# Measuring the distribution of Hume's Scotticisms in the ECCO collection

**Iiro Tiihonen**     **Aatu Liimatta**     **Lidia Pivovarova**     **Tanja Säily**     **Mikko Tolonen**
University of Helsinki
`first.last@helsinki.fi`

## Abstract

This short paper studies the distribution of Scotticisms from a list compiled by David Hume in a large collection of 18th century publications. We use regular expression search to find the items on the list in the ECCO collection, and then apply regression analysis to test whether the distribution of Scotticisms in works first published in Scotland is significantly different from the distribution of Scotticisms in works first published in England. We further refine our analysis to trace the influence of variables such as publication date, genre and author's country of origin.

## 1 Introduction

The 18th century was a period of standardization efforts for the English language, which is reflected in many contemporary texts, including those of philosopher David Hume. One of the widely discussed topics was "Scotticisms", i.e. non-standard words and expressions of Scottish origin (Dossena, 2005, 2012).

Hume, himself a Scot, was concerned with the purity of language and paid a great deal of attention to the language of his writings. He was also involved in editorial work and assisted other writers in polishing their texts. The matter of Scotticisms is mentioned several times in his letters to other writers while discussing their work.

A list of "Scotticisms" was published as an appendix to Hume's *Political Discources* in 1752. The list was reprinted several times during the 18th century and mentioned in works by various authors. However, neither the 1752 edition nor any other work by Hume explains how the list was compiled, to what extent it is representative of the language use in 18th century and what the impact of this work was on further standardization efforts. These characteristics make it an interesting cultural artifact to study quantitatively. It might capture regional, generational, genre-related and other variation and eventual standardization of English, but it is also most likely affected by the experiences and interests of its famous author. For example, previous research has found out that many of Hume's Scotticisms occurred in legal contexts (Cruickshank, 2013), which might be explained by the fact that Hume as a former student of law had been exposed to them. Better understanding of how the use of Hume's Scotticisms varied in eighteenth-century texts can provide insights both to the standardization and variation of historical English, and to the origins, nature and limitations of the list itself.

We perform a large-scale corpus study to determine: (i) How Scottish were Hume's "Scotticisms"? (ii) Who used them and where? (iii) Was there change over time, did efforts like Hume's make a difference? To that end, we search for the Scotticisms on the list in a corpus of 18th century publications from England and Scotland and study their distribution across location, genre and time.

There was a previous attempt to search the items on Hume's list in a limited correspondence corpus (Cruickshank, 2013). However, as far as we are aware, this is the first attempt to analyse the actual usage of these items in a massive database of public discourse.

## 2 Data

The main textual dataset for our analysis is *Eighteenth Century Collections Online* (ECCO) (Tolonen et al., 2021). The texts in ECCO have been made into a machine-readable form using optical character recognition (OCR) technology. However, the ECCO dataset has significant problems with the quality of the OCR texts. Since the dataset itself consists of bitonal microfilm scans of varying quality, the OCR process has often not been able to reproduce the text very well or at all. Thus,

the textual data is often very messy, which can cause problems with analyses of the data (Hill and Hengchen, 2019). Nevertheless, a lot of the ECCO data is of fair quality, and as such the dataset can be used for various kinds of analyses.[1]

Comprising about 200,000 volumes of 18th-century printed works, the ECCO dataset covers roughly half of the surviving printed works from the period. The dataset is not a balanced linguistic corpus, but more an incidental collection of various texts. For instance, ECCO contains many more documents from the later periods of the century, with earlier periods underrepresented in relative terms. Furthermore, different document lengths can dominate different periods. The dataset has not been balanced with respect to other variables, such as genre or register. Moreover, ECCO includes multiple editions of many works, which can confuse quantitative analyses due to the same or very similar content being included multiple times at different time periods.

Our metadata for the texts originates from the *English Short Title Catalogue* (ESTC)[2], harmonized and augmented by the Helsinki Computational History Group (Lahti et al., 2019). Metadata were used to implement two data-filtering steps intended to control for the complexity of the analysis and data quality issues:

- pamphlets and editions other than first were excluded;
- editions must have been published between 1700 and 1799;

And additional steps for the regression analysis:

- the median OCR quality of the pages in the edition had to be at least 80%;
- only editions by authors who were born between 1630–1780 and had at least 5 editions in the ESTC were used;
- the word "Scotticism(s)" must not be mentioned in the text of the edition, as it indicates a linguistic work discussing Scotticisms.

