

A large-scale analysis of the use of Scotticisms

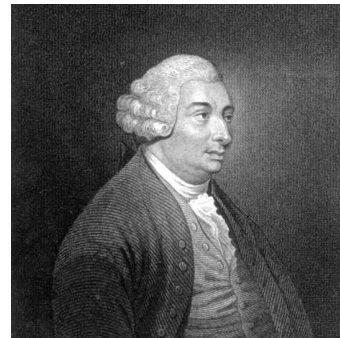
in the eighteenth century



Aatu Liimatta, Iiro Tiihonen, Lidia Pivovarova, Mikko Tolonen, Tanja Säily

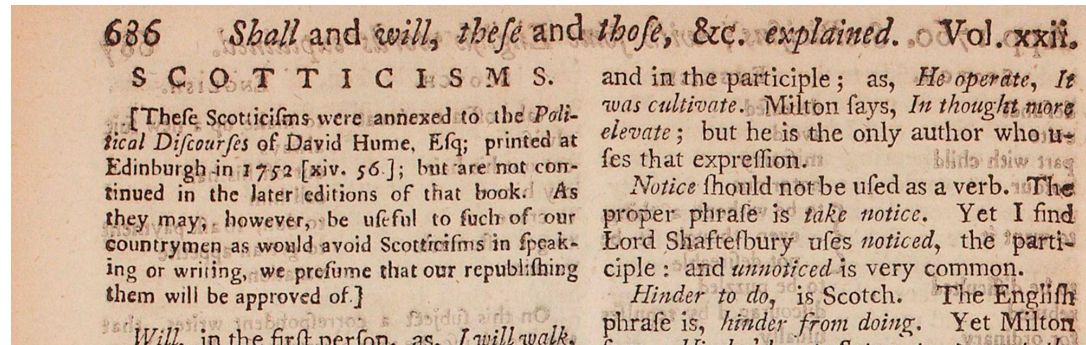
Introduction

- C18: **standardization** of the English language in Scotland became a central topic in public discourse
- Standardization promoted by groups such as the **Select Society of Edinburgh** (fl. 1754–1764; Emerson 1973)
- One of the members of the society was the Scottish Enlightenment philosopher **David Hume** (1711–1776)



Hume's list of Scotticisms

- **Glossary** appended to Hume's *Political Discourses* (1752)
- Listed “Scotch” words and phrases that should be avoided and the preferred “English” equivalents
 - e.g. “Scotch” *to be diffculted* vs. “English” *to be puzzled*
- **Reprinted** a number of times in the eighteenth century
→ seen as important, influenced at least some C18 writers (Cruickshank 2013)



Research questions

- How common was the use of Hume's "Scotticisms" in eighteenth-century British books?
- Who used the Scotticisms and where?
 - Author, author's regional background, publisher, place of publication, genre...
- Was there change over time, did efforts like Hume's make a difference?

Material: *Eighteenth Century Collections Online* (ECCO)

- The 200,000 volumes of ECCO cover half of the surviving printed record of 18th-century Britain
- Challenges:
 - Texts are not clean (OCR noise)
 - Biases in the composition of the collection
- Metadata from the *English Short Title Catalogue* (ESTC)
 - Harmonized by the Helsinki Computational History Group (COMHIS)
 - Dataset restricted to 1st editions of book-length works (altogether **18,252 books**)
 - Additional author + genre metadata produced at COMHIS

Methods: Identification of Scotticisms

— — —

- Hume's list retrieved from [LEME](#)
- Operationalization: textual lookups using **regular expressions**
 - Noun singular/plural, verb conjugation, different pronouns, different determiners, common variant spellings, compound words together/hyphen/space, typical OCR errors (e.g. s/f)...
 - E.g. *cause him do it*
-> `cau(s|f)(es|ed|e|'d|d|ing) (me|you|him|her|it|us|them) do`
 - But: does not find other verbs than *do*, which lowers recall
- Many items on the list would need structural/part-of-speech analysis and/or semantic **disambiguation** for correct identification
 - E.g. *to want it* 'need, lack', not 'desire'; *chimney* 'grate', not 'smokestack'
- We were able to analyse 67 Scotticisms and 58 "Englishisms" out of 106 pairs on Hume's list

Items included (1 / 2)

— — —

Scotch

conform to
friends and acquaintances
Maltreat
advert to
proven
improven
approven
pled
incarcerate
fresh weather
in the long run
notwithstanding of that
a question if
with child to a man
simply impossible
in time coming
nothing else
severals
anent
allenary
alongst
as I shall answer

