

Revisiting the Donation of Constantine

Francesca Frontini¹ and Gerard Lynch² and Carl Vogel²

Abstract. Techniques developed for synchronic text classification problems are applied to a significantly diachronic dataset. The scale of the temporal categories appears to matter. The problem addressed is that of using automated text classification methods to temporally locate *The Donation of Constantine*. The results reported do not contradict the analysis of Lorenzo Valla from 1440, claiming the document a forgery, but suggest that it is a very good forgery. This contributes to establishing the validity of these classification methods as applied to temporal categories and small datasets.

1 BACKGROUND

Some current work in computational linguistics returns to classical problems in historical linguistics [17, 7, 16, 18]. We apply text classification techniques using chronologically determined categories for a corpus of Latin texts. We report on experiments using letter bigram distributions as an index to Latin morphology and language change in Latin over time from extensive use of explicit case marking and free word order towards a period of scant variation in constituent order in written Latin. We focus on *The Donation of Constantine* (hereafter, the DOC) [2, 5].

By this name is understood, since the end of the Middle Ages, the alleged testament of Emperor Constantine the Great (272-373 AD), which is constructed as a bequest to the papacy, and used as a land claim by that institution. This document is without doubt a forgery, fabricated somewhere between the years 750 and 850. Its authenticity has been questioned all through the Middle Ages, but it wasn't until the XV century that its falsity was known and demonstrated. In 1433, in his *De Concordantia Catholica*, Nicholas of Cusa judged it as apocryphal. Some years later (1440) Lorenzo Valla (*De falso credita et ementita Constantini donatione declamatio*, Mainz, 1518 [5]) proved the forgery through the analysis of linguistic, stylistic and content anachronisms. Its authenticity was yet occasionally defended until Baronius, Cardinal and ecclesiastical historian, admitted in his *Annales Ecclesiastici* that the DOC was a forgery, whereafter it was soon universally admitted to be such [10, 12].

Although there appears to be a consensus that the text is a forgery, the document remains a focus of research into dating its language. For example, [6] is a recent work that examines the provenance of a phrase that appears in the document (“*urbis Romae episcopo et pape*”), whether this is best translated as “the bishop of the city of Rome and Pope” or “the bishop and pope of the city of Rome”) due to the relative frequency of this description of the Pope in the 8th century and the resulting pragmatic force.

We revisit the question of dating the DOC using solely features of the text. The extent of text-external features is *a priori* classification

of textual categories by the period of composition. The first analysis classifies the text using categories individuated by authorship (about 248 alternatives; plus 44 texts that are from anonymous sources), and the second considers the text with respect to seven temporal periods. The first analysis is interesting in revealing a 4th century historian as the source of text most similar to the questioned document. This does not contradict Valla, but suggests that as a forgery it potentially had a particular source from the target period as an exemplar. The second set of experiments divide the texts into broad temporal periods, extending to the contemporary period. This supplies additional support for claims of validity of the classification methods by suggesting that the text does not pattern well with Latin texts from 1400 to 1650, and less well with texts composed in the period from then until today.

The paper briefly outlines the methods which are tested here (§2) and describes the composition of the corpus (§3). In section §4, §5 and §6 we detail the experiments and demonstrate the outcomes. We conclude by indicating a range of other sorts of studies that we intend to explore specifically with respect to DOC, but more importantly in examining language change in Latin.

2 METHODS

For purposes of authorship attribution in forensic contexts, it has been suggested that letter unigram distribution analysis provides an anchor into the most reliable classification methods [3, 4]. The intuition behind reliability is the replicability of this level of tokenization: one may elect to count spaces, numbers, punctuation marks or not, but having made that decision there is no dispute about what counts as a letter. This is the opposite end of the spectrum from general linguistic “habits” that may figure into some methods of analysis [8]. Even hand tagging of part of speech gives rise to the need to assess inter-rater reliability. The more abstract the linguistic feature employed, the less agreement there will be about its application in individuating instances. This is in no small part because of gradience within linguistic categories [1].

One can conduct attribution studies with text-internal methods, external methods, or some hybrid [9]. External methods include reasoning about contemporaries and the descriptive content of the text, such as pointing out anachronisms. Here, we focus on the form of words in the texts, not the semantics.

The methods applied to assessing these distributions derive from suggestions about word-level tokenization from [14, 13] in connection with comparing corpora and, in that process, assessing corpus-internal homogeneity. The work here begins by collecting and individuating files of text. The input to the methods is an index of files and their *a priori* category. For the experiments reported in this paper, we balanced the file sizes at about 4KB of lines (using the unix split command). For these experiments, the natural category for a file

¹ Dipartimento di Linguistica, Università degli Studi di Pavia, Italy

² Computational Linguistics Group, Intelligent Systems Lab, Trinity College Dublin, Dublin 2, Ireland {vogel,gplynch}@tcd.ie

is the name of its author³ or the period of Latin during which it was composed. We used letter bigram tokenization and record frequency distributions of letter bigrams in each file. We have used letter bigrams in order to capture the most productive unit of grammatical morphology encoding in Latin; however, for general classification letter unigrams have been proven quite useful [21, 11]. Pairwise similarity indices are constructed for all of the files using the cumulative chi-square value for each of the bigrams. The chi-square value is based on observed occurrences relativized to the file size (the normal computation of chi-square using observed values and expected values); this statistic is summed for each bigram that appears in either of the two files, and then divided by the number of degrees of freedom (essentially, the number of comparisons). In a divergence from the normal use of the chi-square test, we are not using the value to conclude that distributions are distinct or to reject a null hypothesis that they are random selections from the same population [15], but in order to rank their similarity. Thus, neither do we exclude from analyses comparisons with less observed frequencies less than the value five, as one ordinarily does when applying the chi-square as an adversarial test and not as a measure of similarity.

