs. fathers, 1600–1679 (both *I* and *you*)
### Results

**Family letters in CEEC**

<table>
<thead>
<tr>
<th>Role Relation</th>
<th>1600–1679</th>
<th>1680–1800</th>
<th>Entire period</th>
</tr>
</thead>
<tbody>
<tr>
<td>husband &gt; wife</td>
<td>101,015</td>
<td>15,716</td>
<td>116,731</td>
</tr>
<tr>
<td>wife &gt; husband</td>
<td>31,502</td>
<td>51,910</td>
<td>83,412</td>
</tr>
<tr>
<td>father &gt; son</td>
<td>26,756</td>
<td>39,560</td>
<td>66,316</td>
</tr>
<tr>
<td>son &gt; father</td>
<td>29,090</td>
<td>94,931</td>
<td>124,021</td>
</tr>
<tr>
<td>mother &gt; son</td>
<td>41,734</td>
<td>18,902</td>
<td>60,636</td>
</tr>
<tr>
<td>son &gt; mother</td>
<td>25,039</td>
<td>15,381</td>
<td>40,420</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>255,136</td>
<td>236,400</td>
<td>491,536</td>
</tr>
</tbody>
</table>
Importance of relationship categories to corpus design
Results

Fathers vs. women
Results

Father – son vs. all women: I and you (1600–1679)
Results

Husbands vs. women
Results

Husband – wife vs. all women: I and you (1600–1679)
Results

Gender differences within the nuclear family

Arthur Capel, 1st Earl of Essex with his wife, Elizabeth, Countess of Essex
Gender differences within the nuclear family

<table>
<thead>
<tr>
<th>Sender-recipient</th>
<th>Sender-recipient</th>
<th>Period</th>
<th>Pro</th>
<th>+/-</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother-son</td>
<td>Father-son</td>
<td>1600–1679</td>
<td>I</td>
<td>+/-</td>
<td>0.00286</td>
</tr>
<tr>
<td>Wife-husband</td>
<td>Husband-wife</td>
<td>1680–1800</td>
<td>I</td>
<td>+/-</td>
<td>0.00226</td>
</tr>
<tr>
<td>Wife-husband</td>
<td>Husband-wife</td>
<td>1600–1679</td>
<td>You</td>
<td>+/-</td>
<td>0.00040</td>
</tr>
<tr>
<td>Mother-son</td>
<td>Son-mother</td>
<td>1600–1679</td>
<td>You</td>
<td>+/-</td>
<td>0.00005</td>
</tr>
</tbody>
</table>
Results

Gender differences within the nuclear family

- The results point to women’s overuse of pronouns
- Most of the differences observed in our study are strong tendencies
  - Statistically significant results only obtained when women use more pronouns than men
- Husband–wife letters (1600–1679)
  - The only category where men use more / than women
  - Not statistically significant
Husbands vs. wives: I (1600–1679)
**Conclusion**

**Gender-based or relationship-based?**

- Sender–recipient relationship is clearly important
- Gender is a more important factor
  - In inter-gender comparisons (e.g. husband–wife vs. wife–husband) women consistently used more pronouns than men
- Relationship categories should be taken into account in both research and corpus design
Thank you!
References


CEECE = Corpus of Early English Correspondence Extension. Compiled by T. Nevalainen, H. Raumolin-Brunberg, S. Kaislaniemi, M. Laitinen, M. Nevala, A. Nurmi, M. Palander-Collin, T. Säily and A. Sairio at the Department of Modern Languages, University of Helsinki.

CEECESU = Corpus of Early English Correspondence Supplement. Compiled by T. Nevalainen, H. Raumolin-Brunberg, S. Kaislaniemi, M. Laitinen, M. Nevala, A. Nurmi, M. Palander-Collin, T. Säily and A. Sairio at the Department of Modern Languages, University of Helsinki.