English

conformable to
friends and acquaintance
abuse
attend to

pleaded
imprison
open weather
at long run
notwithstanding that
a question whether
with child by a man
absolutely impossible
in time to come
no other thing
several
with regard to
solely
along
I protest or declare

Scotch

cause him do it
marry upon
effectuate
a wright
defunct
evite
part with child
notour
to be diffculted
think shame
in favours of
dubiety
compete
remeed
bankier
adduce a proof
superplus
forfaulture
in no event
common soldiers
debitor
exeeded

English

cause him to do it
marry to

a carpenter
deceast
avoid

notorious
to be puzzled
asham'd
in favour of
doubtfulness
enter into competition

banker
produce a proof
surplus
forfeiture
in no case
private men
debtor
exempted

Items included (2 / 2)

— — —

Scotch

yesternight
big coat
tenible argument
amissing
extinguish an obligation
depone
to inquire at a man
angry at
to send an errand
to furnish goods to him
to open up
Thucydide
Herodot
Sueton
butter and bread
pepper and vinegar
paper, pen and ink
as ever I saw
come in to the fire
alwise
cut out his hair
to get a stomach
vacance

English

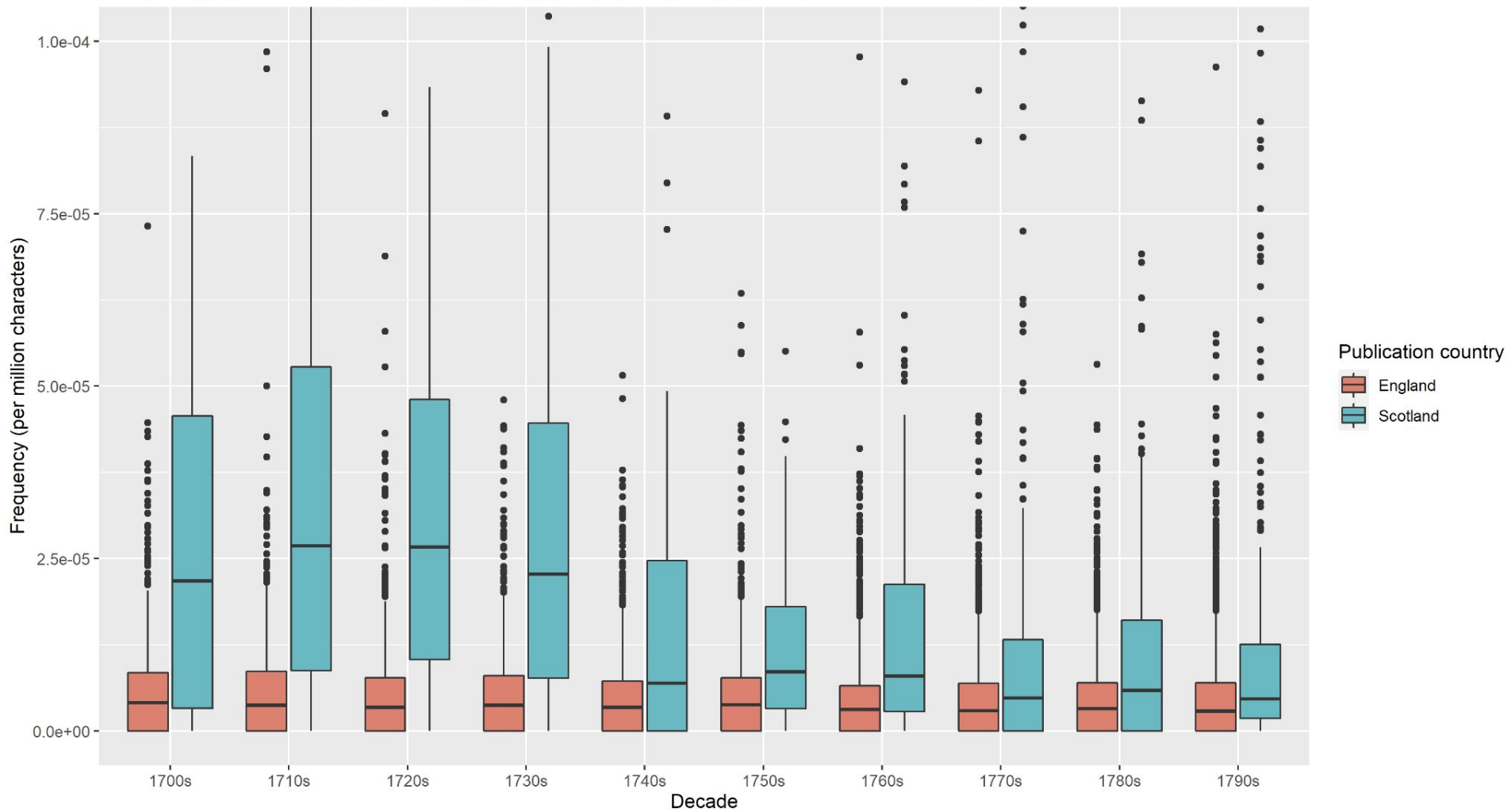
last night
great coat
good argument
cancel an obligation
to inquire of a man
angry with
to send off an errand
to furnish him with goods
Thucydides
Herodotus
Suetonius
bread and butter
vinegar and pepper
pen, ink and paper
as I ever saw
come near the fire
always
cut off his hair
to get an appetite
vacation

