**LINGUISTIC QUESTIONS**

Social meaning of spelling variation in historical periods of English and Finnish

Neologisms in early English correspondence

**TOOLS AND MATERIALS**

Developing a modular research toolkit for sociolinguistic analysis

**RESEARCH GROUP:**

Terttu Nevalainen (PI; University of Helsinki)
Anni Sairio (University of Helsinki)
Tanja Säily (University of Helsinki)
Anna Merikallio (University of Helsinki)

Taru Nordlund (PI; University of Helsinki)
Katja Litola (University of Helsinki)
Johanna Utriainen (University of Helsinki)

Eetu Mäkelä (Aalto University)
Poika Isokoski (PI; University of Tampere)
Harri Siirtola (University of Tampere)
Significance

• To develop novel research methods in digital humanities
• In STRATAS, more specifically, to further historical sociolinguistic research
  → to enable new questions that address social meaning
    • Sociolinguistics studies language in society, with the basic starting points that 1) language variation is never random, but it carries social meaning, and 2) language forms a fundamental means for human expression of identity
    • The project looks for new means to capture this inherent feature of language
  → the new modular language independent tools will also enable new ways of exploring linguistic data, and make linguists ask questions that they never thought before...

UNIVERSITY OF HELSINKI
ERRATAS: checking for orthographical reliability of editions of historical texts (Kaislaniemi et al.)

Historical linguists generally use edited sources, to allow ready access to large amounts of data. But how philologically reliable are these editions? The ERRATAS subproject investigates this, & will create a typology of editorial reliability.

---

The ERRATAS subproject investigates the reliability of edited historical texts. These editions are often used by linguists and scholars to access large amounts of data. However, the question arises: how reliable are these editions? The ERRATAS project aims to address this issue by creating a typology of editorial reliability.
Can we trust the spellings in editions of historical texts?

Even editions perceived as ‘good’ change numerous aspects of the texts. This makes them unreliable for studying spelling – although they can still be used to study e.g. syntax or morphology. But by understanding the degree of orthographical reliability of an edition, we can look at a broader range of linguistic features in its texts, including specific orthographical features. This will allow us to expand the scope of use of all editions.

1. historical spellings modernized
2. word divisions normalized
3. abbreviations expanded
4. capitalisation modernized
5. Punctuation added
6. Transcription errors

- ss => s

mvch l hau
will a dooun
meaco

UNIVERSITY OF HELSINKI
19th century Finland as a socio-linguistic laboratory (Nordlund et al.)

- Modern standard of written Finnish was constructed within a brief time period of the 19th century
  - Connected to modernization, nation-building, industrialization, urbanization and the changes in the world-view
- Finnish makes an interesting case because:
  - The standard was consciously constructed on the basis of different dialects
  - The process was short, and some changes are still ongoing
  - Makes a comparison to languages and countries with different linguistic and societal structures
• The social border had previously been drawn between two languages, Swedish and Finnish, and Swedish was predominantly the written language up until 1809 when Finland became an autonomous grand duchy of the Russian empire
• Finnish was mainly used in religious settings and administration when necessary

→ In the 19th century, Finnish language and its varieties received new types of social meanings, for example, power, prestige, and solidarity
→ because the new standard was currently being built, linguistic variation can reveal how these meanings came to connected to specific linguistic elements

1st stage of the project:
Corpus of letters with social, regional and temporal balance

Photographs: Finnish Literature Archives (Helsinki).
Language standardisation streamlined 18th-century English spelling, but private writing retained variation. Two case studies:

1. The social history of -d. “I coloured, M=r= Donnellan twitterd, D=r= Courayer sutterd half french half English” [1747]
   → social connotations and life span of this inconspicuous form? feature of private (women’s?) writing? In public writing at all?

2. Spelling and social stigma: negative contractions. “if you dont like the letter pray burn it” [1742]
   → trajectory of these stigmatized yet vibrant variants? Gendered writing, social stratification?
Neologisms in early English letters (Säily et al.)

- Previous research: dictionary data, bias towards well-known authors
- Our letter corpus: wide social spectrum, speech-like genre
- Research questions:
  - **Who** are the innovators?  
  - **How** do the new words spread?  
  - **Which semantic domains** do the neologisms represent?  
  - **Why** are the neologisms created and established?
- Challenges:
  - **Identification** of neologisms (spelling variation…)
  - Automatic retrieval of related **lexicographical** data (dates, meanings)
  - **Interface** for pruning the possible neologisms found

Neologisms in early English letters (Säily et al.)
Tool Development

Eetu Mäkelä
Semantic Computing Research Group (SeCo)
Department of Computer Science
eetu.makela@aalto.fi - http://seco.cs.aalto.fi/u/kiemakel/
Context: Enabling a Virtuous Cycle of Research in the Humanities (VISCERA)

0. Formulate research questions

1. Discover relevant data

2. Ingest and integrate data

3. Explore and interpret data

4. Publish interpretations, also as data

How?
Contexts:

