This is the accepted manuscript of the article, which has been published in Pahta P., Skaffari J. & Wright L. (eds) Multilingual Practices in Language History: English and Beyond. Boston/Berlin: De Gruyter Mouton. Language Contact and Bilingualism no. 15. ISBN: 978-1-5015-1381-7. ISSN: 2190-698X.

The social and textual embedding of multilingual practices in Late Modern English: A corpus-based analysis

Arja Nurmi, Jukka Tyrkkö, Anna Petäjäniemi and Päivi Pahta

1. Introduction

Recent research has established that multilingual practices, evidenced in the alternating use of two or more languages, are characteristic of language use in various types of English writings from different historical periods. While several single genres and topic domains have received attention in this body of research, most studies are based on relatively small datasets. We are still lacking a credible overview of the frequency and type of multilingual practices based on systematic corpus-based study. With the current availability of large masses of electronic text from historical periods of English, work on this scale is finally becoming possible.

This corpus-based study sets out to provide baseline evidence of the frequency and typology of multilingual practices in Late Modern England. The data comes from the 34-million-word *Corpus of Late Modern English Texts 3.0* (CLMET3), where the multilingual passages have been identified using a range of complementary automatic and semi-automatic techniques, including a new corpus tool, *Multilingualiser*, developed specifically for processing multilingual data. The corpus has been enhanced with sociolinguistic and text-typological background information to facilitate the analysis of multilingual practices and language-external factors. The enhanced data allows us in this study to present an evidence-based overview of (1) the frequency of foreign-language passages in written English in 1710–1920, (2) the variety of languages used in these texts in addition to English, (3) the connections of multilingual practices and the social variables describing the authors of each text, and (4) the further text-typological features associated with the use of multilingual practices.

2. Multilingual practices in written language in Late Modern England

2.1. Late Modern English and England

The Late Modern period, conventionally defined as covering the eighteenth and nineteenth centuries, was characterised by a rising awareness of English as a major European language. The major changes to English lexis, grammar, spelling and orthography were mostly over by the end of the seventeenth century, and both public and learned discourses concerning the need to standardise and maintain the purity of English emerged for the first time. During the eighteenth century in particular, influenced by Enlightenment ideals, both dictionaries and grammars were published for the first time at great volume, and ordinary men and women felt a growing need to speak “proper”

English and many took it upon themselves to attend lessons and read books that promised to improve elocution and, thereby, one's standing in polite society.

However, just as a good command of educated English was one of the marks of an educated and cultured person, so was a sufficient knowledge of foreign languages, in particular Latin, Greek and French. Latin and Greek were still the universal languages of learning and formed the backbone of young gentlemen's private education, and members of *le ton*, women in particular, were expected to read and speak French. A smattering of Italian, perhaps German as well, was expected to stick during one's grand tour, the common rite of passage for young men of the upper middle class, and Hobson-Jobson words were popular even in the language of those who had never been to the colonies.

A notable aspect in all this is the fact that much of this multilingualism took place within the community of English speakers. Although there were from time to time contacts with continentals and other foreigners, much of the code-switching was intended to index membership in a certain cultured and worldly society. The common attitude is expressed well in a passage from Arthur Conan Doyle's *Rodney Stone*, in which a boy's uncle imparts wisdom about code-switching to his nephew:

> "You sing, don't you, nephew?" he asked, suddenly.
> "Yes, sir, a little."
> "A baritone, I should fancy?"
> "Yes, sir."
> "And your mother tells me that you play the fiddle. These things will be of service to you with the Prince. Music runs in his family. Your education has been what you could get at a village school. Well, you are not examined in Greek roots in polite society, which is lucky for some of us. It is as well just to have a tag or two of Horace or Virgil: '**sub tegmine fagi**,' or '**habet foenum in cornu**,' which gives a flavour to one's conversation like the touch of garlic in a salad. It is not **bon ton** to be learned, but it is a graceful thing to indicate that you have forgotten agood deal. Can you write verse?" (Conan Doyle: *Rodney Stone,* 1896)

The Late Modern period was also the time when the novel emerged as the dominant form of English prose literature. This was a natural result of expanding literacy among the lower middle and the working classes, but also of fundamental change in how reading was increasingly viewed as a pleasurable and leisurely pursuit, rather than something associated with utility or devotional

activity.[1] Novels, along with other popular genres of narrative prose such as travelogues and biographies, were frequently set in foreign lands or featured foreign characters, giving authors reason to use foreign languages as a means of characterisation and adding local flavour to the stories.

2.2. Multilingual practices in writing

The frequency of multilingual passages in texts can be assumed to vary according to genre, audience and writer. The selection of languages available to a writer depends on the writer's educational opportunities and the contemporary trends in language use.

Based on results obtained from the multi-genre *Helsinki Corpus*, the frequency of switched passages on average ranges from 1.8/1,000 words in Old English to 2.2 in Middle English and 1.0 in Early Modern English (Pahta and Nurmi 2006). The vast majority of switched passages were in Latin during all three periods (1.8 for OE, 2.2 for ME and 0.9 for EModE), and the role of other languages was negligible. The second most frequent language of switched passages in Old English and Early Modern English was Greek (0.02 and 0.06 respectively), while during the Middle English period French came second (0.03/1,000 words). Contemporary European languages such as Italian and Spanish only started appearing during the Early Modern English period, and only sparsely then. The topic domains most favouring switching at all ages were religion and science. These results are in agreement with e.g. Voigts (1996), Hunt (2000) and Pahta (2003, 2004a, 2004b) as regards scientific writing, and with e.g. Wenzel (1994), Machan (1994), Iglesias-Rábade (1996), Schendl (2000), Halmari and Adams (2002) and Pahta and Nurmi (2011) in the case of religious texts. For the eighteenth century, the trend with scientific writing containing much multilingual material continues (Nurmi and Pahta 2010), while the domain of religion is increasingly vernacular (Nurmi and Pahta 2010; Tuominen, this volume).

In personal correspondence, similar trends can be observed. In Late Middle English personal letters, the frequency of switching varies, but the average is 0.55/1,000 words for Latin passages and 0.16 for French, other languages only appearing as isolated instances. The writer's social status and profession were decisive for frequency of switching and language choice (Nurmi and Pahta 2004). This can be linked to the daily pursuits of the writers, so that the litigious gentry included legal Latin in their letters, while merchants involved in foreign trade brought elements of French, Italian and Dutch to discussions of business with partners on the continent. The group that stands out most in terms of Latin usage are the clergy, who reach a frequency of 1.85 Latin passages / 1,000 words in the data. The use of foreign languages in the merchants' letters can be related to multilingual practices evident in business documents of the time (see e.g. Wright 1992, 1995, 1998,

[1] The love of reading was an emergent concept across Europe during the eighteenth century. See Mäkinen (2013).

2000, 2001; Rothwell 2000), while the Latin of the domain of law has been noted in Davidson (2003, 2005), as well as Nurmi and Pahta (2013).

One vital conditioning factor in the case of multilingual practices is education. Education was the means of gaining access to both classical languages of higher learning and the contemporary languages, which were needed for purposes of daily work in some cases, but, particularly in the course of the eighteenth century, became increasingly the markers of membership in the educated elites. From the seventeenth century onwards, French was the lingua franca of the educated European, while Italian was the language of music and arts. This is reflected in the multilingual practices evident in personal correspondence in the course of the eighteenth century. In the case of women, the availability of formal education or the means of self-study was dependent on both social status and the benevolence of men, whether fathers or husbands. It is only in the eighteenth century that women begin to show multilingual elements in their writing to a greater extent, the frequency of switched passages ranging from 0.2 in the fifteenth century to 1.0 in the eighteenth (Nurmi and Pahta 2011).

Another relevant feature, in addition to genre, topic domain and the writer's educational opportunities, is the reader. One writer could vary both the languages used and the frequency of switching according to the intended recipient of a letter, for example. So clergyman and Classical scholar Thomas Twining had an average of 6.8 or 6.7 switched passages / 1,000 words, when writing to Charles Burney the father and the son (respectively), but only 3.9 when writing to Fanny Burney (daughter and sister to the two Charleses). The languages Twining chose for each recipient also match his relationship with them: in letters to Charles Burney sr Twining switches frequently into French, but also into Italian, which was the language of their shared interest, music. When writing to the younger Charles Burney, a Classical scholar like Twining himself, the most frequent languages in addition to French are Latin and Greek. In the letters to Fanny Burney, almost all foreign-language passages are in French (Nurmi and Pahta 2010).

Thus, based on earlier research, we have arrived at three main hypotheses concerning the social variables related to multilingual practices. They are:

1. Education will play a role in the use of multilingual resources, because individual multilingualism (as opposed to societal multilingualism) is more common when there is access to structured language teaching (e.g. Nurmi and Pahta 2004, 2010, 2011; Pahta and Nurmi 2009, 2011).
2. Texts that (and authors who) have contacts with foreign environments (a novel set in France, a letter writer living in Italy, travel writing) are more likely to show multilingual practices (e.g. Nurmi and Pahta 2004, 2011; Pahta and Nurmi 2006, 2009).