Methods of analysis

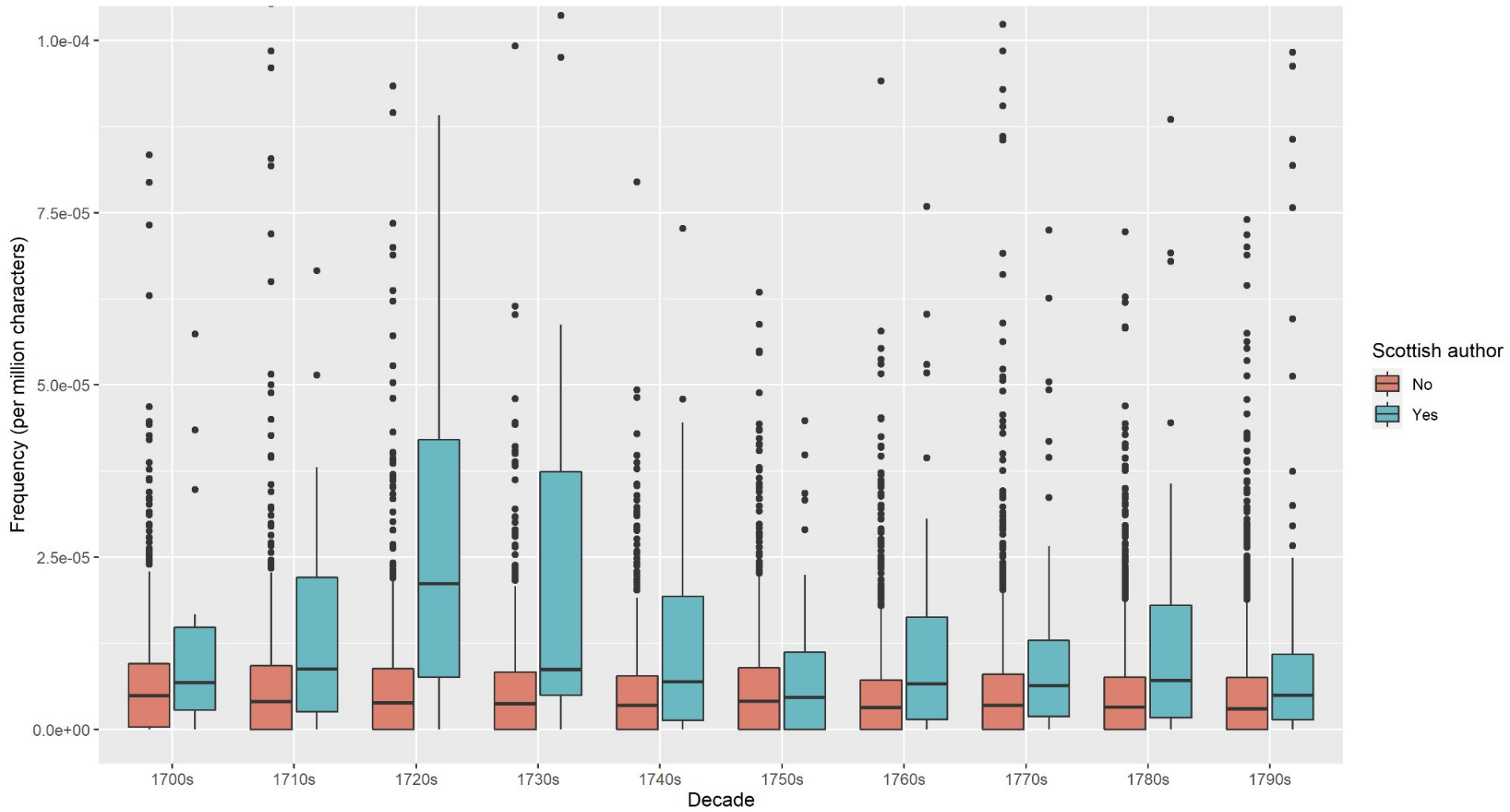
- First iteration: Data **visualization** (e.g. boxplots)
 - Used to study the relationship between Scotticisms and (individual) other variables of interest (e.g. time, nationality, OCR noise)
 - Pros: easy to interpret
 - Cons: Do not allow for more complex analyses
- Second iteration: penalized multivariate **regression**
 - Allows for parallel consideration of hundreds of variables possibly related to the appearance of Scotticisms, while being robust against false signals (LASSO penalization)
 - Pros and cons: interpreting the fitted model produces a more complex argument about Scotticisms