The rank ordering of similarity scores for each pair of files is then input to the Mann-Whitney rank ordering test to examine the significance of categories of files. For each *a priori* category, and each file within that category, we determine which categories the file fits best (the rank order of comparisons with the questioned file and the other files for each possible category is what one reasons with using the Mann-Whitney). Thus, each of the files within a category may or may not come up as significantly well suited to its *a priori* category and also with respect to other categories, simultaneously [20]. For the experiments reported here we are very generous in considering even $p < 0.25$ significance attached to similarity of a file with respect to a category as a relevant level of similarity. We use an additional level of significance testing for assessing category homogeneity—this is in thinking of a testing the fairness of c -sided coins, where c is the number of categories and the number of tosses is determined by the number of files in a category; thus, Bernoulli schema let us decide whether sufficiently many files are attributed to a category to deem the category homogeneous [19]. As sometimes the best fit category for one or more files of an *a priori* category is instead some other category, it is useful to also consider those alternative assignments.

3 CORPUS

The Latin Diachronic Corpus was sourced from the Internet. Latin sections are available on several text repositories on the web, such as Project Gutenberg⁴ and Bibliotheca Latina at IntraText.⁵ More specific sources are: www.thelatinlibrary.com and www.documentacatholicaomnia.eu. The primary source for the texts used here has actually been www.thelatinlibrary.com because it appears to be the most complete and with the widest diachronic span. Further, all of the texts are in the same html format. The documentation for the website indicates that

“Many were originally scanned and formatted from texts in the Public Domain. Others have been downloaded from various sites on the Internet (many of which have long since disap-

peared). Most of the recent texts have been submitted by contributors around the world.”⁶

The whole site has been downloaded using the program wget.

Pages not holding Latin texts were discarded. The html pages were automatically transformed into plain text:

- no accented or special characters
- no arabic numbers (to get rid of page numbers and other indexing numbers)
- no punctuation (which is spurious with respect to authorship in Latin, being added by the editor of the phylogenic edition)
- substitution of U → V and u → v;

These changes are all intended to normalize the texts in a principled way. In modern Latin transcriptions, graphemes <u> and <v> were used to express the allophones of the unique old Latin phoneme /u/, which had an approximant pronunciation (such as in English <w>) and a plain vocalic one. Standards of transcription differ from text to text. This change prevents spurious heterogeneity. Making the substitution does not lose information, since the pronunciation is reconstructable from phonemic context.

The files downloaded were renamed to indicate the age, authorship and work associated with each file. The temporal periods were:

- | | |
|-----------------------------------|---------------------------------|
| 1. archaic age: | early Latin text until 100 B.C. |
| 2. classical age: | 100 A.D. - 250/300 A.D |
| 3. late imperial Latin: | 300 - 600 |
| 4. early middle ages: | 600 - 1000 |
| 5. high middle ages: | 1000 - 1400 |
| 6. humanists: | 1400-1650 |
| 7. modern and contemporary Latin: | 1700 - today |

The division between period 1 and 2 is quite clear from the point of view of language change. Whereas the first period is characterized by a higher degree of instability in the morphosyntactic features, as well as by archaisms in the lexicon, period 2 witnessed the creation and establishment of a standardized literary language whose rules were explicitly codified by grammarians and strictly followed.

Probably the most problematic division is the one between classical age (2) and late empire (3). From the literary and artistic point of view it is certainly objectionable to put classical authors such as Cicero, Virgil and Ovid (*latinitas aurea*), as we have done, in the same category with later authors (the so called *latinitas argentea*),⁷ but from the point of view of linguistic change, especially of a highly conventionalized written language, it seemed plausible to create a broader category. The idea was to separate the mostly pagan classic and imperial Latin from the Christian literature which flourished in the late empire. The limit was set to 313 A.D., the year in which the Edict of Milan was issued by emperors Constantine and Licinius, proclaiming religious tolerance in the Empire, which set the path for Christianity to become a state religion.

Period 4 is characterized by the decay of Latin as spoken language, and by the emergence of Romance Languages. From the point of view of the literary language though, Latin established itself as the official language both of Church and Empire. Study of Latin revived especially in the IX century under Charlemagne (the so called “Carolingian Renaissance”). The boundary between 4 and 5 is not only

³ The name of the author and particular work would also have been reasonable.

⁴ www.gutenberg.org – Last Verified, January 14, 2007.

⁵ www.intratext.com/LATINA/ – Last Verified, January 14, 2007.

⁶ See <http://www.thelatinlibrary.com/about.html> – Last verified, January 13, 2007.