3. Gender and social class are related to educational opportunities, but they will not provide significant results in this data due to the scarcity of female writers and the relative uniformity of informants' social background.

3. Material and methods

3.1. Material

As primary data we used the *Corpus of Late Modern English Texts 3.0* (CLMET3).[2] The corpus comprises 333 full-length texts published in 1710–1920; only texts by contemporary authors are included, so there are no reprints of early modern books. The combined word count is 34.3 million words covering a wide selection of genres ranging from academic treatises and learned essays to travelogues, biographies, novels and plays. Because the main purpose of CLMET3 is to serve as a generic corpus of Late Modern English, the selection of texts was primarily meant to serve the overall representativeness of the period and only secondly to represent the individual genres within it. The texts were harvested from the Project Gutenberg open access archive, edited for corpus use and checked for authenticity. The archive source introduces an opportunistic selection bias by favouring printed and generally prominent texts over a truly random sampling; however, in the present study this works in our favour because the primary research question concerns multilingualism in texts written for a general readership and one of the key objectives is to establish baseline evidence for the frequency of code-switching in the types of books an average reader might have encountered in the late modern period. Some authors, e.g. Jane Austen, Edmund Burke and Lewis Carroll, are represented in the corpus by more than one text, but none has more than three. All authors are native speakers of English and nearly all flourished in Britain. CLMET3 is quite representative of gendered writing, with 71 texts (21 per cent) written by female authors and 259 (78 per cent) by male authors. The mean age of authors is 41.6 years. The corpus also includes three samples from periodicals with numerous authors and some co-authored texts which for the purposes of sociolinguistic analysis were assigned to the perceived main author; for example, the novel *Diary of a Nobody* (1888) was co-authored by brothers George and Weedon Grossmith, but we list it under the elder brother George. The periodicals were left out of the present analysis, as they represent a highly multi-authored text type and the number of samples is too small to allow reasonable generalisations to be made.

[2] CLMET3 was compiled by Hendrik de Smet, Hans-Jürgen Diller and Jukka Tyrkkö. A CQP-ready version of CLMET3.0 was released in October 2015 by Hendrik De Smet, Susanne Flach and Jukka Tyrkkö. The new version of the corpus, CLMET3.1, also comes with a new cleaned-up version of part-of-speech tagging. Like CLMET3.0, the new corpus is freely available from Hendrik De Smet at https://perswww.kuleuven.be/~u0044428/.

The accurate and uncontroversial assigning of genre labels is a notoriously difficult task.[3] Although ambiguity and uncertainty are to some extent an accurate assessment of the difficulty inherent to invariably simplistic genre labels, for the purposes of corpus linguistic analysis such highly detailed descriptors can render genre an effectively moot variable. Thus, for the purposes of the present study, the genre system in CLMET3 was cleaned up by collapsing certain categories into one, resolving double classifications, and assigning a category to texts which were originally categorised as X, or ambiguous. For example, several texts were classified in CLMET3 as both Biographies and Travelogues or as both Histories and Treatises, and these were revisited and the texts were assigned to the genre category that appeared the most appropriate. The pruned genre system comprises twelve genres, which when necessary can be further combined into the three macro-genres of *Drama*, *Fiction* and *Non-fiction* (see Table 1); although CLMET3 includes a small number of periodical articles as well, we leave them aside in the present study. The overall text counts are typically very modest in most individual genres and thus the macro-genre level is most useful when it comes to statistical analysis of frequency data.

Table 1. Macro-genres and genres in CLMET3 with the number of each.

Macro-genre	**Genre**	**1710–1780**	**1780–1850**	**1850–1920**	**Total**
Drama					**74**
	Drama	25	19	30	74
Fiction					**131**
	Novels	36	35	56	127
	Children's story	0	1	3	4
Non-fiction					**125**
	Biography	4	8	7	19
	Essay	3	2	1	6
	History	3	2	5	10
	Instruction	1	0	1	2
	Letters	4	6	4	14
	Religious	0	0	2	2
	Travel	1	5	4	10
	Treatise	18	20	24	62

Our focus on sociolinguistic metadata rises, firstly, from the research tradition of stratificational sociolinguistics and, secondly, from issues specific to the study of individual multilingualism

[3] See e.g. Diller, De Smet, and Tyrkkö (2011).

identified in previous research by Pahta and Nurmi. In the stratificational tradition (e.g. Labov 1994, 2001; Chambers 2003; Nevalainen and Raumolin-Brunberg 2003), we have tracked the author's age (operationalised as year of birth), gender (as a binary variable; overwhelmingly male) and place of birth (in the broad categories of *South England*, *North England*, *Ireland*, *Scotland*, *Wales* and *Abroad*). We also made an initial classification into social classes, but this proved unworkable, since the clear majority of the writers represent the middle classes. In order to track some differences among the authors, they were given a main occupation following the information found in the *Oxford Dictionary of National Biography*, and these were classified into four main categories: *professional writer* (either of fiction or non-fiction, e.g. Charles Dickens, Henry Fielding), *academic* (e.g. Charles Darwin and Charles Babbage), *cultured* (a catch-all category including people who made their living in the arts, but whose main occupation was not writing, including e.g. Horace Walpole and Edward Gibbon) and *other* (e.g. Charlotte Brontë and Henry Rider Haggard). An individual writer's occupation has been classified varyingly along their lifespan, since many began in one category (particularly *cultured* or *other*) before they found success in writing and were able to become full-time authors.

For the study of multilingual practices, it is important to be familiar with the writer's linguistic profile, and with extralinguistic factors affecting it. In order to track these, we looked at both educational background and travel history. Both depend on sometimes haphazardly preserved information, and may not have been adequately described in our main source, the *Oxford Dictionary of National Biography*. The educational background of an author has been tracked in the form of five binary categories: *grammar school*, *university*, *other formal education*, *private tuition* and *education abroad*. These give us some indication of the languages available for our authors to learn, as, for example, a university education still meant the study of Latin. However, while education provides us with some idea of language teaching, it is difficult to estimate how much any individual would have actually learned during the course of their education. Likewise, it is possible that an individual may have acquired language skills which have not been recorded in biographies.The other element of a writer's linguistic profile was their travel history. We have tracked this in broad categories of *Britain* (i.e. the writer never travelled abroad), *Europe* (except French-speaking countries and Italy), *Inner Circle* (following Kachru's classification) and *World* (with the possibility of contact with non-European languages). *French-speaking countries* and *Italy* were singled out, since French and Italian were two frequently occurring foreign languages, and we wanted the opportunity to observe the impact of e.g. travelling in Italy on a writer's use of Italian phrases in their writing. All these variables were tracked as binary options, and, with the exception of the category Britain, one writer could have "yes" in more than one category. Finally, we have made an estimate of the writers' overall linguistic profiles, listing the main foreign languages a

writer would probably know based on their education and travel. We have binary categories for *Latin*, *Greek*, *French*, *Italian* and *German*.

3.2. Data extraction and the analytical procedure

The first order of business when analysing the frequency of switched passages is to identify them in a corpus. While such a task is doable by manual analysis when the corpus is suitably small, it becomes prohibitively time-consuming and labour-intensive when the corpus comprises 34 million words. On the other hand, a fully automated computational approach would be likely not to reach the required level of analytical prowess, either missing words and passages that a human reader considers code-switches (false negatives) or, conversely, falsely assigning as code-switches words and passages which are not (false positives). To avoid systematic problems of precision and recall, we decided to use a semi-automatic method by designing a multilingualism detection tool which would identify and tag potential multilingual passages at a reasonably high level of precision and then turn the data over for manual pruning. The tool, named *Multilingualiser*,[4] makes use of an iterative, stepwise algorithm which starts with simple dictionary look-up using one or more of the built-in wordlists (Latin, French, Italian and German) or a user-defined wordlist. The tagger then analyses the results of the first pass looking for strings of tags of the same type and their untagged collocates. The number of closely proximate tags increases the likelihood that the tagging is correct and that the untagged items in the same string should also be tagged. For example, take the following sentence from *Red Pottage* (1899) by Mary Cholmondeley:

(1) Rachel was not by nature *de celles qui se jettent dans l'amour comme dans un précipice*.

In this instance the software should understand that the entire end of the sentence, beginning with *de*, is French. What makes the task difficult is that *nature* is a visual diamorph and could be either French or English, and that *de* and *un* could also be Spanish, Portuguese or Italian (even if the latter would be quite rare).[5] Furthermore, in this particular instance the French dictionary did not include the word *jettent*, the third person plural present and subjunctive form of *jeter* ('throw'). Thus, to work properly the software needs to understand that the unambiguously French function words *celles*, *qui*, *se*, *dans* and *comme*, as well as the equally clear French content words *l'amour* and *précipice*, signal that *de* and *un* have a very high likelihood of also being French, and likewise that

[4] *Multilingualiser* was developed by Jukka Tyrkkö in LiveCode and will run in OS X, PC and Linux. The tool will be available to the research community free of charge once the development reaches the first stable version. A forthcoming development will see the inclusion of first-dating information based on the *Historical Thesaurus of English*. The algorithms and the statistical issues involved are discussed in detail in Tyrkkö, Nurmi, and Tuominen (forthcoming 2017).