- Reassembling the Republic of Letters (EU COST Action)
- Cultures of Knowledge (Oxford)
- Linked Open Data Science Service (Aalto)
- Humanities + Design Laboratory @ Stanford, the University of Colorado, the University of Helsinki, Finnish public libraries, …
- STRATAS

Tools:

- Palladio, VISU, CORE, Aether, ARPA, HistOmor, Recon, Mare, …
- Khepri

http://seco.cs.aalto.fi/u/jiemakel/viscera
CLEMENT PASTON to JOHN PASTON on 25 August, 1461

<Q A 1461 FN CPASTON> X CLEMENT PASTON >P 1.193> || [118. TO JOHN PASTON I 1461, 25 AUGUST] || "To his rythe reuerent and worshipful broder John Paston." Rythe reuerent and worshipful broder, I recomand youe to gowre good broderhood, desiring to here of gowre welfare and good prosperite. The quyche I pray God ancresse to his pleasure and gowre hertys hease; currying gowre - tat I have spak wyth John Avass, and Maytor spak wyth hym bothe, on Fryday be fore Seynt Ioseph o sum were, tat - tat gé weold have hynm hom, + te quyche xwicd cause hym not to be haddle in fauor; and also men wold thynke + tat he were put owle of service. Also W. Peck tellythe me + tat hys mony is spent, and not rightes but wissy and disclers, fore + te covty is gretter in - te Kyngys house chase be rythye + tat - tee xwicd hedde ba, as Wyllam Pekorye can tell + gow. And - te trea muste gowre + c. at + te last, as by Wyllam Pekorye saynyng, and + gat + tat will be to lytyll. And I wolt well we kan not ge xi d. of Cristofyro Hansweyn, so I will be lawl to lend it him of my owne sluer. If I knew verly + gour entent were + tat he xwicd hom I wolde send hym non. There I will doo as me thynketh + xwicd be best spelyd, and + tat, me thynketh, is to send him + te sluer. + Terefore I pray + gow as hastily as + ge may sende me a - gen y mark, + te remanweyte I trow I will get vp on Cristofyro Hansweyn and Lwket, I pray + gow sende me it as hastily as + ge may, fore I will leue my sylfe rythe bare; and I pray + gow sende me a letter how - go well + tat he will be demenyed. Wyrthyn on Tuesday by Seynt Bartheleine, &c. (Christus vos observer.) By Cl[tent Paston]"
Text Variation Explorer: linguistically oriented text visualization tool

Harri Siirtola & Poika Isokoski
Visual Interaction Research Group (VIRG)
TAUCHI Research Center
School of Information Sciences
University of Tampere
SUBPROJECTS: TVE

• How to visually explore corpora?
• Exploration increases chances of finding something unexpected: hypothesis generation

• **Text Variation Explorer**: interactive visualisation tool for exploring & comparing corpora
• Language-independent

• Siirtola et al. (2014)
therin more at large. And as touching your domeonyng in marriage, that for special causes greatly resteth in our mynde and pleasure, we have in likewise shewed unto hym the same by our said instructons, to whom in declaring therof, and of every other thing concerning the premisses, we desire you to reve unto hym ful feith and credence, and with al effect applye and endeavor you to thexeucucion and performyng of the same, as our great trust is in you. Yeven, &c. the xxix. day of Septembre.

<Q. RER 1484 RICHARD3>
<A. RICHARD III>
<P 1,74>

[TO GERALD FITZGERALD]
[A. D. 1484. September.]
To Thele of Kildare.
Right trusty, &c. Certifieng you that as touching the

Word break: +/=/#%;.;":"
Sentence break: !?
Word count: 488309
Sent. count: 19508
Frag. count: 391

Text Variation Explorer
(CEEC Sampler, 1410–1680)
(Build 244)
good. My sister gives you thankes for seending him to her. I pray you remember that I reckon the days you are away; and I hope you are nowe well at Heariford, where it may be, this letter will put you in minde of me, and let you knowe, all your friends heare are well; and all the newes I can send you is, that my Lo. Brooke is nowe at Beaathams Court. My hope is to see you heare this day or next, or to-morowe sene, and I pray God give us a happy meeting, and preserue you safe, which will be the great comfort of

Your most true affectionate wife, Brilliana Harley.

Reply the 30 of Sep. 1625 A.D.

<Q HAR 1625:BHARLEY>
<Q A LADY BRILLIANA HARLEY>
TVE 2.0

• From single document model to multiple data sources
  - Comparisons between texts/subcorpora

• Access to metadata
  - Filtering/comparison by time, author, author’s social class etc.

• Better data export
  - All data should be exportable for further analysis (other STRATAS tools, R, ...)

• Text filtering
  - Exclude markup and commentary

• Richer interaction with line graphs
  - Zoom in on certain kinds of words (e.g. words ending in ‘-ed’ or ‘-d’)
  - Showing types and hapaxes found in each text fragment

• Technical improvements: better UI
  - From Swing to JavaFX