Analysis

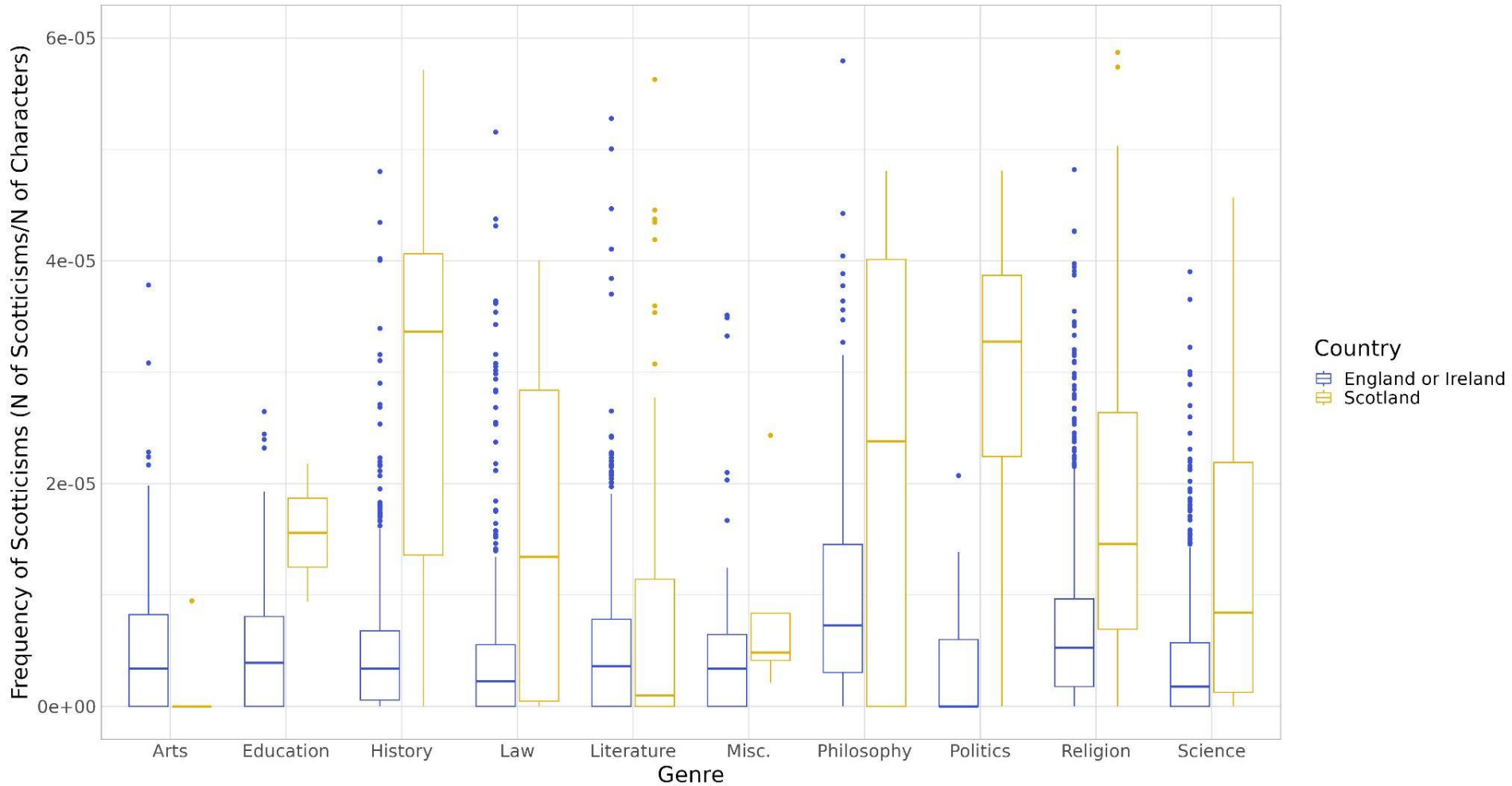
Frequency of scotticisms in books by country of publication per decade of the 18th century