[5] For an introduction to and further discussion of the term visual diamorph, see ter Horst and Stam, this volume.

the likelihood of *jettent* being French is also very good, especially since it is also not an English word. By contrast, because *nature* is both an English word and a French word and it occurs in a sequence-initial position, it may be assigned as ambiguous with a high likelihood of French – which, in this particular case, it is in fact not. The tool is designed to give the user control over the likelihood weights of a variety of collocate positions within the sequence of items, as well as the freedom to set other related parameters, such as how to treat sentence breaks, capitalised items, known function words, and so on. Additional retrieval methods in *Multilingualiser* include a character n-gram-based method for discovering non-English words using word-initial and word-final character trigram sequences; for example, the sequence *ips-* appears word-initially only in the word *Ipswich* in English. *Multilingualiser* can also ignore proper nouns unless they occur within a longer sequence of foreign items.

The first round of tagging and manual pruning was performed using *Multilingualiser* and its in-built tag editor. The initial results of potential switched passages revealed that one-word instances showed a poor recall rate, since many of them were established loans, such as *auditorium* or *inferno*, or English words which share a form with a foreign item, most typically French or Latin (See Tyrkkö, Nurmi, and Tuominen forthcoming 2017 for more on this). This lead to our focusing – at least at this stage of research – on stretches of two or more words in a language besides English. Obviously, this method disregards some genuine code-switches, but our overall view, based on analysis of the initial datasets, is that the number of instances overlooked by this method is negligible. The segments of foreign words were then manually pruned for any remaining items that members of the research team considered to be English despite appearances, the results of the language identification were checked and corrected when necessary, and the rare languages were manually identified.

The switched passages were then manually classified into three main categories, to a large extent based on previous work by Nurmi and Pahta (see e.g. Nurmi and Pahta 2010, 2013; Pahta and Nurmi 2009). The categories present a continuum from more established to less established switches. Conventionalised passages, such as *cara sposa*, *fille de chambre* or *terra firma*, are typically 2–3 words long, appear frequently in English texts and may be familiar to a reader with very poor language skills (see examples 2–3). There are many expressions which could be classified as terms, including *a priori*, *beau monde*, *carte blanche*, *canto fermo* and *ipso facto*. We have chosen to include all such borderline cases at this stage of our study.

(2) The fair *fille de chambre* came close up to the bureau where I was looking for a card (Sterne, *A Sentimental Journey*)

(3) Now I assert, that whoever reasons after this manner, does *ipso facto* believe the actions of the will to arise from necessity (Hume, *A Treatise of Human Nature*)

The second category, requiring somewhat more linguistic skill from both reader and writer is that of prefabricated expressions. Typical examples include (4) and (5). These are usually quotations, proverbs and maxims; we have classified instances in this category only if the quotation could be identified as coming from a recognised source or the expression was listed as a proverb or otherwise appears to have been in general use. Also identification as a quotation within the corpus text itself has been accepted as proof of the prefabricated status, as in (4). Quotations are not necessarily of written data: they can also be reported speech (6).

(4) that it may always apply to itself that celebrated passage in **Lucan**, *Nec quenquam jam ferre potest Caesarve priorem, Pompeiusve parem*. Indeed, … (Fielding, *Amelia*)
(5) All which, from the words, *De gustibus non est disputandum*, and whatever else… (Sterne, *The Life and Opinions of Tristram Shandy*)
(6) D'Aubreu, the pert Spanish minister, said the other day at court to poor Alt, the Hessian, '*Monsieur, je vous félicite, Munster est pris*.' (Walpole, *Letters*)

Finally, the third category of switched passages contains all instances not classifiable as conventionalised or prefabricated. These are often longer than the other two, and require a higher level of command of the other language included as well. Reported speech in fiction (example 7) is included in this category, since it is the author's own production, not that of an actual interlocutor of his or hers. Similarly, a writer reporting his/her own words in conversation is regarded as free switching, even if they were first put together at an earlier time (8).

(7) 'Mr Western,' answered the lady, 'you may say what you please, *je vous mesprise de tout mon coeur*. I shall not therefore be angry.' (Fielding, *The History of Tom Jones*)
(8) he asked me if Mr. Pitt was like his sister, I told him, '*Qu'ils se ressembloient comme deux gouttes de feu*.' (Walpole, *Letters*)

The analysis of frequency data will be presented in Section 4, but a brief general comment about the statistical nature of this type of data is in order here. As is well-attested, corpus linguistic data is rarely, if ever, derived from truly random samples and the linguistic features under investigation are even more rarely seen to be normally distributed, that is, linguistic data often does not follow a Gaussian curve. Consequently, both the descriptive and inferential statistical methods used ought to

be robust. To take a simple example, the mean (or average) frequency of a small dataset can be easily skewed by one of two texts that show an unusually high frequency. While some linguists choose to interpret this as merely an artefact of the fickle nature of linguistic data, it is often wiser to use methods which deal with outliers in a more organised fashion – or at the very least, it is important to be aware of outliers in the data. The normality of a dataset can be examined in a number of different ways, but one of the most convenient of these is a visual examination of a quantile-quantile plot where the quantiles of a primary dataset are plotted against normal quantiles. In a normally distributed dataset the data points follow a diagonal straight line, while in a skewed dataset they do not. As Figure 1 illustrates, the distribution of switched chunks in CLMET3 is far from normal: there are many texts with very few if any tags and then a smaller number with high frequencies.

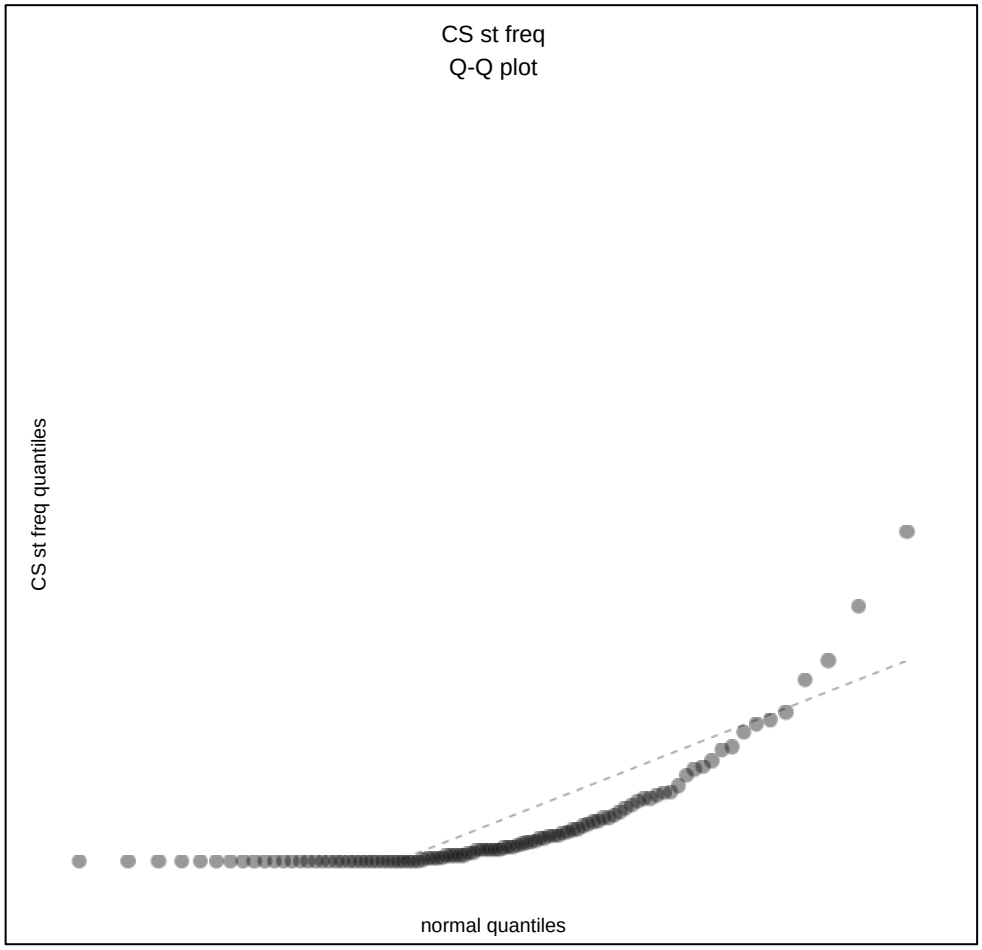

Figure 1. Quantile-quantile plot showing the non-normality of the distribution of code-switched segments (of two words or longer) in CLMET3.

