ERRATAS: Charting the philological reliability of modern editions of English historical texts

1. **AIM: QUALITY CONTROL**

Historical corpora of manuscript (hand-written) texts have primarily been compiled from (modern) printed editions: trade-off of quality for quantity, not all linguistic features of manuscript texts retained in corpora. Text features ignored or modernised in editions include e.g. layout, script, spelling.

Therefore, historical corpora are generally considered unsuitable for studying spelling. ERRATAS aims to remedy this by evaluating printed editions of historical letters to try and categorise degrees of editorial intervention, and, thus, determine the level of philological reliability of printed editions.

2. **MATERIAL: Corpus of Early English Correspondence (CEEC)**

Linguistic corpus of English personal letters: 5.1m words in 12,000 letters written by 1,200 writers between 1402-1800. Designed for the sociolinguistic study of morphology and syntax: comes with metadata on social backgrounds of correspondents (age, gender, social rank, education, etc). Compiled from 192 printed ‘original-spelling’ editions (see www.helsinki.fi/varieng/CoRD/corpora/CEEC/).

3. **FOCUS: SPELLING**

What we know of the history of English spelling is based on printed texts. Little is known about English private spelling practices 1400–1800, as no long-term philologically reliable data is available.

Historical English spelling is characterised by variation. In Shakespeare’s English, the letters ⟨u⟩ and ⟨v⟩ both could stand for the sounds /u/ and /v/: e.g. “vp” for up, and “giue” for give was normal and accepted. Usage was not categorical: all writers showed variation: e.g. writing giue but also ever.

Old forms (giue, vp) were common until late 1600s, but English spelling began to be standardised in the 1600s. This first happened in printed texts: this chart shows how the word USEFUL in printed texts (EEBO) was first spelt vsefull; then usefull; and finally useful.

4. **RESEARCH QUESTIONS**

- What do editors say they have done to spellings? What have editors actually done?

In ‘original-spelling editions’, editors commonly claim that texts “have been published precisely as written”. Yet this is usually followed by a list of features where this principle was not followed: usually in punctuation, some spelling variation, abbreviations (commonly expanded), and capitalisation.
Do editorial practices form a hierarchy?

If editorial practices cluster – if feature A is modernised, what about features B and C? – we can form a typology to allow us to evaluate, based on looking only at certain textual features, how faithfully an edition retains manuscript reality. Although we hope to uncover trends that can be applied to all editions of historical texts, within the STRATAS project the primary question is:

Can CEEC be used to study English private spelling practices 1400-1800?

5. METHODS

Manual checking: editorial principles and practices are recorded into a database, using an exhaustive bespoke checklist of textual features known to occur in historical English. Editions are also spot-checked against manuscripts in various archives to establish the veracity of our deductions. The resulting database can then be incorporated into a corpus as metadata, and texts in the corpus given ratings of philological reliability.

6. INITIAL FINDINGS

- Editors are unreliable: We have found no direct correlation of explicit editorial principles and the philological reliability of editions. Yet the age of the edition does not appear to be a factor either.
- The ERRATAS method works, and can be used to identify which editions can be used to study manuscript spelling.
- Example: <u/v>-variation (giue vs give, vp vs up). In CEEC (all 17C letters), the new form (give, up) is dominant: the old form occurs 13% of the time. After identifying the best editions used in CEEC, in the resulting corpus (a subset of the previous), the old form occurs 31% of the time.

With editorial interference thus reduced, the results of corpus searches are closer to manuscript reality (actual private spelling practices).