⁷ The *latinitas aurea* is latin for “the Golden Age of Latin” (75 B.C. to 14 A.D.); the “Silver Age of Latin” refers to the later period from from 14 A.D. to 200 A.D.

motivated by historical reasons—the new millennium being a conventional turning point from the early to the high Middle Ages—but also by linguistic ones. This period saw the coming of age of the new national languages used (both in the Romance as well as in the Germanic area) for the first time in the written medium as well. With the birth of a literature in the modern languages, most of the authors belonging to this period became not only bilingual in speech (as already in period 4) but also in their written production (e.g. Dante).

As for period 6, it is known that Humanism is a movement, rather than a period, characterized by the rediscovery of the classics and the adherence to the Ciceronian style in the production; it spread from Italy and its influence reached the different parts of Europe in different moments between 1400 and 1650. In this category, we include authors who might show distinctive humanistic (that is classical) features in their style, thus avoiding the influence of the medieval category and upon the last category, which is composed by a series of very heterogeneous texts, dominated by the scientific Latin of the 18th, 19th century, as well as by the Latin of the church.

In this classification, the purported period for the Donation of Constantine is category 3 for late imperial Latin, and the accepted actual period of the text is category 4 of the early Middle Ages. However, finding similarity to the classical period is clearly relevant. No effort is made to balance the corpus by genre.

4 EXPERIMENT 1

Because the text of DOC amounts to only about 20K bytes, we individuated all of the files into approximately 4K chunks; this meant splitting DOC into six files. Removing files that were too small (less than 1K) left 19,963 files to be classified by author. In any run we examined five files of each category, 79 categories. Sources with too few files of sufficient size were eliminated in arriving at the 79 categories. This meant that we had balanced files by file size and categories for their number of constituent files. We analyze the aggregate results across twenty such random samplings. The homogeneity index for a category is the average number of times that files from the category fit best with the category itself.

Some of the more homogeneous categories are listed in Table 1. To consider significance for this one wonders how many tosses out of five coming up to the same side of a 79-sided die, in repeated experiments, it would take to conclude that the die is not fair: three is rather significant ($p < 0.005$). If the die is not fair, it is safe to reject the hypothesis that the files are clustering with their category because of random chance and accept that it is for reasons of meaningful similarity among the files. The rank ordering is based on similarity, after all. Note a preponderance of 4th century sources among the most homogeneous listed in Table 1. This might lead one to expect this temporal category to be the most heterogeneous. This is not the case, as is shown in the next section.

Thus, clearly the DOC is self-homogeneous as a document to have so high an index (2.737) after so many runs when being considered with respect to over 300 sources. However, recall from §2 that it is useful to examine the categories which are best fits for files of some category when their *a priori* category is not the best fit. In the case of the DOC, it finds Albert of Aix, Addison, Cassiodorus, Gregory of Tours and William of Tyre one time each in total across the 20 samplings. However, the remarkable thing is that the texts find Ammianus as a most similar alternative source 13 times. Note that Ammianus is particularly homogeneous in these experiments. This is striking given that the category is divided into 248 files, in contrast with the six files of the DOC. With regard to the dating of the document,

Table 1. Most homogeneous author categories

Source	Approximate Period	Homogeneity
Addison	19th C AD	5.0
Albertanus	13th C AD	2.526
AlbertofAix	12th C AD	3.738
Ammianus	4th C AD	4.632
Apicius	4th C AD	4.316
Arnobius	4th C AD	2.842
Bultelius	16th C AD	4.368
Claudian	4th C AD	3.211
Commodianus	3rd C AD	4.632
DOC	8th C AD	2.737
Descartes	17th C AD	4.053
Gestafrancorum	12th C AD	4.053
Juvenal	late 1 early 2nd C AD	3.263
Kempis	15th C AD	3.789
Kepler	16-17th C AD	2.895
Walter	12th C AD	2.842

while this does not refute accepted claims that the text was composed during the early middle ages, it is very interesting that the most reliable alternative source dates to the late imperial period, in which the text was originally claimed to have been written. This could perhaps point to Ammianus a source of language directly influencing the forgery.

5 EXPERIMENT 2

We also wished to consider the text from the point of view of *a priori* temporal categories. We used the seven categories described in §3. Naturally, the category corresponding to the DOC is skewed in size in comparison to the other categories, but it is still useful to examine its texts in this way.

5.1 Temporal Categories with the *Donation* Isolated

As in the first experiment we took samples of five files from each category, however we ran 1000 such experiments. Table 2 shows the results. It is unsurprising that the DOC is the most homogeneous since it is a single text split into six constituent files from which five are chosen in any one experiment, while the other categories have many more from among which to choose five.

Table 2. Homogeneity of Temporal Categories

Period	Homogeneity
1archaic	1.569
2classic	2.247
3late	0.452
4earlyMA	1.252
5highMA	0.771
6humanism	1.018
7modern	0.779
donation	4.99

It is again very useful to consider the alternative assignments. First of all, Table 3 demonstrates that the best alternative category for the category defined by the DOC, out of 1000 runs, finds the classical period seven times (the period close to the claimed time of writing) and the high middle ages, twice. We emphasize that both are extremely small figures: in 1000 experiments involving five files each, only nine

occasions was a DOC file not most similar to the DOC as a category.