To deal with the distributional properties of our data, we will therefore use robust non-parametric methods of inferential analysis without making assumption of normality. Although it may be argued that some parametric tests such as one-way Anova are not particularly sensitive to non-normality,

our data is too strongly skewed for us to consider them. In practice then, we use the Wilcoxon ranked sum test (also known as the Mann-Whitney U test) in our monofactorial analysis, comparing each pair of factor levels separately. Similarly to the parametric independent sample Student's t-test, the null hypothesis in the Wilcoxon test is that the two samples tested come from the same population and the alternative hypothesis is that the populations are different. For the Wilcoxon test the values of the continuous variable (here, standardized frequencies) are ranked, that is, all the datapoints are combined into a single dataset, the values are organized according to size and assigned ordinal ranks, after which the ranks are sorted back to the original groupings. We report the means and standard deviations as well as the z-scores used in calculating the approximate p-value. The null hypothesis is rejected at the significance level of 0.05; in the notation, we denote $p<0.05$ with *, $p<0.01$ with ** and $p<0.001$ with ***.

4. Results: Overview of languages and their frequencies, genre findings, sociolinguistic findings

The analysis reveals that the mean frequency of foreign-language segments of two words or longer (hereafter CS segments) is 0.14/1,000 words and the median is 0.04/1,000 words (Figure 2).[6] This is clearly lower than the result for the Early Modern English period of the *Helsinki Corpus*, as well as the frequencies attested in personal correspondence in the eighteenth century. These numbers hide a wide range, however: the highest frequency of CS segments, 2.59/1,000 words, is found in Laurence Sterne's *A Sentimental Journey through France and Italy* (1768). The mean chunk length in the corpus is 13 in words and the median 4 words. The longest chunk in the corpus is 2,944 words in *Stories from the Italian Poets* (1845) by Henry James. There are 104 texts with no CS segments, though it is important to remember that single-word switches are not counted here. There is no correlation at all across the timeline, which means that in terms of both frequency and switch type CS segments occurred in a steady manner throughout the late modern period.

[6] Although switching back to English from a foreign language segment is technically also a code-switch, we count switch-points exclusively as transitions from English to another language.

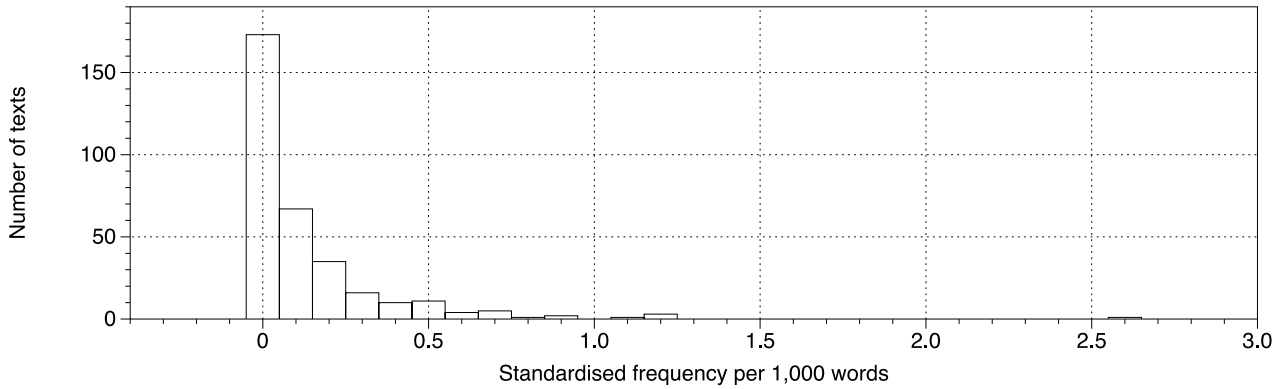

Figure 2. Frequency of code-switched segments (of two words or longer) in CLMET3 by text.

The majority of the CS segments in CLMET3 are in French or Latin; the two are essentially equal when it comes to standardised frequencies of chunks (see Table 2). There are twenty-one different languages attested altogether; the identification of five is uncertain.

Table 2. CS segments in CLMET3.

Language	**CS segments**	**Total word count in the CS segments**
Arabic	18	53
Buginese?	1	6
Dutch	3	9
Egyptian?	2	6
French	1937	21924
Gaelic	1	4
German	110	2630
Greek	270	3018
Hawaian?	1	5
Hindi	1	2
Indonesian	6	12
Irish	6	21
Italian	358	17527
Latin	2156	23544
Lenape?	1	18
Malay	1	2
Malay?	1	2

Portuguese	19	108
Romani / Caló	8	41
Samoan	1	8
Sinhalese	1	2
Spanish	26	154

As Table 2 shows, the foreign languages can be roughly divided into two groups: high-frequency languages and low-frequency languages. Latin and French are by far the most common languages in terms of number of individual foreign-language segments, but Italian comes relatively close in word count, mainly due to several extremely long switches. The rare languages typically occur in a single text in CLMET3, and consequently any differences observed between the low-frequency languages are essentially random artefacts and they cannot be considered indicative of language-specific frequency differences; for example, although CLMET3 happens to feature Malay or (possibly invented) Egyptian in small frequencies, one should not expect to encounter these languages in Late Modern English as a matter of course. The correlation between the number of segments and word count is illustrated in Figures 3 and 4.

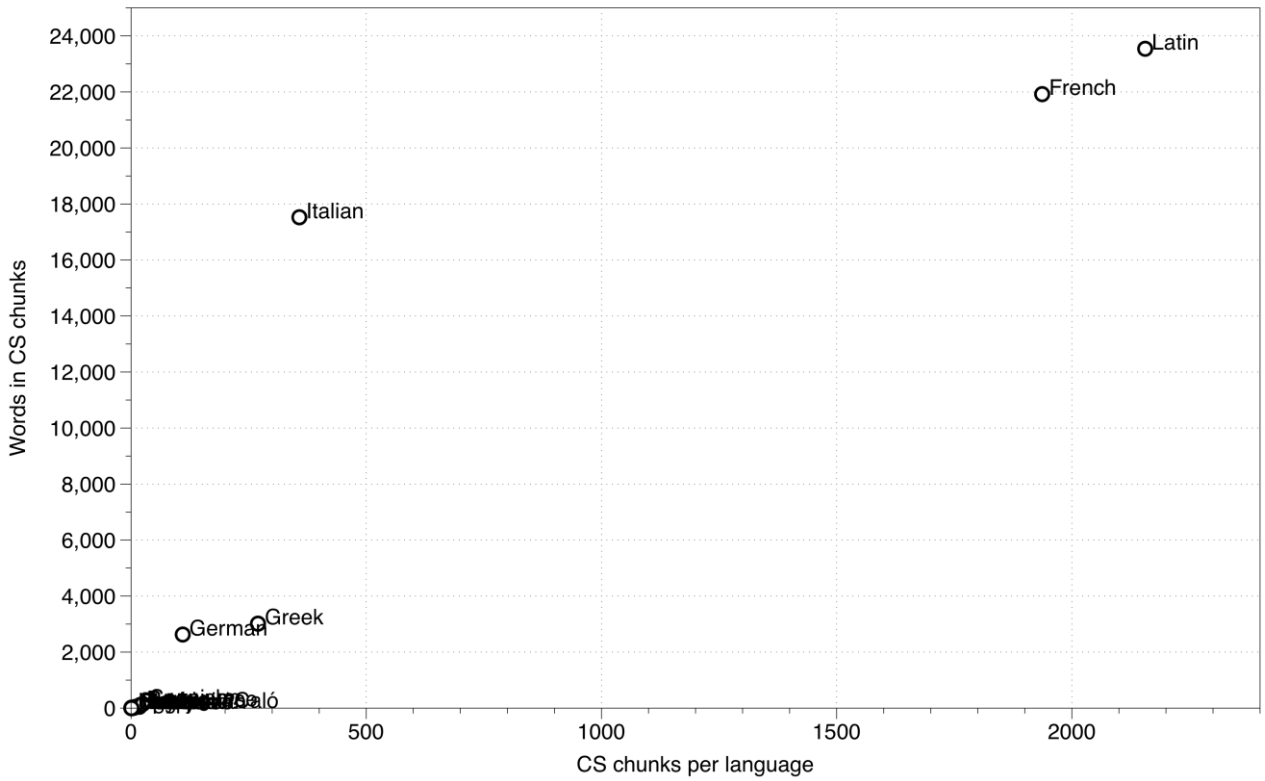

Figure 3. Scatter plot displaying correlation between number of CS segments and their combined word counts.