---

## 3 Method

### 3.1 Scotticism Extraction

We retrieved Hume's list of Scotticisms in plaintext format from the Lexicons of Early Modern English (LEME).[3] To identify Hume's Scotticisms in the texts in ECCO, we operationalized the Scotticisms on Hume's list as regular expressions, which enables us to look for various alternative versions of the Scotticism in question. For example, in the case of *cause him do it*—which Hume says should be *cause him to do it* instead—there are two major varying components. First, we included different conjugations of the verb *to cause*, such as *causes*, *caused*, or *causing*. Second, we included other object pronouns, such as *her*, *me*, and *us*, in addition to the pronoun *him*. Furthermore, we accounted for varying spelling conventions by including some common variant spellings, such as *caus'd* and *causd* for *caused*. Finally, cognizant of the varying OCR quality of the data, we included provisions for some common OCR errors, most importantly for the tendency of the long *s* being erroneously recognized as an *f*. After these considerations, and judging that the final *it* is not central to the structure, the regular expression lookup for *cause him do it* was the following:

```
cau(s|f)(es|ed|e|'d|d|ing) (me|you|him|her|it|us|them) do
```

For other items, where necessary, we also considered other potential varying components such as plural and singular forms for nouns, different determiners, and multi-word expressions which could be written separately, together, or hyphenated.

While these kinds of lookups using regular expressions do increase the recall, they do not find all possible instances of the relevant construction. For example, in the case of *cause him do it*, the object *him* could also be any noun phrase, and *do* could be any verb. However, identification of such constructions would require part-of-speech tagging and structural parsing of the ECCO corpus. This procedure by itself may introduce additional errors and would require estimation of tagging and parsing performance of existing tools for historical OCRed data, which is out of the scope of this paper. Similarly, some items on Hume's list of Scotticisms would only be possible to identify using parsed data, and therefore had to be excluded from our analysis.

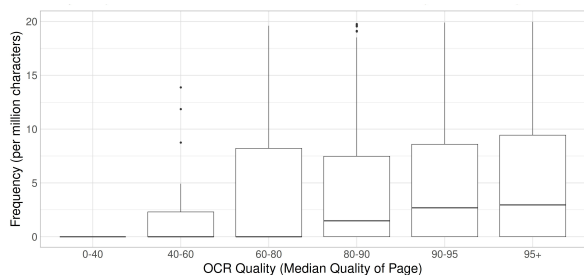For some items, Hume speaks against the use of

---

Figure 1: The frequency of Scotticisms in relation to OCR quality of the documents (specified by the collection distributor, i.e. Gale). 364 observations omitted from the visualization due to failing outside visualization range.

a more widely used word in a specific sense. For instance, while *chimney* is widely used in English to refer to a smokestack, Hume rejects its Scottish use to refer to the grate under a fire in a fireplace. As semantic disambiguation would be required to identify the word being in the Scottish sense while ignoring its use in other senses, such items were excluded from our analysis. In total, we were able to identify at least to some degree 67 Scotticisms out of the 106 items on Hume's list. The list of extracted Scotticisms and corresponding regular expressions is presented in Appendix B.

Since the OCR process often leads to changes in spelling and to other forms of textual 'noise', it is possible that we fail to find many occurrences of Scotticisms. Our analysis of the relation between OCR quality and observed Scotticisms in a document is illustrated in Figure 1. Low level of OCR quality is associated with a lower frequency of Scotticisms. Hence we hypothesize that the real differences in the use of Scotticisms in 18th-century Britain might have been even higher than what we have measured. The figure tentatively suggests that median OCR page quality of below 60 percent leads to a drastic drop in observed Scotticisms, and from there on the number of Scotticisms in a typical edition increases from 0 for the 60–80 percent OCR quality bracket to around 3 per million characters for the 90–95 percent bracket.

## 3.2 Regression Analysis

We analysed the relation of Hume's Scotticisms to other characteristics of publications with *multivariate regression analysis*. This was done to verify which factors related to Scotticisms would hold when other potentially related variables were also controlled for.