Table 3. Alternative fits for the Donation of Constantine

Period	Fits out of 1000
1archaic	0
2classic	7
3late	0
4earlyma	0
5highma	2
6humanism	0
7modern	0

5.2 Temporal Categories Including the *Donation*

The homogeneity values for the temporal categories in Table 2 suggest a replication of that experiment with just seven categories: once with the files of the DOC recorded in category 4, the high Middle ages where consensus dates the document; and once with the files recorded with category 3. With repeated sampling, again 1000 experiments, and these texts recorded with a larger category, such that the DOC will not figure into the random selection of each and every experiment (as the construction from this section forces). Significance for this construction is as follows: eight categories and five files within a category are selected; for three to be assigned to its *a priori* category is only approaching significance ($p < 0.15$), but four is statistically significant ($p < 0.01$). Thus, only the Donation as a category is homogeneous. We do not expect any great changes to the values recorded for the proper temporal categories in Table 2 for either of these permutations.

Consider the case in which the files of the DOC are categorized with those of the late imperial period. Table 4 shows that the resulting homogeneity of the revised categories is not significantly changed (with seven categories and five choices, here, four out of five is significant $p < 0.05$). Only the classical period approaches significance. Both the late imperial period and the early Middle ages increase in homogeneity when the texts of the Donation are considered part of the late imperial period. The same is true when the texts are added to the category of the early Middle Ages (see Table 5). However, here the effect is much smaller. Further, both the late imperial period and the early Middle Ages form more homogeneous categories when the files of the donation are added to the late imperial period than when they are added to the category of the Middle Ages. This suggests greater compatibility with the earlier of the two periods.

Table 4. Temporal Homogeneity with the Donation in 3late

Period	Homogeneity
1archaic	1.657
2classic	2.218
3late	0.520
4earlyMA	1.349
5highMA	0.917
6humanism	1.184
7modern	0.814

Then it makes sense to consider the succession of tables 6-12. Initially we discuss the results in the first column of alternative fits (“Exp 2.1”); the other two columns correspond to the repeated 1000

Table 5. Temporal Homogeneity with the Donation in 4earlyMA

Period	Homogeneity
1archaic	1.681
2classic	2.254
3late	0.495
4earlyMA	1.282
5highMA	0.94
6humanism	1.05
7modern	0.811

experiments with the files of the DOC in 3late (“D3”), and then the 1000 experiments with those files in the category 4earlyMA (“D4”), instead. As with Table 3 the entries are on the same scale: an entry like 453 “Fits out of 1000” experiments means that with 5000 selections of files from the period given by the table, 453 were alternatively assigned to the category provided by the row as a better fit than the four other files in the *a priori* selection of five files for that experiment. The bottom row in each of the tables provides the total sum out of the experiments out of 1000 samples of five files per category in which a file in the category that the table records (stated in its caption) some find a category given by the earlier rows as a better alternative fit than the category in the caption.

First notice that with the DOC as a separate category, in fact, the classical period is the best alternative for each of the temporal categories. This speaks to the influence of that period on the language, and apart from the immediately preceding and following periods its best alternatives distribute more or less evenly, although with greatest overlap with humanism. This is consistent with Latin being a dead language by the time of the 6th century; from that period Latin was learned from written texts rather than ambient spoken language. During the humanism period, classical texts were consciously sought and rediscovered, influencing further the Latin in use during the humanism period. The pattern of similarity of periods of Latin to earlier and later periods is consistent with accepted philological thought. Table 6 shows that the archaic period finds as its best alternative the immediately following classical period in nearly $\frac{1}{10th}$ of the tests.

The viability of the DOC as an alternative temporal category peaks with the High Middle Ages (Table 10). This analysis supplies no evidence supporting the claim that the DOC belongs to the late imperial period. Support for assignment to the late imperial period is roughly equal to that of the early middle ages.

Table 6. Alternative fits for the Archaic Period

Period	Fits out of 1000		
	Exp 2.1	D3	D4
2classic	483	453	468
3late	15	15	27
4earlyma	92	92	80
5highma	48	53	67
6humanism	199	211	175
7modern	90	104	107
Donation	4	n.a.	n.a.
Total Alternatives:	931	928	924

Inspection of the D3 and D4 columns reveals no change in the trends observed for the Exp 2.1 column of each of the tables. The differences that do exist are most interesting for the results reported in Table 7, Table 8, Table 9, and Table 11. Where differences of interest occur, it is because the location of the DOC texts within either cate-

Table 7. Alternative fits for the Classical Period

Period	Fits out of 1000		
	Exp 2.1	D3	D4
1archaic	2	2	4
3late	48	44	62
4earlyma	199	223	181
5highma	104	101	116
6humanism	299	331	304
7modern	186	181	195
Donation	31	n.a.	n.a.
Total Alternatives:	869	882	862

Table 8. Alternative fits for the Late Imperial Period

Period	Fits out of 1000		
	Exp 2.1	D3	D4
1archaic	0	0	0
2classic	474	483	505
4earlyma	145	148	141
5highma	117	113	130
6humanism	116	150	121
7modern	103	102	101
Donation	43	n.a.	n.a.
Total Alternatives:	998	996	998

Table 9. Alternative fits for the Early Middle Ages

Period	Fits out of 1000		
	Exp 2.1	D3	D4
1archaic	1	2	2
2classic	487	483	476
3late	40	34	42
5highma	139	146	184
6humanism	133	158	140
7modern	115	129	114
Donation	46	n.a.	n.a.
Total Alternatives:	961	952	958