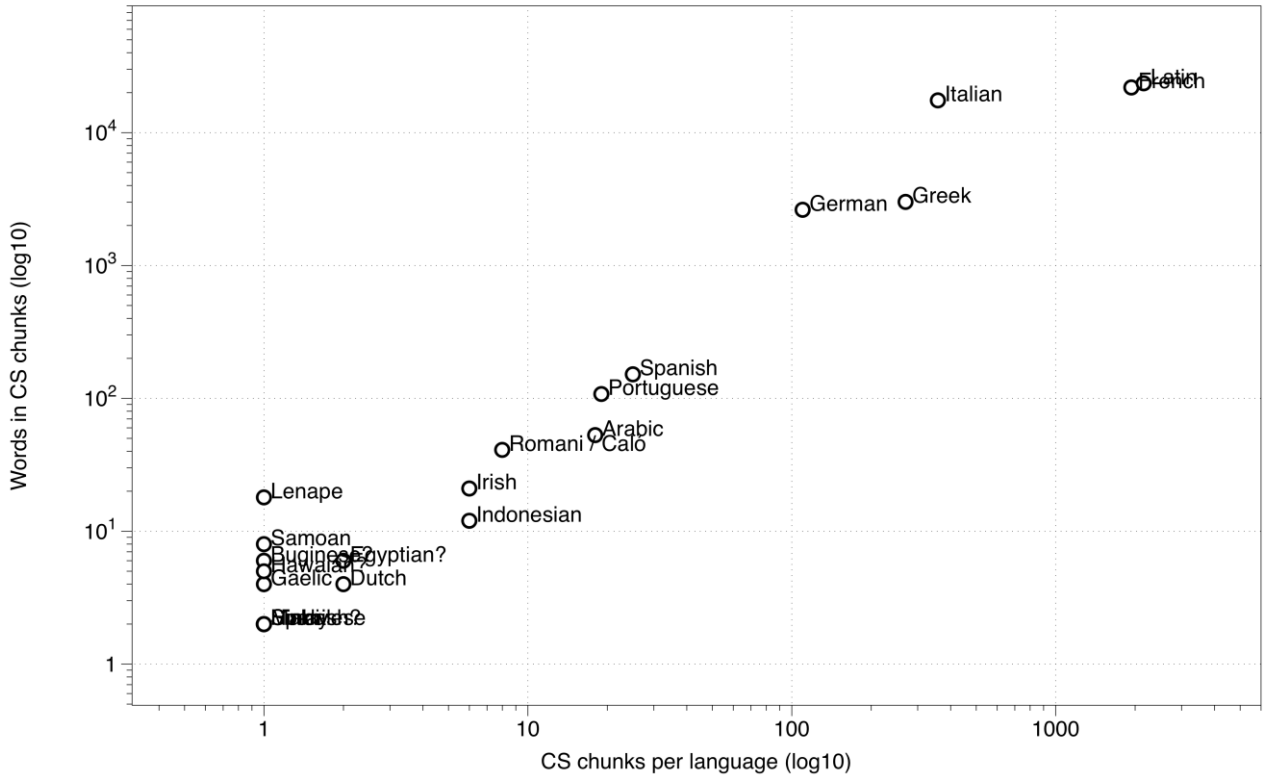

Figure 4. Scatter plot displaying correlation between number of CS segments and their combined word counts using $\log_{10}$ scale.

The switch type variable reveals that the vast majority of CS segments are pre-fabricated in nearly all the languages (see Figure 5). With the exception of Hindi and Irish, both of which are extremely infrequent, French is the only language where the majority of CS segments are not pre-fabricated, but rather conventionalised. Perhaps the most striking observations concern Latin which, despite endless years of cramming in school, is hardly ever used in a free and original fashion: only 77 (3.5%) chunks out of 2,154 have been categorised as free switches.

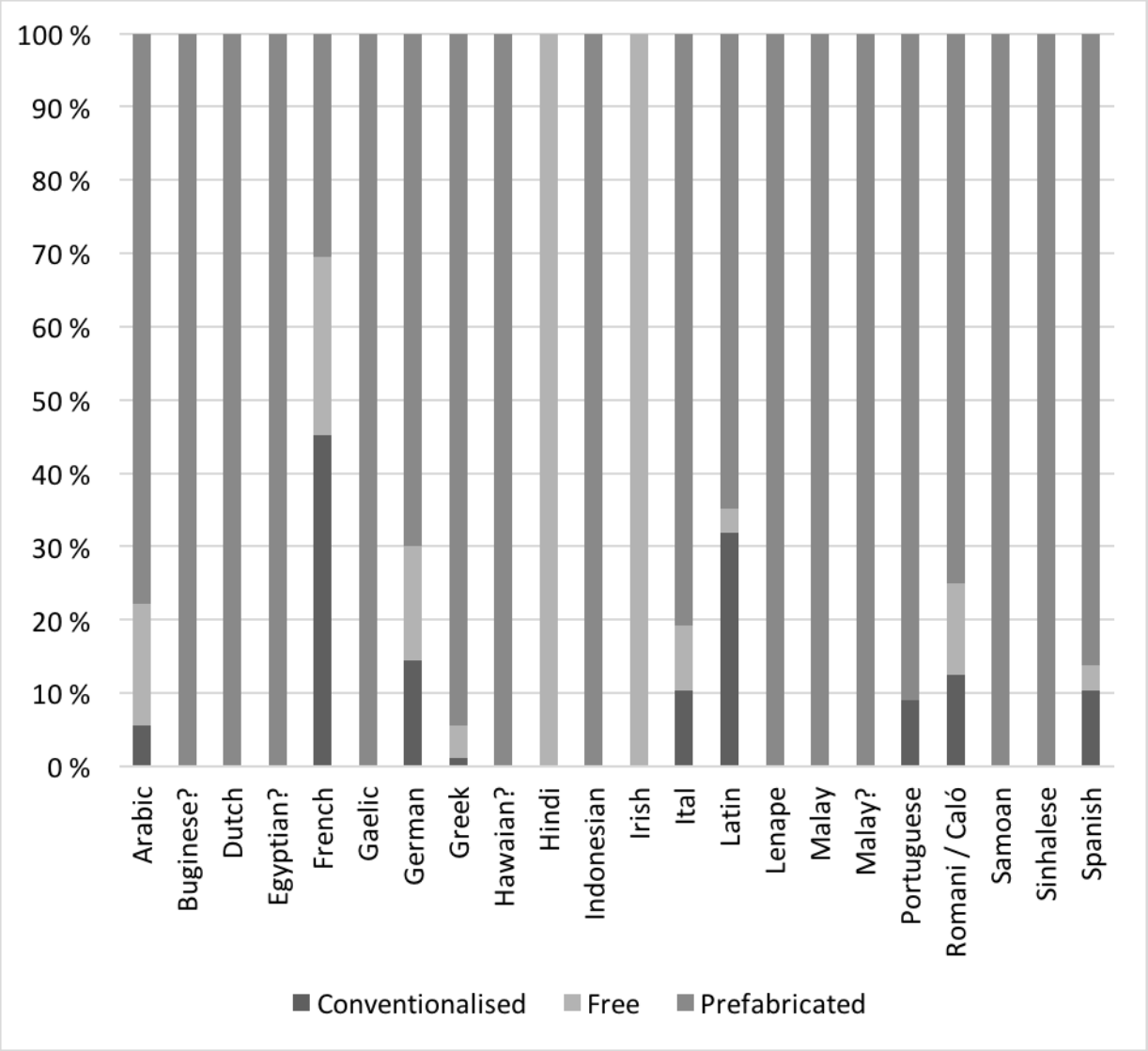

Figure 5. Switch type proportions by language.

Moving on to monofactorial analysis and starting with the macro-genres, we see that Non-fiction texts generally contain more switched passages than either Drama or Fiction, though there are notable outliers in both; the differences between Non-fiction and both Drama and Fiction are significant. This finding agrees perfectly with the previous observations concerning the prevalence of pre-fabricated switches: the most typical function of foreign language use in Non-fiction is quoting from original sources.

Table 3. All CS segments in macro-genres.

Macro-genre	**n**	**Overall mean (/1,000**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Non-Fiction vs. Fiction Z=5.44, p=***

		words)		Non-Fiction vs. Drama
Drama	74	0.11	0.21	Z=5.22, p=***
Fiction	131	0.10	0.25	Fiction vs. Drama
Non-fiction	124	0.20	0.25	Z=1.69, p=0.20