Univariate analyses would suggest that a higher average rate of Scotticisms in a group of works tends to correlate with a much higher variation in their number, a phenomenon not captured by many standard regression models. We also suspected that there might be more zero-Scotticism instances than standard statistical models assume. And, by making use of the publisher and author information, our data became relatively high-dimensional, creating the risk of overfitting. We need to assess uncertainty in the model while making it sufficiently complex to incorporate all of these properties. Our solution was to implement a Bayesian zero-inflated Negative Binomial regression model with the R-package BRMS (Bürkner, 2017).[4] Four chains of 3,000 iterations (half of these samples were warm-ups for each chain) were run on STAN (Stan Development Team, 2023) via BRMS to obtain an approximation of the posterior distribution.

Negative Binomial distribution allows us to model the significant increase of variance as a function of the mean (heteroscedasticity), zero-inflation addresses the problem of possible over-representation of zeroes, and by setting a horseshoe-prior to the fixed effects, we can guide the model to by default, e.g. in absence of significant evidence in the data, be in favour of considering none of the effects to matter, which most likely is the case. Similarly, it is easy to group variables together in the Bayesian framework, making them share information in the model fitting process. And, as the posterior distribution is simulated in our Bayesian approach, the model fitting process also produces estimates about the reliability of our findings (effect sizes).

In total, 8,948 editions were used for the regression analysis. Of these, 8,000 were used to fit the model, and the rest were reserved for model evaluation. 812 of the observations reserved for evaluation had an author that was also present in the data used to fit the model, making predictions—and hence the kind of evaluations conducted—with the Bayesian model possible.

We modeled the relation of the number of Scotticisms in an edition to its metadata features. The length of the book in characters was used as an offset to normalize for the length of the book. We included two types of variables in the model. Population-level variables affected all observations with an equal impact. Hence, population-level vari-

---

[4] https://CRAN.R-project.org/package=brms

ables can be associated with fixed effects of regular linear models, as their effect (e.g. that of the OCR quality) only varies by the value of the variable itself, not by some other variable. The impact of the group-level effects varied by a grouping variable. That is to say that they (e.g. the time of publication of an edition) affected the target variable (Scotticisms) differently based on the value of some other variable(s) (e.g. place and genre of the publication) called the grouping variable(s). In our case, the group-level effects are a constant grouped by the author and a quarter-century specific effect grouped by the publication place and genre. In other words, we model the Scotticisms as being affected by a constant unique to each author and progress of time that was conditioned by the combination of genre and time: e.g., Scottish law having its own temporal trajectory.

## 4 Results

We evaluated the reliability of our model by comparing how well it predicted the frequency of Scotticisms in a test set compared to a null model that only considers the document length in characters to predict a number of Scotticisms.[5] The comparison of the predictions of the model to the real number of Scotticisms in the test data is shown in Figure 2. There is a clear difference in favour of the full model, as it is able to more accurately detect those instances in which the number of Scotticisms in a document is very high. Our model is able to capture such variation in the number of Scotticisms that generalizes beyond the training set.

The results of the regression analysis are presented in Table 1, as well as in Tables 2 and 3 in Appendix A. These tables report the approximated marginal posterior-distributions of the model. They communicate estimates of how well the whole range of possible parametrizations of the model explain the data and align with priors using posterior-likelihood as the measure. The probabilities related to any given parameter values express how big a proportion of the posterior-likelihood (compared to all possible parametrizations) is concentrated on those parameter values. For example, if some parameter's (e.g. the effect of OCR quality) marginal posterior distribution at 2.5th quantile is 0.1 and at 97.5th quantile 0.2, we can say that 95 percent of the posterior-likelihood (or posterior-probability) is
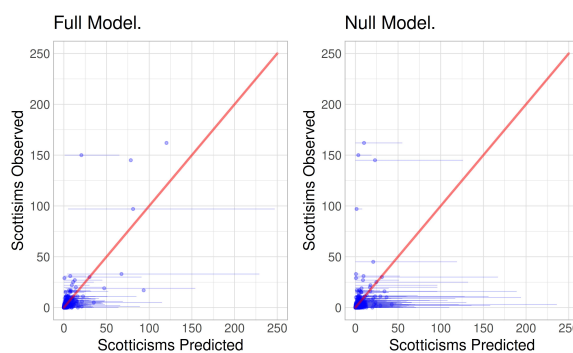


Figure 2: The real and predicted (mean) number of Scotticisms as predicted by the full and null models. When possible within the limits of the x-axis, the range of predictions from 2.5th to 97.5th quantile is illustrated with blue horisontal lines. 3 observations omitted from the full and 1 from the null model due to zooming of the image.

concentrated on models that propose that the parameter is positive and between 0.1 and 0.2. The tables allow us to identify those parameters for which most (95 percent) of the posterior-probability is supporting either a positive or a negative effect on Scotticisms.