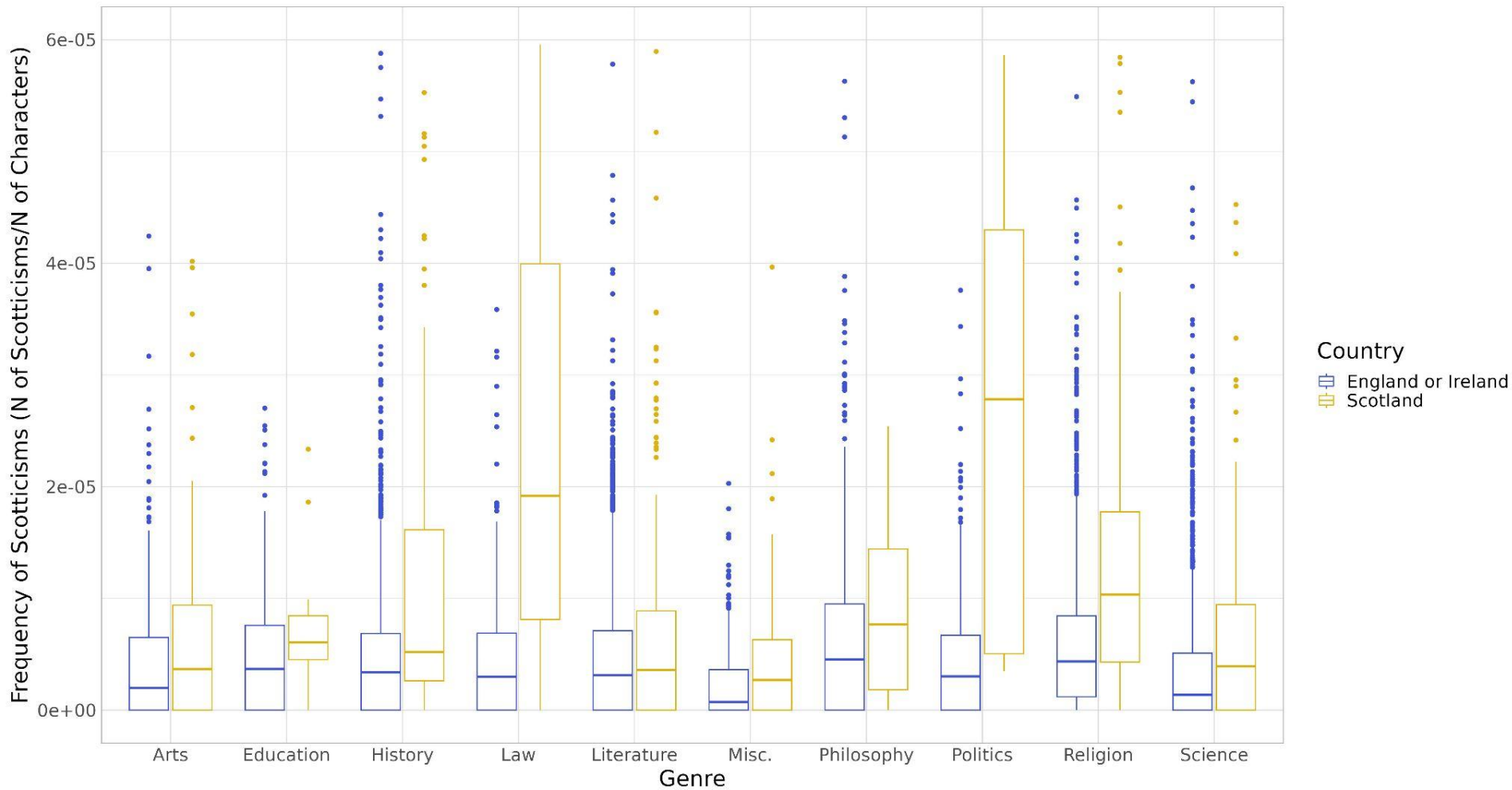
Frequency of scotticisms in books by author per decade of the 18th century