Table 10. Alternative fits for the High Middle Ages

Period	Fits out of 1000		
	Exp 2.1	D3	D4
1archaic	0	0	0
2classic	437	413	450
3late	37	34	41
4earlyma	209	254	223
6humanism	141	164	142
7modern	95	107	112
Donation	56	n.a.	n.a.
Total Alternatives:	975	972	968

Table 11. Alternative fits for the Humanism Period

Period	Fits out of 1000		
	Exp 2.1	D3	D4
1archaic	1	2	4
2classic	600	592	597
3late	20	30	25
4earlyma	112	134	107
5highma	75	77	99
7modern	148	135	141
Donation	22	n.a.	n.a.
Total Alternatives:	978	970	973

Table 12. Alternative fits for the Modern Period

Period	Fits out of 1000		
	Exp 2.1	D3	D4
1archaic	2	1	0
2classic	560	546	548
3late	24	32	40
4earlyma	123	138	103
5highma	75	77	96
6humanism	192	187	182
Donation	6	n.a.	n.a.
Total Alternatives:	982	981	969

gory 3 or 4 cause those categories, or the bellwether categories of the classical period or humanist periods, to change in overall homogeneity. This is represented in these tables by heterogeneity in increase or decrease in alternative assignments of the relevant categories.

Consider the overall force of these results. Take Experiment 2.1 (the first column in each of Table 6 through 12) as a baseline. In this paragraph, we summarize the trends reflected in the bottom line of each table. Of interest is the set of trends in homogeneity for each of the *a priori* categories when the text of the Donation is folded into either the late imperial period or the early Middle Ages. When the numbers in these tables increase, it is a sign that the category in the caption of the table has increased in heterogeneity. This is because these tables reflect the best alternative fits for the captioned category. Table 6 shows that when the DOC is added to the late imperial period (D3), the archaic period becomes more homogeneous, but more still when it is added to the early Middle Ages (D4). In the case of the classical period (Table 7), the category becomes less homogeneous when the DOC is in the late imperial period, but more so when it is in the early Middle Ages. The late imperial period (Table 8) becomes marginally more homogeneous when the DOC is within it (D3), and remains constant if the DOC is in the early Middle Ages (D4). The early Middle Ages (9) becomes more homogeneous with the DOC in it (D4), but is still more homogeneous with the DOC in the late imperial period (D3). The high Middle Ages (Table 10) is more homogeneous when the DOC is placed in either of the two periods, but slightly more so with it in the early Middle Ages. The humanism period (Table 11) is also more homogeneous with the DOC in either the late imperial period or the early Middle ages than in the baseline, but here, more so with it in the late imperial period (D3). Finally, the modern period (Table 12) is also more homogeneous with the DOC in either of the categories, but much more so with the DOC in the early Middle Ages (D4).

The tables for the late imperial period (Table 8) and the early Middle ages (Table 9) are the most critical. Both the late imperial period and early Middle Ages are more homogeneous when the DOC is

considered as part of the late imperial period, the period in which the DOC claims itself to have been composed. Table 8 shows that the late imperial period finds the early Middle Ages as a best alternative marginally more when the DOC is in the late imperial period than when it is in the early Middle Ages. On the other hand, the early Middle Ages (Table 9) finds the late imperial period as a best alternative slightly less when the DOC is part of the late imperial period, and marginally more when the DOC is part of the early Middle Ages. Thus, there is equivocal support for locating the DOC in either of the two periods—slightly stronger similarity with the late imperial period emerges.

6 EXPERIMENT 3

As a control, we conducted a comparable analysis isolating the text of Apicius, who actually wrote in the 4th century, the period we are using as our 3rd category, Late Imperial Latin (3late). Recall from Experiment 4 that this was one of the relatively homogeneous sources, albeit based on a small number of files (15). The motive for choosing the source is partly that it actually is from exactly the Late Imperial Period in which the Donation was claimed to have been written.

6.1 Temporal Categories with Apicius Isolated

As in isolating the DOC we first consider the text from the point of view of *a priori* temporal categories. We used the seven categories described in §3. Again, we ran 1000 experiments, sampling five files from each of the temporal categories plus Apicius as a category, each time. The DOC was left categorized with the early Middle Ages as per scholarly consensus. Table 13 shows the results. It is unsurprising that Apicius is the most homogeneous since it consists of fifteen files from which five are chosen in any one experiment, while the other categories have many more from among which to select five. There is little change in the significance of homogeneity of the temporal categories with the texts of Apicius isolated.

Table 13. Homogeneity of Temporal Categories

Period	Homogeneity
1archaic	1.636
2classic	2.143
3late	0.464
4earlyMA	1.235
5highMA	0.852
6humanism	1.035
7modern	0.751
Apicius	4.815

Table 14. Alternative fits for Apicius

Period	Fits out of 1000
1archaic	1
2classic	58
3late	1
4earlyma	43
5highma	34
6humanism	23
7modern	24

Table 14 is to be compared with Table 3. Notice that when it is considered on its own as a category, it has considerably more fits with

alternative categories than the DOC. Its most frequent alternative fit is the classical period, and strikingly, its best alternative fit is rarely its natural category (3late), just as for the DOC on both points.