Several findings of the preceding univariate visualizations are supported by the posterior distribution of the fitted model:[6] good OCR quality, the author being a Scot born in the 17th century and especially Scottish legal publications are associated with an increased rate of Scotticisms. Additionally, several authors are associated with an increased rate of Scotticisms. The most consistent factors associated with a lower rate of Scotticisms are those depicting the difference between the first and 3rd/4th quarters of the 18th century for non-legal Scottish publications.

Hence, the regression model offers support for three major claims:

1. For genres other than law, the rate of Hume's Scotticisms decreased in Scottish publications during the later 18th century.
2. Scottish legal publications used Hume's Scotticisms at a much higher rate than other types of publications.
3. Author-to-author variation in the use of Scotticisms was significant, and the 17th-century Scottish generations used them more.

Based on these claims, we can draw two higher-

---

[5]The number of iterations and chains for the null model was the same as for the full model.

[6]Here, we only discuss such effects for which the tails (2.5th, 97.5th quantiles) of the posterior are both either positive or negative. That is, we focus on effects that the model sees as very likely positive or very likely negative.

| Variable | 2.5th q. | Median | 97.5th q. |
|---|---|---|---|
| Intercept | 0.68 | 1.9 | 3.7 |
| OCR quality | 0.094 | 0.13 | 0.16 |
| Scot a. 17th c. | 0.007 | 0.53 | 0.85 |
| Other a. 18th c. | -0.6 | -0.24 | -0.0001 |

Table 1: Posterior distributions of population-level effects. Includes only those effects for which the sign of the 2.5th and 97th quantiles was the same (i.e., the effect is highly likely either positive or negative).



Figure 3: Frequency of Scotticisms per decade in works published in England (red) and Scotland (blue).

level conclusions. First, the overall process of standardization was best resisted by the often formulaic legal genre of Scottish texts, which did not show robust signs of decreased use of Scotticisms even at the end of the century. In previous research it has been suggested that Scottish legal language got replicated throughout society because law had a daily impact on the lives of most Scots (Cruickshank, 2013, 39). Our results imply the opposite: while legal texts remained Scotticism-heavy, literary culture as a whole was heavily impacted by standardization.

The other major conclusion is that individuals differed in their use of Scotticisms to a remarkable degree. Even after accounting for other factors that were related to variation in the use of Scotticisms (among them the overall difference between authors born in the seventeenth vs. the eighteenth century), some authors used them orders of magnitude more than others. While the analysis of the use of Scotticisms by specific individuals is beyond the scope of this paper, this differentiation as a general phenomenon has historical implications. Hume and fellow-minded advocates of standardization were focusing on removing characteristics of English that did divide authors.

It is worth noting that most items from Hume's list were more frequently used in their standard form, even in Scottish writings. For example, the Scotticism *alwise* was found 173 times in our corpus of works published in Scotland, while the standard form *always* was used 44,972 times in this corpus. Thus, *alwise* could serve as a strong predictor of Scottish work, but even in Scottish works the standard form was much more frequent. Therefore, the standardization was not a complete transformation of Scottish English but an attempt to eliminate what was seen as regionally specific language mistakes.

Taken as a whole, the changes in the use of Scotticisms were remarkable. Figure 3 shows that Scotticisms are indeed prevalent in books published in
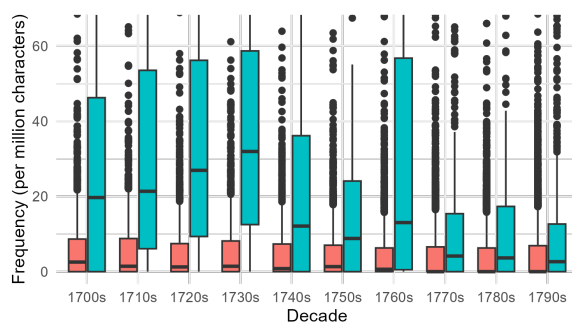
Scotland and that even there their number gradually decreases during the 18th century, with the 1760s being an outlier. We checked those documents from 1760s that have the biggest number of Scotticisms and found out that all of them were legal documents. The peak in the number of Scottish legal documents published during that time could have some historical explanation or could be a mere corpus artefact. We leave this for further investigation.