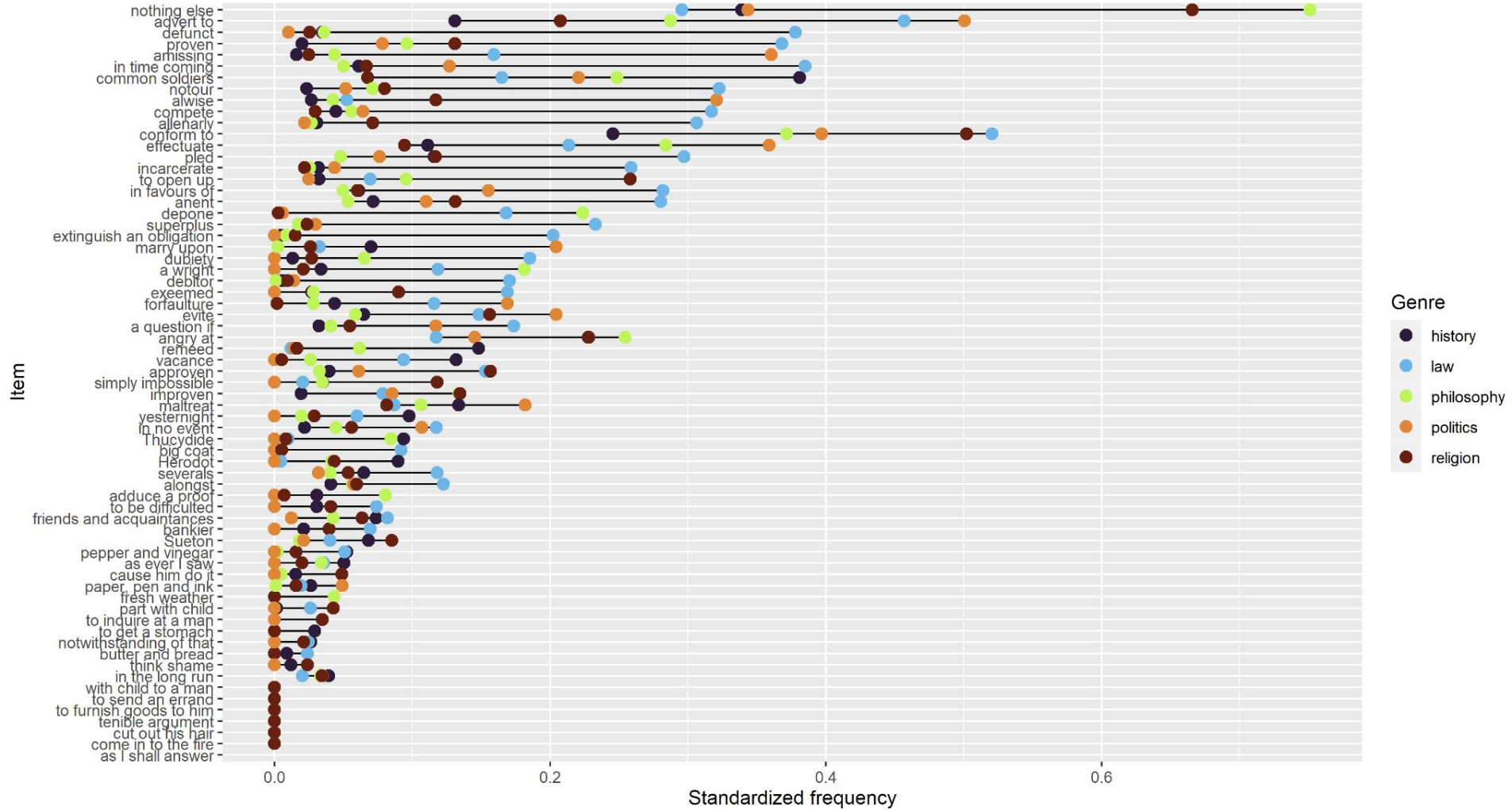
Scotticisms in Books by Genre and Country of Publication. First Half of the Eighteenth Century