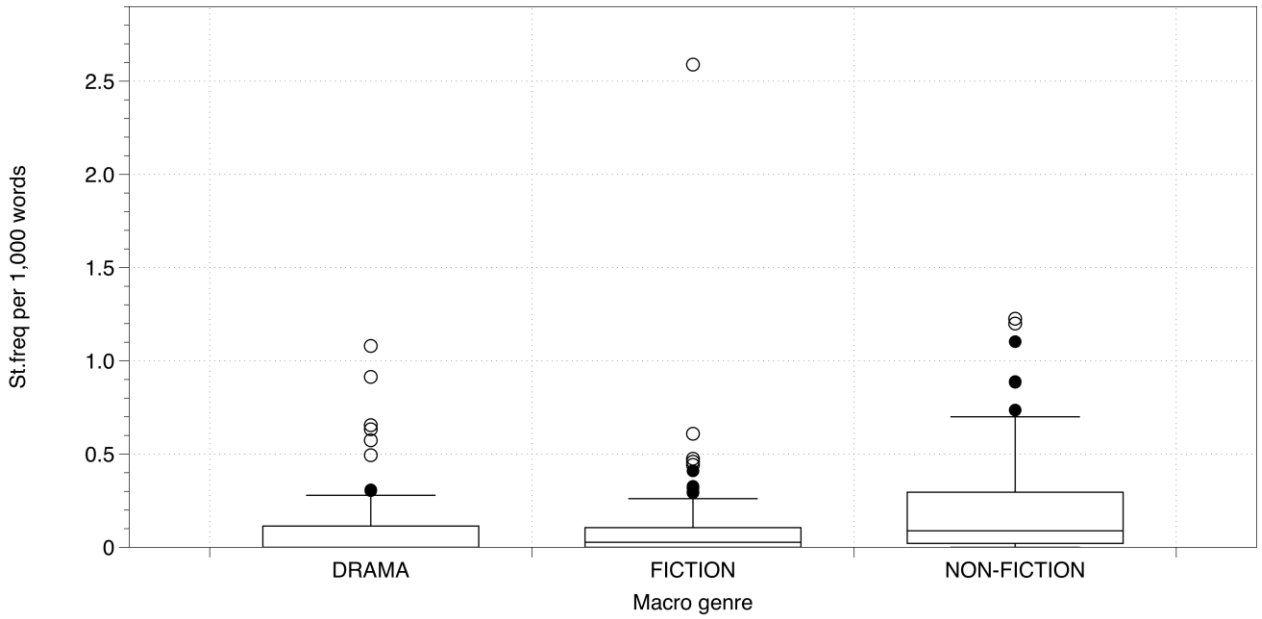

Figure 6. All CS segments in macro-genres.

Predictably, Non-fiction writing contains Latin, Greek and German in greater frequencies than Fiction or Drama, while French is used more in Drama and in Fiction. Perhaps against expectations, Italian appears most frequently in Non-fiction. Although Italian was a fashionable language among the upper classes, few possessed real mastery of the language. Knowing this, authors of Fiction or Drama usually limited the use of Italian to conventionalised greetings and exclamations. By contrast, Non-fiction authors who use Italian, such as Edward Gibbon, do so knowing that their readers are likely to understand, or at least appreciate, long quotes in the original language.

Table 4. Appearance of most frequent languages in the three supergenres.

LATIN

Macro-genre	**n**	**Overall mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Non-Fiction vs. Fiction Z=4.55, p=*** Non-Fiction vs. Drama
Drama	74	0.07	0.19	Z=4.05, p=***
Fiction	131	0.02	0.05	Fiction vs. Drama

Non-fiction	124	0.16	0.66	Z=1.28, p=0.19

FRENCH

Macro-genre	**n**	**Overall mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Non-Fiction vs. Fiction Z=2.22, p=* Non-Fiction vs. Drama Z=2.07, p=** Fiction vs. Drama Z=1.51, p=0.13
Drama	74	0.14	0.54	
Fiction	131	0.03	0.07	
Non-fiction	124	0.22	1.94	

GERMAN

Macro-genre	**n**	**Overall mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Non-Fiction vs. Drama Z=1.93, p=* Non-Fiction vs. Fiction Z=1.04, p=0.29 Fiction vs. Drama Z=1.21, p=0.22
Drama	74	0.002	0.013	
Fiction	131	0.002	0.007	
Non-fiction	124	0.006	0.03	

ITALIAN

Macro-genre	**n**	**Overall mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Non-Fiction vs. Drama Z=3.17, p=*** Non-Fiction vs. Fiction Z=2.24., p=* Fiction vs. Drama Z=1.49, p=0.13
Drama	74	0.003	0.02	
Fiction	131	0.004	0.02	
Non-fiction	124	0.02	0.10	

GREEK

Macro-genre	**n**	**Overall mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Non-Fiction vs. Drama Z=2.57, p=*** Non-Fiction vs. Fiction Z=2.84., p=*** Fiction vs. Drama
Drama	74	0.01	0.05	
Fiction	131	0.002	0.02	

Non-fiction	124	0.017	0.07	Z=0.50, p=0.61

Links between sociolinguistic parameters and code-switching confirm many of the hypotheses noted earlier in Section 2.2. For example, it appears clear that while gender or place of birth do not predict the multilingual practices in any way, the author's occupation plays a central role in the use of foreign languages, with Academic authors using multilingual resources most frequently and Professional Writers the least. Statistically significant differences can be observed between Professional Writers and Academics, and between Professional Writers and Cultured authors. By contrast, Academic and Cultured authors cannot be said to differ, nor Professional Writers and Others. This division can be understood in terms of the authors' professions, but also as a reflection of their respective audience designs.

Table 5. All CS segments by occupation.

Occupation	**n**	**Overall Mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:**
Academic	30	0.16	0.18	Cultured vs. Academic Z=0.25, p=0.79 Professional Writers vs. Others Z=1.04, p=0.29 Others vs. Cultured Z=1.69, p=0.09 Others vs. Academic Z=1.75, p=0.08 Professional Writers vs. Cultured Z=2.66, p=*** Professional Writers vs. Academic Z=2.66, p=***
Cultured	56	0.22	0.30	
Other	73	0.12	0.20	
Professional Writer	170	0.11	0.25	

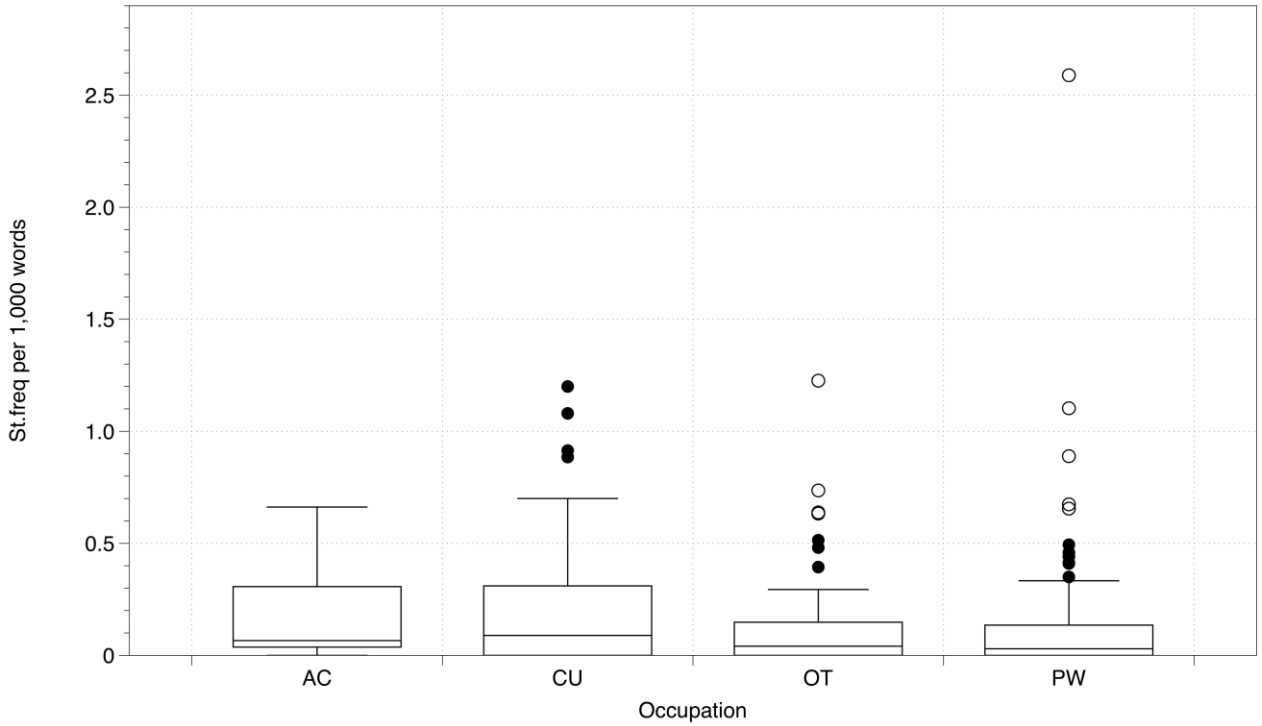

Figure 7. Box plot of multilingual practices and occupation.

The situation changes slightly when we turn to specific languages. Focusing on the two most frequently used foreign languages, we see that Academics and Others use Latin the most frequently, and that the differences in Latin use are statistically significant between Academics and Others, Academic and Cultured authors, and Professional Writers and Academics. By contrast, French is used most frequently by Cultured authors and the least by Academics.

Table 6. Most frequent languages by occupation.