## 5 Conclusion and Further Work

Our analysis confirms that David Hume was familiar with the peculiarities of language use in his time. The overall trend towards standardization was resisted by Scottish legal texts and modified by significant variation between authors. The specific contribution of David Hume to this process is a matter for further research.[7]

Further work would include both refining the methods and taking into account a broader set of materials. The former line of research may include structural analysis of the data—including part of speech tagging and parsing—and efforts to find "Scotticisms" in a data-driven way without any predefined list. The latter would involve studying Scotticisms compiled by other 18th century writers, and studying other types of data outside ECCO, such as newspapers and personal correspondence. We also plan to apply methods based on contextualized embeddings for semantic disambiguation and post-OCR correction of items from the list.

## Acknowledgments

---

[7] A paper discussing these results from a humanities perspective is currently under review (Tolonen et al., 2024).

# References

Paul-Christian Bürkner. 2017. BRMS: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

Janet Cruickshank. 2013. The role of communities of practice in the emergence of Scottish Standard English. In Joanna Kopaczyk and Andreas H. Jucker, editors, *Communities of Practice in the History of English*, pages 19–45. John Benjamins, Amsterdam.

Marina Dossena. 2005. *Scotticisms in Grammar and Vocabulary: 'Like Runes Upon a Standin' Stane'?* John Donald, Edinburgh.

Marina Dossena. 2012. Scottishness and the book trade: Print and Scotticisms. In Warren McDougall and Stephen W. Brown, editors, *The Edinburgh History of the Book in Scotland*, volume 2, pages 545–550. Edinburgh University Press, Edinburgh.

Mark Hill and Simon Hengchen. 2019. Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. *Digital Scholarship in the Humanities*, 34(4):825–843.

Leo Lahti, Jani Marjanen, Hege Roivainen, and Mikko Tolonen. 2019. Bibliographic data science and the history of the book (c. 1500–1800). *Cataloging & Classification Quarterly*, 57(1):5–23.

Stan Development Team. 2023. *Stan Modeling Language User's Guide and Reference Manual*. Version 2.31.

Mikko Tolonen, Aatu Liimatta, Lidia Pivovarova, Iiro Tiihonen, and Tanja Säily. 2024. Hume's list of Scotticisms in eighteenth-century British context.

Mikko Tolonen, Eetu Mäkelä, Ali Ijaz, and Leo Lahti. 2021. Corpus linguistics and Eighteenth Century Collections Online (ECCO). *Research in Corpus Linguistics*, 9(1):19–34.

# A Regression Analysis Results

| Variable | 2.5th q. | Median | 97.5th q. |
|---|---|---|---|
| Law England (intercept) | -2.04 | -0.24 | 1.05 |
| Law Scotland (intercept) | 0.99 | 2.94 | 4.80 |
| Other England (intercept) | -2.14 | -0.39 | 0.82 |
| Other Scotland (intercept) | -0.30 | 1.51 | 2.76 |
| Religion and Philosophy England (intercept) | -1.93 | -0.17 | 1.04 |
| Religion and Philosophy Scotland (intercept) | -0.41 | 1.39 | 2.65 |
| Law England 2nd quarter | -1.61 | -0.83 | -0.02 |
| Law Scotland 2nd quarter | -1.77 | -0.44 | 0.53 |
| Other England 2nd quarter | -0.24 | -0.07 | 0.09 |
| Other Scotland 2nd quarter | -0.53 | -0.02 | 0.48 |
| Religion and Philosophy England 2nd quarter | -0.30 | -0.13 | 0.02 |
| Religion and Philosophy Scotland 2nd quarter | -0.81 | -0.36 | 0.04 |
| Law England 3rd quarter | -1.03 | -0.29 | 0.38 |
| Law Scotland 3rd quarter | -0.90 | 0.22 | 1.20 |
| Other England 3rd quarter | -0.39 | -0.20 | -0.03 |
| Other Scotland 3rd quarter | -1.78 | -1.30 | -0.83 |
| Religion and Philosophy England 3rd quarter | -0.52 | -0.33 | -0.15 |
| Religion and Philosophy Scotland 3rd quarter | -1.30 | -0.88 | -0.46 |
| Law England 4th quarter | -0.75 | -0.13 | 0.46 |
| Law Scotland 4th quarter | -2.09 | -0.85 | 0.21 |
| Other England 4th quarter | -0.30 | -0.12 | 0.07 |
| Other Scotland 4th quarter | -1.95 | -1.44 | -0.97 |
| Religion and Philosophy England 4th quarter | -0.44 | -0.24 | -0.04 |
| Religion and Philosophy Scotland 4th quarter | -1.56 | -1.13 | -0.71 |