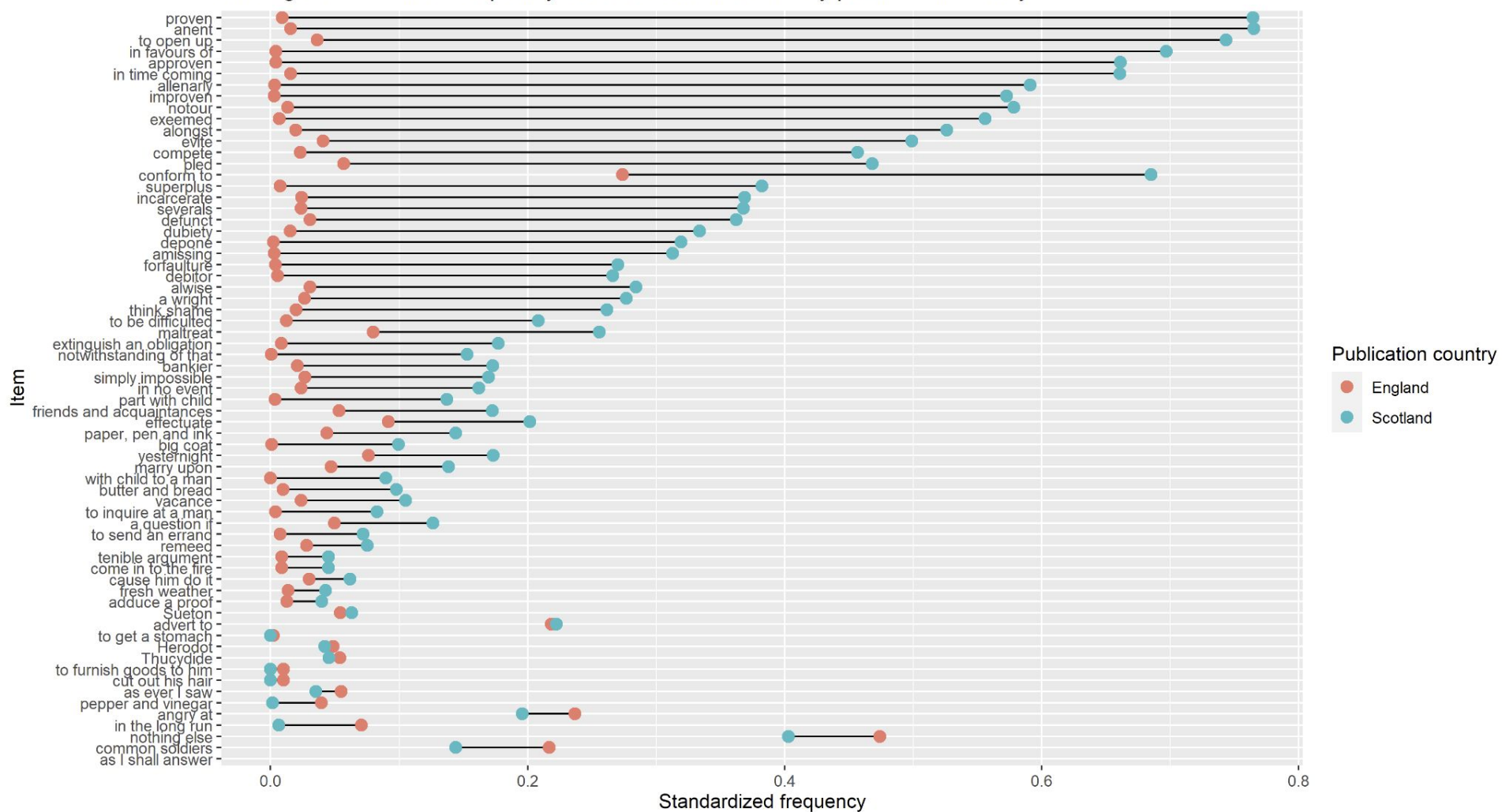
Scotticisms in Books by Genre and Country of Publication. Second Half of the Eighteenth Century