LATIN

Occupation	**n**	**Overall Mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:**
Academic	30	0.11	0.16	Cultured vs. Academic Z=1.85, p=* Professional Writers vs. Others Z=0.69, p=0.48 Others vs. Cultured Z=0.42, p=0.67 Others vs. Academic Z=2.38, p=** Professional Writers vs. Cultured Z=1.11, p=0.26 Professional Writers vs. Academic
Cultured	56	0.08	0.18	
Other	73	0.17	0.85	
Professional Writer	170	0.04	0.12	

				Z=3.25, p=***

FRENCH

Occupation	**n**	**Overall Mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:**
Academic	30	0.03	0.04	Cultured vs. Academic Z=1.25, p=0.20 Professional Writers vs. Others Z=-1.01, p=0.82 Others vs. Cultured Z=-1.63, p=0.3 Others vs. Academic Z=-0.21, p=0.82 Professional Writers vs. Cultured Z=-2.78, p=*** Professional Writers vs. Academic Z=-0.88, p=0.37
Cultured	56	0.18	0.59	
Other	73	0.34	2.53	
Professional Writer	170	0.04	0.11	

The switch types correlate with the author's occupation as well. When it comes to conventionalised CS segments, all occupations use them more or less equally and there are no statistically significant differences. Free switches are used the most by Cultured authors and a statistically significant difference is seen between them and Others. Finally, pre-fabricated CS segments are used similarly by Academic and Cultured authors, and Professional Writers and Others, respectively, the two pairs being statistically different.

Table 7. Switch types by occupation.

CONVENTIONALISED

Occupation	**n**	**Overall Mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:**
Academic	30	0.06	0.19	Cultured vs. Academic Z=-0.19, p=0.84 Professional Writers vs. Others Z=-0.61, p=0.54 Others vs. Cultured Z=-0.95, p=0.33
Cultured	56	0.08	0.20	
Other	73	0.05	0.09	

Professional Writer	170	0.05	0.12	Others vs. Academic Z=-0.61 p=0.52 Professional Writers vs. Cultured Z=-1.70, p=0.26 Professional Writers vs. Academic Z=-1.11, p=0.26

FREE

Occupation	**n**	**Overall Mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Cultured vs. Academic Z=1.81, p=0.06 Professional Writers vs. Others Z=1.62, p=0.10 Others vs. Cultured Z=-2.15, p=* Others vs. Academic Z=0.24, p=0.80 Professional Writers vs. Cultured Z=-0.96, p=0.33 Professional Writers vs. Academic Z=1.39, p=0.16
Academic	30	0.003	0.01	
Cultured	56	0.03	0.11	
Other	73	0.008	0.03	
Professional Writer	170	0.02	0.12	

PREFABRICATED

Occupation	**n**	**Overall Mean (/1,000 words)**	**Standard deviation**	**Wilcoxon Each Pair Comparisons:** Cultured vs. Academic Z=-0.10, p=0.91 Professional Writers vs. Others Z=-1.17, p=0.24 Others vs. Cultured Z=-1.85, p=* Others vs. Academic Z=-1.72, p=0.08 Professional Writers vs. Cultured
Academic	30	0.09	0.16	
Cultured	56	0.10	0.17	
Other	73	0.06	0.15	
Professional Writer	170	0.04	0.12	

				Z=-3.32, p=*** Professional Writers vs. Academic Z=-2.90, p=***

If we treat education as a single factor, it is immediately clear that it is perhaps the single best predictor when it comes to high rates of multilingual practices. Even grammar school prepares a person for including multilingual elements in their writing – 0.14/1,000 words (n=60) vs. 0.12/1,000 words (n=273), p=* – but university education is the real threshold. The mean frequency of switches in texts written by university graduates is 0.17/1,000 words (n=139) compared to 0.09/1,000 words (n=194) in texts written by those who did not attend a university; the difference is statistically significant (Figure 8).

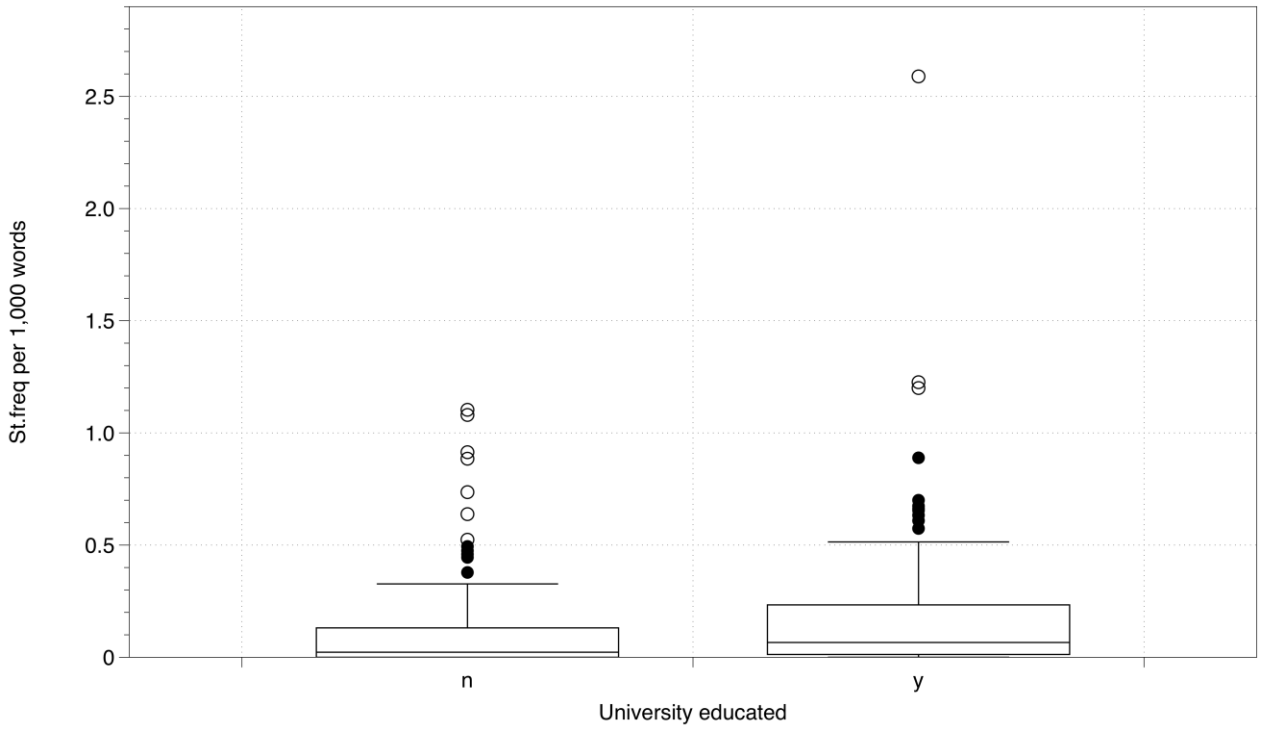

Figure 8. CS segments and university education.

When it comes to specific languages, Latin and Greek are strongly associated with university education, but French, Italian and German show no statistically significant differences between the two groups (Table 8). Note that although the frequency differences between the two groups appears quite large when it comes to French, the difference is not statistically significant because there is great variation within each group. Naturally, there are significant co-variances here when it comes

to the author's education and the types of text they are likely to produce; for example, Academic texts are usually written by university graduates.

Table 8. Most frequent languages and university education.

Language	**St.freq. among university educated**	**St. freq, among non-university graduated**	**Z**	**p**
Latin	0.13	0.05	3.33	***
French	0.21	0.07	0.65	0.51
Italian	0.01	0.009	0.67	0.50
German	0.005	0.002	-0.01	0.99
Greek	0.011	0.009	2.07	*

Education abroad also predicts an apparent tendency for multilingual practices. Those educated abroad have an overall mean frequency of 0.22/1,000 words (n=33) while those educated in Britain show a frequency of 0.12 (n=296); z=2.98, p=***. As for specific languages, education abroad has a statistically significant effect on the use of Latin, French and Greek, but not on Italian and German. Interestingly, the higher frequency is associated with the use of conventionalised switches but not pre-fabricated or free switches. The data also shows that spending time in a French-speaking country increased one's likelihood of using French, Latin, Italian and even German, but not Greek, while spending time in Italy increased the likelihood that a person code-switches to any of the five most common foreign languages.

5. Conclusion

To our knowledge, the present study is the first corpus linguistic study of historical multilingual practices to attempt a frequency-based analysis of code-switching using a medium-sized, genre-stratified corpus. One of the main motivations behind this continuing study is the desire to go beyond hand-picked examples and the idiolects of specific authors, and to understand multilingualism as more general feature of language use. By establishing evidence-based baseline frequencies for foreign content in English texts we provide a backdrop against which the findings of more intensively focused studies can be evaluated, but also to recognise outliers and significant predictors which smaller datasets cannot provide and to test hypotheses formed on the basis of small-scale studies.

Based on earlier studies of historical multilingualism, we knew that the overall frequency of foreign content is generally relatively low in English texts, and consequently a semi-

automatic method of discovery was developed for that purpose in the course of the project (see e.g. Tyrkkö, Nurmi, and Tuominen forthcoming 2017). Our experience with the Multilingualiser suggests that further training of the tool and analysis of the results may help in automatic separation of visual diamorphs, improving precision. Other issues that we will continue to work on involves developing a more data-driven method for disambiguating between conventionalised short phrases that ought to be considered part of the English lexicon and those that remain effectively foreign, and exploring the context-dependent nature of "foreignness" across genres and registers.