Table 2: The effect of different combinations of genre and place on the rate of Scotticisms (intercept) and how the effect changes in the 2nd, 3rd, and 4th quarters of the 18th century.

| Variable | 2.5th q. | Median | 97.5th q. |
|---|---|---|---|
| Hamilton, James Hamilton, Duke of, 1724-1758. | 2.01 | 2.70 | 3.45 |
| Palmer, Thomas Fyshe, 1747-1802. | 1.89 | 2.53 | 3.25 |
| Law, William, 1686-1761. | 1.50 | 2.08 | 2.72 |
| Cullen, Francis Grant, Lord, 1658-1726. | 1.40 | 2.05 | 2.80 |
| Cardonnel, Adam de, -1820. | 1.35 | 2.01 | 2.74 |
| Simson, John, 1668?-1740. | 1.27 | 1.97 | 2.76 |
| Mackenzie, Alexander, 1735-1805. | 1.30 | 1.84 | 2.41 |
| Kirkby, John, 1705-1754. | 0.99 | 1.78 | 2.63 |
| Aberdeen, George Gordon, Earl of, 1722-1801. | 1.08 | 1.77 | 2.52 |
| Blackstone, William, Sir, 1723-1780. | 0.89 | 1.66 | 2.50 |
| Mitford, William, 1744-1827. | 0.87 | 1.56 | 2.36 |
| Maittaire, Michael, 1667-1747. | 0.95 | 1.56 | 2.26 |
| Roscoe, William, 1753-1831. | 0.90 | 1.52 | 2.19 |
| Howie, John, 1735-1793. | 0.86 | 1.51 | 2.26 |
| Roxburghe, John Ker, Duke of, 1740-1804. | 0.71 | 1.51 | 2.38 |
| Cockman, Thomas, 1675?-1745. | 0.85 | 1.47 | 2.17 |
| Badeslade, Thomas. | 0.74 | 1.47 | 2.30 |
| Okely, Francis, approximately 1719-1794. | 0.66 | 1.47 | 2.27 |
| Cockburn, William, Sir, 1662-1751. | 0.72 | 1.43 | 2.20 |
| Baretti, Giuseppe, 1719-1789. | 0.77 | 1.41 | 2.13 |
| Eachard, John, 1636?-1697. | 0.71 | 1.39 | 2.12 |
| Moray, James Stuart, Earl of, 1708-1767. | 0.57 | 1.38 | 2.34 |
| Gib, Adam, 1714-1788. | 0.72 | 1.36 | 2.01 |
| Guyon, Jeanne Marie Bouvier de La Motte, 1648-1717. | 0.73 | 1.35 | 2.00 |
| Middleton, Conyers, 1683-1750. | 0.88 | 1.31 | 1.77 |
| Robe, James, 1688-1753. | 0.64 | 1.31 | 2.07 |
| Forrester, Thomas, 1635?-1706. | 0.54 | 1.31 | 2.16 |
| Anderson, James, 1739-1808. | 0.71 | 1.30 | 1.92 |
| Grove, Henry, 1684-1738. | 0.85 | 1.27 | 1.72 |
| Coote, Charles, 1761-1835. | 0.74 | 1.21 | 1.73 |
| Heathcote, Ralph, 1721-1795. | 0.51 | 1.19 | 1.88 |
| Tucker, Abraham, 1705-1774. | 0.60 | 1.19 | 1.81 |
| Jackson, John, 1686-1763. | 0.71 | 1.18 | 1.69 |
| Brown, John, 1722-1787. | 0.75 | 1.14 | 1.53 |
| Ireland, Samuel, -1800. | 0.50 | 1.13 | 1.77 |
| Hare, Francis, 1671-1740. | 0.58 | 1.09 | 1.62 |

Table 3: Authors with the highest "random" effect on the rate of Scotticisms.

