Women scientific discourse: compiling a corpus, analysing the language.

Begoña Crespo
Muste group
www.udc.es/grupos/muste
University of A Coruña, Spain
bcrespo@udc.es
RESEARCH PROJECT funded by Xunta de Galicia:

WOMEN AND SCIENTIFIC DISCOURSE: SOCIAL CONTEXT AND VARIATION IN ENGLISH (1700-1930)

Research team: 6 people (but actually 3 working on it)
- Rationale
- Main aim and procedure
- Corpus compilation
- Exploitation
  - Pilot Studies
Rationale

- Increasing number of studies claiming that women have contributed to the development of science ever since the very beginning
  - no academic training
  - domestic roles /roles of minor social consideration.

- A male-dominant view of the history of science has not always credited these women with their earned recognition (Solsona I Pairó, 1996; Schiebinger, 1989; 2006; Hunter, 1997, 2005; Hutton, 1997).

- Social prejudices of the times
  - worked with husbands or male relatives, publishing their works anonymously or with a pseudonym.
Aim and procedure I

- to study the characteristics of scientific English by female writers of science (1700-1930)

HOW TO PROCEED

- compilation of two corpora:
  - FeSciT, Female Scientific Texts, works
  - PreWoS, Prefaces by Women Scientists, prefaces to texts
Procedure II

**FeSciT, Female Scientific Texts**

1. In-depth description of female scientific language (different levels of analysis, syntactic, morphological, semantic, discursive and pragmatic…)
2. *Coruña Corpus of English Scientific Writing*
Comparison between male and female scientific Writing.

Contribution to the history of

*English for Specific Purposes !!!*
Procedure III

*PreWoS, Prefaces by Women Scientists*

- Position of the scientist in society,
- relationship with the readership
- strategies used in the communicative process
- how the genre of the texts they precede may have an effect on the writer-reader relationship.

Contribution to the history (of science, women)
Corpus compilation I

Principles

Time-span: 1700-1930
English-speaking women scientists:

• Authors/writers
• Scientific assistants to their husbands or other male relatives
• Editors and translators
• Botanic illustrators, collectors of fossils, etc.

SOURCES

*The Biographical Dictionary of Women in Science (BDWS)*, (Ogilvie & Harvey, 2000)
*Oxford Dictionary of National Biography (ODNB)*, (Several Authors, 2004)
Corpus compilation II
What can we do with all these data?
create a webpage containing information about these women (lives and works)

Create a search tool with keywords sociolinguistic information.
<table>
<thead>
<tr>
<th><strong>Surname, married (single)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Birth and Death</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Place of education 1</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Place of education 2</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Manner of education</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Post/occupation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Connection with science</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Family: father’s occupation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mother’s occupation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Single/married</strong></td>
<td>S/M</td>
</tr>
<tr>
<td><strong>Marriages</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Husband’s name</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Husband’s occupation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Children</strong></td>
<td>Y/N</td>
</tr>
<tr>
<td><strong>Science writer</strong></td>
<td>Y/N</td>
</tr>
<tr>
<td><strong>Field/discipline</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Contacts, mentors...</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Titles of works</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Work selected</strong></td>
<td></td>
</tr>
<tr>
<td>** Discipline**</td>
<td></td>
</tr>
<tr>
<td><strong>Year of publication</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Place of Publication</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Genre</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Preface</strong></td>
<td>Y/N</td>
</tr>
<tr>
<td><strong>Dedicatory</strong></td>
<td>Y/N</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td>Y/N</td>
</tr>
<tr>
<td><strong>Foreword</strong></td>
<td>Y/N</td>
</tr>
<tr>
<td><strong>Additional information</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td></td>
</tr>
</tbody>
</table>
The results of this work can be seen in

www.womenscientistsudc.com
Corpus compilation III

Second database (min. two authors per decade)
Corpus compilation IV

COMPILATION PRINCIPLES

Both *FeSciT* and *PreWoS* share most compilation principles:

1. Women who wrote and published between 1700 and 1930.
2. Only English-speaking women (educated in English).
3. Any scientific discipline according to the UNESCO classification of sciences.
4. Journal articles have been disregarded for the compilation of *PreWoS*.
5. As a starting point, two texts as well as prefaces per decade have been selected and they have been included in toto.

*Coruña Corpus*

- Edited and printed prose texts only
- First editions or others within 30-years period (Kytö, Rudanko and Smittenberg’s 2000: 92)
- No translations
As we have access to texts and conveniently save them in the format in which we find them (pdf, txt., html). Conversion to txt. files.

These can be easily uploaded into one of these search engines: **CQP web, wmatrix** and **antconc**.

wmatrix-3: the semantic tagging of texts.

To create an interface with the two corpora, use **CC as a reference corpus**
Pilot Studies


Material: women authors writing texts on astronomy, philosophy, life sciences and history during the period from 1700 to 1900

Findings: statistically significant reduction in the use of conditionals
They also classified the functions of conditionals in the text using Warchal's (2010) typology,
and found that the majority of conditionals were epistemics (conditionals represent a process of reasoning).

Replicated


Material: eighteenth-century male and female-authored philosophy and life sciences texts

Findings: it was found that content conditionals were much more frequently used by men than by women,
Pilot Studies

Research question: persuasion/argumentation
Four linguistic strategies: Conditional subordinators, Suasive verbs, Predictive modals and Necessity modals

Material: Texts in CEPhiT: 18th c.

Findings: Different kinds of persuasive strategies

• male: modality
• female: conditional and suasive verbs (more subtle)
Research question: Involvement in prefaces
Material: Eighteenth & nineteenth c. prefaces to scientific texts by women

Linguistic features: contractions, first-person pronouns, pronoun *it*, second person pronouns, *not*-negation, demonstrative pronouns, emphatics, causative subordination, discourse particles, indefinite pronouns, hedges, amplifiers, private verbs, other adverbial subordinators and possibility modals.

**Findings**: most abundant feature personal pronouns > followed by the pronoun *it*, possibility modals and private verbs in descending order. **ACTIVE INVOLVEMENT**

- overt authorial presence in these texts manifesting the importance of reasserting themselves and interacting with the reader.

- combination of first person pronouns with private verbs (for the expression of attitudes, feelings and inner thoughts) strengthens the writer’s personal commitment with her scientific work.

- Closeness to the reading public use of colloquial and general proforms such as *it* in seeking the reader’s confidence and reliability.
Research question: Involvement in prefaces and texts
Material: Eighteenth & nineteenth c. prefaces to scientific texts by women
Findings:
overall proportion of features expressing involvement in main texts surpasses that of cases found in front matter

predominance of involvement features in the main texts, creating a sense of closeness between author and reader ???

Texts: wh-words, private verbs and 1\textsuperscript{st} and 2\textsuperscript{nd} person pronouns appear in this order.
Prefaces: order is reversed, 1\textsuperscript{st} and 2nd person pronouns are the most abundant elements, followed by private verbs and wh-words.
Thank you for your attention!

bcrespo@udc.es