It is hardly surprising that the main findings follow what has been established in earlier studies. The prevalence of Latin and French and the link of multilingual practices to the author's university education and profession, as well as to the intended readership of the text all play a role in the big picture of when and how multilingual practices are activated in the conscious or subconscious selection of resources from a writer's repertoire. Consequently, we argue that the three hypotheses set up in Section 2.2 are confirmed: an author's education has been shown to play a major role in the frequency of multilingual practices, texts directly associated with foreign countries likewise features highest frequencies of foreign content, and gender and social class cannot be adequately examined due to the skewed nature of the present dataset. In addition to these findings, a key observation concerns the complexity of code-switching as a phenomenon: specific foreign languages and switch types are used at different frequencies depending on the macro genre and the author's sociolinguistic background variables. However, it is important to note the numerous co-variances within the metadata, particularly between the strongest predictor variables listed above. This does not mean that the individual factors would not predict the use of multilingual resources, but we need to be careful to note that the dynamics are more complex than they may appear. These findings clearly show that a monofactorial analysis needs to be explored further with more sophisticated multifactorial approaches (see e.g. Tyrkkö and Nurmi 2017).

Questions still requiring answers include the language proficiency of the reading public, which is difficult to chart in variables that are easy to describe consistently. Tyrkkö, Nurmi, and Tuominen (forthcoming 2017) describes some attempts in this direction, but further means of connecting the intended and actual reading public to the texts are needed. Another major question is the changing global context. In our data, the expanding number of languages over the timeline and the inclusion of exotic languages echoes of the building of the British Empire and the exploration of the world in the spirit of the Enlightenment, but approaching these wider questions would require a more specific dataset, connected to those particular endeavours in more intimate detail. At the same time, it should be noted that the cultural context of even the home environment of the writers and readers of the texts we have studied here keeps changing. Processes such as democratisation, mediatisation and secularisation bring changes not only to languages deemed important to study,

but to access to education. What was a correct way of writing kept changing, including the foreign prestige languages, and the increasingly important role of English as a language of significance both at home and abroad should not be overlooked when considering these political, social and philosophical shifts.

References

Chambers, J. K. 2003. *Sociolinguistic theory: Linguistic variation and its social significance*. Malden: Blackwell.

CLMET3 = *Corpus of Late Modern English Texts* 3.0. https://perswww.kuleuven.be/~u0044428/.

Davidson, Mary Catherine. 2003. Code-switching and authority in late medieval England. *Neophilologus* 87. 473–486.

Davidson, Mary Catherine 2005. Discourse features of code-switching in legal reports in late medieval England. In Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen & Brita Wårvik (eds.), *Opening windows on texts and discourses of the past*, 343–351. Amsterdam & Philadelphia: John Benjamins.

Diller, Hans-Jürgen, Hendrik De Smet & Jukka Tyrkkö. 2011. A European database of descriptors of English electronic texts. *The European English Messenger* 19. 21–35.

Halmari, Helena & Robert Adams. 2002. On the grammar and rhetoric of language mixing in Piers Plowman. *Neuphilologische Mitteilungen* 103. 33–50.

Helsinki Corpus. http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/.

Historical Thesaurus of English. http://historicalthesaurus.arts.gla.ac.uk/.

Hunt, Tony. 2000. Code-switching in medical texts. In D. A. Trotter (ed.), *Multilingualism in Later Medieval Britain*, 131–147. Cambridge: D.S. Brewer.

Iglesias-Rábade, Luis. 1996. The multi-lingual pulpit in England (1100–1500). *Neophilologus* 80. 479–492.

Labov, William. 1994. *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change. Volume 2: Social factors*. Oxford: Blackwell.

Machan, Tim William. 1994. Language contact in Piers Plowman. *Speculum* 69. 359–385.

Mäkinen, Ilkka. 2013. Why people read: Jean-Jacques Rousseau on the love of reading. In Dorrit Gustafsson & Kristina Linnovaara (eds.), *Essays on libraries, cultural heritage and freedom of information*, 127–136. Helsinki: National Library of Finland.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Pearson Education.

Nurmi, Arja & Päivi Pahta. 2004. Social stratification and patterns of code-switching in early English letters. *Multilingua* 23. 417–456.

Nurmi, Arja & Päivi Pahta. 2010. Preacher, scholar, brother, friend: Social roles and code-switching in the writings of Thomas Twining. In Päivi Pahta, Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *Social roles and language practices in Late Modern English*, 135–161. Amsterdam & Philadelphia: John Benjamins.

Nurmi, Arja & Päivi Pahta. 2011. Multilingual practices in women's English correspondence 1400–1800. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse*, 44–67. New York & London: Routledge.

Nurmi, Arja & Päivi Pahta. 2013. Multilingual practices in the language of the law: Evidence from the Lampeter corpus. In Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds.), *Ex philologia lux: Essays in honour of Leena Kahlas-Tarkka*, 187–205. Helsinki: Société Néophilologique.

Oxford Dictionary of National Biography. http://www.oxforddnb.com/.

Pahta, Päivi. 2003. On structures of code-switching in medical texts from medieval England. *Neuphilologische Mitteilungen* 104. 197–210.

Pahta, Päivi. 2004a. Code-switching in medieval medical writing. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical and scientific writing in Late Medieval English*, 73–99. Cambridge: Cambridge University Press.

Pahta, Päivi. 2004b. 'So seiß idem comentator': Code-switching and organisation of knowledge in John Trevisa's translation of De proprietatibus rerum. In Alicia Rodríguez Álvarez & Francisco Alonso Almeida (eds.), *Voices on the past: Studies in Old and Middle English language and literature*, 35–48. A Coruña: NetBiblo.

Pahta, Päivi & Arja Nurmi. 2006. Code-switching in the Helsinki Corpus: A thousand years of multilingual practices. In Nikolaus Ritt, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky, *Medieval English and its heritage*, 203–220. Frankfurt: Peter Lang.

Pahta, Päivi & Arja Nurmi. 2009. Negotiating interpersonal identities in writing: Code-switching practices in Charles Burney's correspondence. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)* , 27–52. Amsterdam & Philadelphia: John Benjamins.

Pahta, Päivi & Arja Nurmi. 2011. Multilingual discourse in the domain of religion in medieval and early modern England: A corpus approach to research on historical code-switching. In Herbert Schendl & Laura Wright (eds.), *Code-switching in Early English*, 219–251. Berlin & New York: Mouton de Gruyter.

Rothwell, William. 2000. Aspects of lexical and morphosyntactical mixing in the languages of medieval England. In In D. A. Trotter (ed.), *Multilingualism in Later Medieval Britain*, 213–232. Cambridge: D.S. Brewer.

Schendl, Herbert. 2000. Syntactic constraints on code-switching in medieval texts. In Irma Taavitsainen, Terttu Nevalainen, Päivi Pahta & Matti Rissanen (eds.), *Placing Middle English in context*, 67–86. Berlin & New York: Mouton de Gruyter.

Tyrkkö, Jukka & Arja Nurmi. 2017. Analysing multilingual practices in Late Modern English: Parameter selection and recursive partitioning in focus. In Sebastian Hoffmann, Andrea Sand & Sabine Arndt-Lappe (eds.), *Exploring recent diachrony: Corpus studies of lexicogrammar and language practices in Late Modern* English (Studies in Variation, Contacts and Change in English 18). Helsinki: Varieng. http://www.helsinki.fi/varieng/series/volumes/18/tyrkko_nurmi (accessed 8 June 2017).

Tyrkkö, Jukka, Arja Nurmi & Jukka Tuominen. Forthcoming 2017. Semi-automatic discovery of multilingual elements in English historical corpora: Methods and challenges. In Arja Nurmi, Tanja Rütten & Päivi Pahta (eds.), *Challenging the myth of monolingual corpora*. Leiden: Brill.

Voigts, Linda Ehrsam. 1996. What's the word? Bilingualism in late-medieval England. *Speculum* 71. 813–826.

Wenzel, Siegfrid. 1994. *Macaronic sermons: Bilingualism and preaching in late-medieval England*. Ann Arbor, MI: The University of Michigan Press.

Wright, Laura. 1992. Macaronic writing in a London archive, 1380–1480. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New methods and interpretations in historical linguistics*, 762–770. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 1995. A hypothesis on the structure of macaronic business writing. In Jacek Fisiak (ed.), *Medieval dialectology*, 309–321. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 1998. Mixed-language business writing: Five hundred years of codeswitching. In Ernst Håkon Jahr (ed.), *Language change: Advances in historical sociolinguistics*, 99–118. Berlin & New York: Mouton de Gruyter.

Wright, Laura. 2000. Bills, accounts, inventories: everyday trilingual activities in the business world of later medieval England. In D. A. Trotter (ed.), *Multilingualism in Later Medieval Britain*, 149–156. Cambridge: D.S. Brewer.

Wright, Laura. 2001. Models of language mixing: Code-switching versus semicommunication in medieval Latin and Middle English accounts. In Dieter Kastovsky & Arthur Mettinger (eds.), *Language contact in the history of English*, 363–376. Frankfurt am Main: Peter Lang.