6.2 Temporal Categories Including Apicius

The next tables consider the homogeneity of the seven temporal categories with the DOC where consensus locates it, but with the texts of Apicius varying between the late classical period and the early Middle Ages. The significance values are as in Experiment 2. In Table 15, homogeneity of 3late is lessened with Apicius in that category. Table 16 shows that the homogeneity of 3late increases when Apicius is classified as belonging to the early Middle Ages, and that period decreases.

Table 15. Temporal Homogeneity with the Apicius in 3late

Period	Homogeneity
1archaic	1.624
2classic	2.222
3late	0.473
4earlyMA	1.268
5highMA	0.962
6humanism	1.231
7modern	0.839

Table 16. Temporal Homogeneity with the Apicius in 4earlyMA

Period	Homogeneity
1archaic	1.749
2classic	2.273
3late	0.524
4earlyMA	1.217
5highMA	0.880
6humanism	1.070
7modern	0.852

Tables 17-23 show the for the other seven periods what the frequency of best alternative fits were. The first column represents the best alternatives when Apicius is considered in isolation as a temporal category (note that, unlike the DOC, it is never a best alternative). The second column displays the best alternative frequencies when the texts of Apicius are considered as part of the late Imperial Period, and the third column shows the results when Apicius is treated as part of the early Middle ages.

Table 17. Alternative fits for the Archaic Period

Period	Fits out of 1000		
	Exp 3.1	D3	D4
2classic	476	460	474
3late	28	14	31
4earlyma	90	83	94
5highma	55	54	39
6humanism	179	214	183
7modern	97	98	96
Apicius	0	n.a.	n.a.
Total Alternatives:	925	923	917

Just as Experiment 2 did for the DOC, consider the results of analyzing the temporal periods with the texts of Apicius isolated as

a baseline. The question, for each temporal period, is whether it is more or less homogeneous with the texts of Apicius considered as part of the late imperial period or as part of the early Middle Ages. Again, as before, we first consider the ramifications of bottom line f or each table. The archaic period (Table 17) is marginally more homogeneous with Apicius in the late imperial period and slightly more so with those texts in the early middle ages. The classical period (Table 18) shows greater homogeneity with Apicius in the late imperial period than in the early Middle Ages, but both are improvements over the baseline. The late imperial period (Table 19) is only marginally more homogeneous than the baseline, regardless of which of the two periods the texts are placed in. The early Middle Ages (Table 20) is significantly less homogeneous than the baseline whether the texts of Apicius are in the late imperial period or in the early middle ages. The high Middle Ages (Table 21) are much more homogeneous with Apicius in the late imperial period than in the early Middle Ages, and the same is true of the humanism period (Table 22). The modern period (Table 23) decreases in homogeneity from the baseline, regardless of which period Apicius is placed in.

Table 18. Alternative fits for the Classical Period

Period	Fits out of 1000		
	Exp 3.1	D3	D4
1archaic	1	4	4
3late	50	49	69
4earlyma	206	193	203
5highma	111	121	111
6humanism	326	309	293
7modern	181	178	189
Apicius	0	n.a.	n.a.
Total Alternatives:	875	854	869

Table 19. Alternative fits for the Late Imperial Period

Period	Fits out of 1000		
	Exp 3.1	D3	D4
1archaic	0	0	1
2classic	504	469	529
4earlyma	149	133	137
5highma	116	141	106
6humanism	137	149	114
7modern	92	105	109
Apicius	0	n.a.	n.a.
Total Alternatives:	998	997	996

Table 20. Alternative fits for the Early Middle Ages

Period	Fits out of 1000		
	Exp 3.1	D3	D4
1archaic	0	2	1
2classic	508	479	501
3late	26	28	36
5highma	137	174	148
6humanism	167	172	145
7modern	112	110	134
Apicius	0	n.a.	n.a.
Total Alternatives:	950	965	965

Table 21. Alternative fits for the High Middle Ages

Period	Fits out of 1000		
	Exp 3.1	D3	D4
1archaic	2	0	0
2classic	451	431	432
3late	34	35	52
4earlyma	232	228	247
6humanism	147	163	132
7modern	131	103	112
Apicius	0	n.a.	n.a.
Total Alternatives:	997	960	975

Table 22. Alternative fits for the Humanism Period

Period	Fits out of 1000		
	Exp 3.1	D3	D4
1archaic	1	0	0
2classic	618	584	600
3late	29	27	26
4earlyma	118	102	109
5highma	86	103	78
7modern	122	140	158
Apicius	0	n.a.	n.a.
Total Alternatives:	974	956	971

Table 23. Alternative fits for the Modern Period

Period	Fits out of 1000		
	Exp 3.1	D3	D4
1archaic	4	1	0
2classic	553	541	572
3late	34	20	36
4earlyma	114	116	106
5highma	87	100	82
6humanism	183	202	180
Apicius	0	n.a.	n.a.
Total Alternatives:	975	980	976