# B Scotticisms and Regular Expressions

| SCOTTICISM | REGULAR EXPRESSION |
|---|---|
| conform to | conform(s\|ed\|'d\|d\|ing)? to |
| friends and acquaintances | friends and acquaintances |
| maltreat | maltreat(s\|ed\|'d\|d\|ing)? |
| advert to | advert(s\|ed\|'d\|d\|ing)? to |
| proven | proven |
| improven | improven |
| approven | approven |
| pled | pled |
| incarcerate | incarcerat(es\|ed\|e\|'d\|d\|ing) |
| fresh weather | fre(s\|f)h weather |
| in the long run | in the long run |
| notwithstanding of that | notwith(s\|f)tanding of that |
| a question if | a que(s\|f)tion if |
| with child to a man | with child to a man |
| simply impossible | (s\|f)imply impo(s\|f)(s\|f)ible |
| in time coming | in time coming |
| nothing else | nothing el(s\|f)e |
| nothing else | no thing el(s\|f)e |
| severals | (s\|f)everals |
| anent | anent |
| allenarly | allenarly |
| alongst | along(s\|f)t |
| as I shall answer | as I (s\|f)hall an(s\|f)wer |
| cause him do it | cau(s\|f)(es\|ed\|e\|'d\|d\|ing)(me\|you\|him\|her\|it\|us\|them) do |
| marry upon | marr(ying\|y'd\|ies\|ied\|yd\|y) upon |
| effectuate | effectuat(es\|ed\|e\|'d\|d\|ing) |
| a wright | a wright |
| defunct | defunct |
| evite | evit(es\|ed\|e\|'d\|d\|ing) |
| part with child | part(s\|ed\|'d\|d\|ing)? with child |
| notour | notour |
| to be difficulted | (am\|is\|are\|was\|were\|been\|being\|be) difficult(ed\|'d) |
| think shame | (thinking\|thinks\|think\|thought) (s\|f)hame |
| in favours of | in favou?rs of |
| dubiety | dubiet(ys\|y's\|y\|ies) |
| compete | compet(es\|ed\|e\|'d\|d\|ing) |
| remeed | reme(de\|ed\|id\|ad)(s\|ed\|'d\|d\|ing)? |
| bankier | bankier(s'\|'s\|s)? |
| adduce a proof | adduc(es\|ed\|e\|'d\|d\|ing) a proof |
| superplus | (s\|f)uper-?plu(s\|f)((s\|f)?es)? |
| forfaulture | forfaulture(s'\|'s\|s)? |
| in no event | in no event |
| common soldiers | common (s\|f)oldier(s'\|'s\|s)? |
| debitor | debitor(s'\|'s\|s)? |
| exeemed | exee?m(ed\|'d\|d) |
| yesternight | ye(s\|f)ternight |
| big coat | big coat(s'\|'s\|s)?\| big-?coat(s'\|'s\|s)? |
| tenible argument | tenible argument(s'\|'s\|s)? |
| amissing | a-?mi(s\|f)(s\|f)ing |
| extinguish an obligation | extingui(s\|f)h(es\|ed\|'d\|d\|ing)? (an\|the\|(my\|your\|his\|her\|its\|our\|their)) obligations? |
| | \| extingui(s\|f)h(es\|ed\|'d\|d\|ing)? obligations? |
| depone | depon(es\|ed\|e\|'d\|d\|ing) |
| to inquire at a man | (e\|i)nquir(es\|ed\|e\|'d\|d\|ing) at a (man\|person) |
| angry at | angry at |
| to send an errand | (s\|f)en(ding\|ded\|ds\|d\|t) (an\|the) \| (s\|f)en(ding\|ded\|ds\|d\|t) errands? |
| to furnish goods to him | furni(s\|f)h(es\|ed\|'d\|d\|ing)? goods to (me\|you\|him\|her\|it\|us\|them) |
| to open up | open(s\|ed\|'d\|d\|ing)? up |
| Thucydide | thucydide |
| Herodot | herodote? |
| Sueton | sueton |
| butter and bread | butter and bread |
| pepper and vinegar | pepper and vinegar |
| paper, pen and ink | paper,? pen,? and ink |
| as ever I saw | as ever (I\|you\|he\|she\|it\|we\|they) saw |
| come in to the fire | (comes\|come\|coming\|came) in to the fire |
| alwise | alwi(s\|f)e |
| cut out his hair | (cut\|cuts\|cutting) out (my\|your\|his\|her\|its\|our\|their) hair |
| to get a stomach | (gotten\|getting\|get\|got) a stomach (for\|to) |
| vacance | vacance(s'\|'s\|s)? |

Table 4: Scotticisms from Hume's list and regular expressions used to find them in ECCO.