Average standardized frequency of Scottisms in books of selected genres



Average standardized frequency of Scotticisms in books by publication country



Top 6 books using Scotticisms

	Author	Year
State of the conjoined processes, betwixt Sir Alexander MacDonald, of MacDonald, baronet, and others, and Normand MacLeod, of MacLeod, Esq; and others	NA	1766
State of the process of division of the common muirs or commonries of Reddingrig and Whitesiderig. Lying in the Shire of Stirling. 1763	NA	1763
State of the process of reduction, declarator and damages: the Governor and Company of Undertakers for raising the Thames Water in York-Buildings; against Alexander Mackenzie, writer to the Signet	NA	1791
The principles of book-keeping explain'd	Macghie, Alexander	1718
Essays upon several subjects in law	Kames, Henry Home, Lord	1732
State of the processes depending against Mr. John Simson	Simson, John	1728

Example

because tho' the first two Ad- judgers were preferable as to their 10000 Merks, and in so far as the Annualrenter draw from them, the Inhibiter was not prejudged, yet as to the **Superplus** of 2000 Merks, the Competition falling betwixt the Inhibiter and a posterior Annualren- ter, the Inhibiter is intirely preferred.

Essays upon several subjects in law (Henry Home, Lord Kames, 1732)

Penalized regression

- Data: 8500 observations
 - 7200 for model fitting, 1300 for evaluation
- Predictors (total = 651)
 - OCR quality (continuous)
 - Publication country x quarter century (dummy)
 - Nationality of author x birth century (dummy)
 - Publication country x genre x half-century (dummy)
 - Presence of a given publisher with more than 50 editions (dummy)
- Model details
 - Negative binomial model (to handle overdispersion, $\phi=2$)
 - Target = rate of Scotticisms in a volume
 - Offset = number of characters in a volume

Penalized regression – results (1 / 2)

- The model fitting process resulted in 16 variables with non-zero effect size
- Associated with **higher rate of Scotticisms:**
 - Scotland as a publication place, effect decreases through the century
 - Law texts
 - Scottish authors born in the 17th century
 - R. Dodsley (publisher)

Penalized regression – results (2 / 2)

- Associated with **lower rate of Scotticisms**:
 - Low OCR quality
 - Late 18th-century scientific books published in England
 - Scottish authors born in the 18th century
- Reliability of findings
 - Penalization should produce robust results
 - Analysis of residuals suggests that important variables might be omitted, and penalty can also produce “false negatives”

Did Hume use Scotticisms?

-
- Sinclair (1782): *Observations on the Scottish Dialect*
 - Many examples from Hume, e.g. *you was* for *you were* (“This is an impropriety which even Mr. Hume was guilty of”, p.58)
 - Hume’s **publications**:
 - *you was* 0
 - *you were* 15
 - Hume’s **letters** (Past Masters collection):
 - *you was* **41**
 - *you were* 17

→ Used (some) Scotticisms in private writing!



Discussion

Summary of results

- Some of the items listed by Hume are quite common
 - E.g. *conform to*: 84,938 instances
 - Some more Scottish than others, many quite infrequent
- Used more by Scottish authors, books published in Scotland
 - Many especially common in legal texts
- Decrease over time
 - Heaviest users born in the 17th century
 - Change at its fastest around mid-18th century?

→ Efforts like Hume's accelerated ongoing standardization?

Why did Hume focus on legal Scotticisms?

- Legal Scotticisms was a natural context for Scotticisms because of the localized nature of legal language
- Hume came from a family of lawyers
 - Was supposed to become a lawyer himself but ill suited
- All his patrons were lawyers
 - E.g. Henry Home, Lord Kames
- Army position, diplomat, 1752 librarian at the Faculty of Advocates
 - Editing/correcting close to his heart

Conclusion

- Hume and the Select Society of Edinburgh contributed to an ongoing process of standardization
- Hume's glossary was not the only one of its kind produced at the time
 - Ideology of standardization
- Future work
 - Compare actual usage with other lists of Scotticisms, e.g. Sir John Sinclair, James Beattie
 - Add a purely data-driven approach?

References

- Cruickshank, Janet. 2013. The role of communities of practice in the emergence of Scottish Standard English. In Joanna Kopaczyk & Andreas H. Jucker (eds.), *Communities of Practice in the History of English*. John Benjamins, 19–45.
- Emerson, Roger L. 1973. The social composition of Enlightened Scotland – The Select Society of Edinburgh 1754–1764. *Studies on Voltaire and the Eighteenth Century* 114: 291–329.
- Hume, David. 1752. *Political Discourses*. Edinburgh: Printed by R. Fleming, for A. Kincaid and A. Donaldson.