As in §5, it is the difference for the late imperial period and the

early Middle Ages that is of most relevance. Within Table 19, one can see that with Apicius in the late imperial period, the number of times that the early Middle Ages as a best fit alternative for the late imperial period as a category decreases (column D3), and fits with the high Middle Ages increase. If the texts are in the early Middle Ages (column D4), then the category for the late imperial period still has fewer best fit alternatives to the early Middle Ages, and also fewer to the high Middle Ages, but more to the classical period. Best fit alternatives for the early Middle Ages (20) as a category also change from the baseline: when Apicius is part of the late imperial period (D3), alternatives for the early Middle Ages decrease with respect to the classical period from the baseline, stay about the same for the late imperial period, and increase for the high Middle Ages. With the Apicius texts in the early Middle Ages (D4), that period finds slightly fewer best alternative matches to the classical period, more alternative matches to the late imperial period and more to the high Middle Ages. These results show that the late classical period is a good fit for the texts of Apicius: the late imperial period finds the early middle ages as a best fit slightly more often when Apicius is within the early Middle Ages than when it is within the late imperial period, and the early Middle Ages finds the late imperial period as a best alternative more when Apicius is placed within the early Middle Ages than when it is in the late imperial period. Nonetheless, it is striking that Apicius, a larger category than the DOC, is never the best alternative for any of the other categories, in any of its 1000 samplings of five files.

7 CONCLUSIONS

First we try to synthesize the results of Experiment 2 and Experiment 3. Both experiments constructed a baseline, by considering the 7 temporal periods and a sub-corpus considered in isolation. In both experiments we add the sub-corpus to the late imperial period (the D3 column in the tables) and to the early Middle Ages (the D4 columns).

Hypothetically, if the sub-corpus is placed in its actual time period, one might expect, relative to the baseline, the category for its time period to have a reduced number of best fits with the competing category, and one expects the competing category to have a reduced number of best fits with the actual category. Moreover, if the sub-corpus is placed in the incorrect time period of the two competitors, then, relative to the baseline, one could expect the actual category to have an increased number of best fits with the incorrect period, and one expects the incorrect category to have an increased number of best fits with the actual category.

Examining a control, if Apicius is part of the late imperial period (and it is), then when it is placed in the late imperial period, it should, relative to the baseline, have a reduced number of best fits for the early Middle Ages (and it does) and one expects the early Middle Ages to have a reduced number of best fits with the late imperial period (which it does not). If Apicius is placed in the early Middle Ages (which is incorrect), then, relative to the baseline, there should be a greater number of best fits for the late imperial period with the early Middle Ages (there are not), and an increased number of best fits for the early Middle Ages with the late Imperial period (which there are). Thus, the control does not actually fit expectations. This points to the inconclusiveness of the sort of stylometric methods we are using.

However, carrying the argument through, if the DOC is part of the late imperial period, then when it is placed in the late imperial period, it should, relative to the baseline, have a reduced number of best fits for the early Middle Ages (and it does not) and one expects

the Early Middle Ages to have a reduced number of best fits with the late imperial period (which it does not). If the DOC is placed in the early Middle Ages, then, relative to the baseline, there should be a greater number of best fits for the late imperial period with the early Middle Ages (there are), and an increased number of best fits for the early Middle Ages with the late Imperial period (which there are).

Thus, both Apicius and the DOC have dissociations from expectations, if we have reasoned expectations correctly, but the dissociations are in different places. On one hand this speaks to the temporal influence of one period upon another, independently, and on the other hand it points out that the study to date is simply incomplete. It suggests that there is as much reason to doubt the temporal attribution of Apicius as there is the Donation of Constantine, or equally that there is no more reason to doubt the claim of the Donation than there is the provenance of Apicius. Apicius was chosen from within the late imperial period because of its size, it is certainly not a random sample—it merits additional investigation in its own right.

The name "Apicius" actually refers to a collection of Roman cookery recipes, usually thought to have been compiled in the late 4th or early 5th century AD and written in a language that is in many ways closer to Vulgar than to Classical Latin. In the earliest printed editions it was given the overall title *De re coquinaria* ("On the Subject of Cooking"), and was attributed to an otherwise unknown "Caelius Apicius", an invention based on the fact that one of the two manuscripts is headed with the words "API CAE". Recall that Table 1 records this sub-corpus among the most homogeneous, and more so than *The Donation of Constantine*, its actual content make it rather distinctive and perhaps not the most ideal on that basis to use as a control in a study like this.

This paper reports on results to date in automatic analysis of corpora constructed around diachronic categories. These experiments on *The Donation of Constantine* show that internal analysis alone is compatible with the possibility that the document is not a forgery. External analysis derives from consensus in the literature that it is a forgery. Thus, it must be a very good one from the point of view of morphological similarities, whether or not they were intended as such. The letter bigram analysis conducted here was intended to discern patterns of Latin morphology, and as a sub-lexical treatment it explicitly abstracts over textual features that one might consciously control. Authors tend to make lexical decisions, not orthographic ones. Most texts are not lipograms. It is worth pursuing a possibility that Ammianus provided the source language input that shaped the forger's concept of fourth-century Latin. There is however, substantial reason to doubt such direct influence. Although Ammianus was Greek and his native language was Greek, he composed *History* in Latin, as the work was intended for Roman readers. The work consisted of 31 books and earned the author a considerable reputation in his day. It maintained at least some of its popularity until the 6th century, but then fell into neglect and is not mentioned during the Middle Ages. His work, given scholarship methods of the time, would not have been natural for an 8th century forger to stumble upon.

To further these experiments we intend to replicate them with letter trigram distributions. We also intend to record results for letter unigrams and word unigrams. It would be appropriate to normalize letters to upper-case. One obvious orthogonal advance would involve expanding the corpus with works without contention in their association to Constantine. This would give the problem a closer semblance to an authorship attribution task than to a temporal location task. A greater amount of text-external reasoning could perhaps be implemented by restricting genre within the sampling.

This is one thread of our ongoing work in text classification with

temporal categories. The corpus itself is useful in providing a source of data with which to test our classification methods with respect to other established claims in the literature, towards settling the validity of the methods that we have been exploring. A goal for the research is to reliably quantify certainty about attributions of texts to categories on the basis of text-internal considerations. This is a necessary exercise for forensic purposes if evidence from linguistic analysis is to be acceptable to courts of justice.

ACKNOWLEDGEMENTS

Science Foundation Ireland (RFP 05/RF/CMS002) has supported this work. We are grateful to Jerom Janssen for useful advice and discussion. We also thank Rodger Kibble, Sarah Raugas and the anonymous reviewers for encouragement and helpful feedback.

REFERENCES

- [1] Bas Aarts, *Syntactic Gradience: The Nature of Grammatical Indeterminacy*, Oxford University Press, 2007.
- [2] Anonymous, *Das Constitutum Constantini (Konstantinische Schenkung). Textausgabe*, Monumenta Germaniae Historica: Fontes iuris Germanici antiqui 10, Hannover, 1968.
- [3] Carole Chaski, 'Linguistic authentication and reliability', in *Proceedings of National Conference on Science and the Law*, pp. 97–148, (1999).
- [4] Carole Chaski, 'Empirical evaluations of language-based author identification techniques', *Forensic Linguistics*, **8**(1), 1–65, (2001).
- [5] Christopher Coleman, *The Treatise Lorenzo Valla on the Donation of Constantine: Text and Translation*, New York: Russell & Russell, 1971. First published 1922.
- [6] M. J. Edwards, 'Constantine's donation to the 'bishop and pope of the city of rome'', *Journal of Theological Studies*, **56**(1), 115–120, (2005).
- [7] Mark Ellison and Simon Kirby, 'Measuring language divergence by intra-lexical comparison', in *Proceedings of the ACL*, pp. 273–280, (2006).
- [8] Jill M. Farrington, *Analysing for Authorship*, Cardiff: University of Wales Press, 1996. With contributions by Morton, A.Q., M.G. Farrington and M.D. Baker.
- [9] Don Foster, *Author Unknown. On the trail of Anonymous*, Macmillan: London, Basingstoke and Oxford, 2001.
- [10] H. Fuhrmann, 'Konstantinische Schenkung', in *Lexikon des Mittelalters. 5: Hiera-Mittel bis Lukanien*, ed., R. H. Bautier, 1385–1387, Artemis, Mnchen and Zrich, (1991).
- [11] Patrick G. T. Healey, Carl Vogel, and Arash Eshghi, 'Group dialects in an online community', in *DECALOG 2007, The 10th Workshop on the Semantics and Pragmatics of Dialogue*, eds., Ron Arnstein and Laure Vieu, pp. 141–147, (2007). Università di Trento (Italy), May 30 – June 1, 2007.
- [12] C.G. Herbermann, E.A. Pace, C.B. Pallen, T.J. Shahan, and J.J. Wynne, *The Catholic Encyclopedia: An International Work of Reference on the Constitution, Doctrine, Discipline, and History of the Catholic Church*, Appleton, 1907.
- [13] A. Kilgarriff and R. Salkie, 'Corpus similarity and homogeneity via word frequency', in *Proceedings of Euralex 96*, (1996).
- [14] Adam Kilgarriff, 'Comparing corpora', *International Journal of Corpus Linguistics*, **6**(1), 97–133, (2001).
- [15] Adam Kilgarriff, 'Language is never, ever, ever random', *Corpus Linguistics and Linguistic Theory*, **1-2**, 263–275, (2005).
- [16] Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tank, and Martin A. Nowak, 'Quantifying the evolutionary dynamics of language', *Nature*, **499**(11), 713–716, (2007).
- [17] John Nerbonne, T. Mark Ellison, and Grzegorz Kondrak, 'Computing and historical phonology', in *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 1–5, Prague, (2007).
- [18] Mark Pagel, Quentin Atkinson, and Andrew Meade, 'Frequency of word-use predicts rates of lexical evolution throughout indo-european history', *Nature*, **499**(11), 717–720, (2007).
- [19] Carl Vogel, 'Corpus homogeneity and bernoulli schema', in *Mining Massive Data Sets for Security*, pp. 93–94, (2007). NATO Advanced Study Institute.
- [20] Carl Vogel, 'N-gram distributions in texts as proxy for textual fingerprints', in *The Fundamentals of Verbal and Non-Verbal Communication and the Biometrical Issue*, eds., Anna Esposito, Eric Keller, M. Marinaro, and Maja Bratanić, 189 – 194, Amsterdam: IOS Press, (2007).
- [21] Carl Vogel and Sandrine Brisset, 'Hearing voices in the poetry of brenndan kennelly', *Belgian Journal of English Language & Literature*, 1–16, (2007).