Brian J. Reilly
The Johns Hopkins University

Moira R. Dillon
Harvard University

Virtuous Circles of Authorship Attribution through Quantitative Analysis

Chrétien de Troyes's Lancelot

Authorship attribution often suffers from charges of vicious circularity, of conflating authorship and style. We argue for a more virtuous circularity for attribution arguments made through the quantitative analyses of stylometry. Our study investigated temporal lexeme usage, which has been suggested as a mark of authorship style, in the corpus traditionally attributed to Chrétien de Troyes. Our analyses revealed different usage patterns for these lexemes after line 6150 of the Lancelot, possibly indicating the signature of another author in the text (traditionally presumed to be that of Godefroy de Lagny). By embracing probabilistic arguments instead of certainties, digital philology can employ stylometry to consider questions of authorship attribution for medieval texts.

Stylometry—the quantification and measure of an author's style—can be deadly business. Traditionally used to investigate the attribution of a text to a particular author, stylometry often causes scandal, as when works by Molière get attributed to Corneille[1]—yet one more way to proclaim the death of an author. The data behind such results tend to be dismissed within the humanities as unrelated to the "true" marks of a writer's style (Love 154–56).[2] Debates over the effectiveness of stylometry in identifying authorship are played out in academic journals where its results are judged against corpora of certain attribution. Still, there is at least one way to know that stylometry is working without formal proof: the rise of adversarial stylometry, i.e., the attempt to use stylometry against itself to preserve anonymity (or to suggest a false identification) (Brennan and Greenstadt). In this digital age with its wars on terror and revolutionary springs, identifying authors (who may be dissidents, terrorists, targets, etc.) has taken on new urgency. There would be no need to invest in the creation of computer code to deceive stylometric programs if those programs did not actually threaten the anonymity of the author. Deadly business indeed.

It would be a mistake to think that these all too modern concerns have no relation to debates over the authorship attribution of medieval texts. Techniques of adversarial stylometry have long been used by authors who only intuitively grasped the metrics involved. Today, for example, we define an "imitation attack" as the attempt "to write a document such that [the] writing style will be recognized as that of another specific author" (Brennan and Greenstadt 60); medievalists might recognize this practice as the school exercise of *imitatio*. We may also be haunted by the possibility that an unknown scribe made changes to a text in such a way that his style would be identified not as his own but as that of the original author, rendering the text we read less the product of a single intention than we should like or than our arguments require. Advances in stylometric analyses have begun to allow us to disentangle such continuators from original authors.[3] The same is true for collaborative efforts. Some Shakespeareans, for example, invoke stylometric techniques (not without controversy) to argue for the authorship attribution of this or that line of dialogue in several plays that Shakespeare wrote collaboratively (Vickers). Might we invest digital philology with the use of stylometry on medieval texts to distinguish among such compositional collaborators as continuator, copyist, jongleur, and—dare we say it—author?

While it is true that the author has long since been declared dead in literary studies, the debate over whether a medieval author can be identified by his (or, less often, her) style has been rekindled with the appearance in 2011 of *Thinking Through Chrétien de Troyes*, a collaborative work by the playfully self-styled "Chrétien Girls."[4] This collective, Stahuljak et al., interprets Chrétien de Troyes "not as an author identified with a historical figure but as a body of texts attributable to this name"(2).[5] There are many ways to proclaim the "death of author," and here we see one of the most reasonable and productive: the historical author (singular) is now a living body of texts (plural). The fruits of such a position are found throughout this study and were already in seed in Kay's 1997 "Who Was Chrétien de Troyes?"

Once the textual plurality of "Chrétien de Troyes" in the title *Thinking Through Chrétien de Troyes* is understood, we can see an implicit rejection of a work like James-Raoul's 2007 *Chrétien de Troyes, la griffe d'un style*, which emphasizes the uniqueness of the author in terms of his style. *The* author is linked to *the* signature of *a* style: *la* griffe d'*un* style. One need not go very far into this thick volume of stylistic analyses to see its relation to the question of authorship attribution. James-Raoul's "Introduction" opens:

> Non pas les procédés, la technique ou la rhétorique, mais le style. Il s'agit de montrer que la griffe « Chrétien de Troyes » caractérise une manière d'écrire faisant sens, particulière, c'est-à-dire reconnaissable en soi ou par comparaison et s'observant dans un ensemble d'œuvres assez nettement pour identifier un auteur.
>
> Neither the practices, the technique, nor the rhetoric, but the style. The attempt is to show that the signature "Chrétien de Troyes" characterizes a meaningful and unique way of writing, which is to say that it is recognizable by itself or by comparison and that it can be seen in a group of works clearly enough to identify an author. (11)

The plurality of texts is unified (*un ensemble*), as is the identifiable author.

With due respect for this truly "thorough and detailed study" (4), Stahuljak et al. nevertheless draw the inevitably sharp contrast between it and their understanding of Chrétien's authorship: "James-Raoul's study helps contextualize the Chrétien corpus in twelfth-century intellectual and literary history, but it also reveals a dangerous circularity between the conceptions of 'authorship' and 'style'" (5). This dangerous circularity—dangerous presumably because vicious—is thought to inhere in the very idea of establishing the attribution of a unified corpus to a single author named Chrétien de Troyes:

> The construction of a unified corpus is grounded in the premise of single authorship, subtended by an implicit claim to oneness: one author—Chrétien de Troyes—to whom was securely attributed a corpus of five romances. The focus on authorized authorship has tended to marginalize other works associated with or attributed to Chrétien—the lyrics, the short text *Philomena*, and *Guillaume d'Angleterre*—and in a circular logic, to draw from the romance corpus an identifying style that either renders other works marginal to the Chretien œuvre or excludes them altogether. (1)

Scholars construct (presumably by tradition or inference) a grouping of texts and attribute them to Chrétien de Troyes; they do not consider texts like *Philomena* or (especially) *Guillaume d'Angleterre* to be within that grouping; they then analyze the attributed corpus to determine the stylistic signature of its author; and finally, they use that signature as a measure by which to judge other texts, e.g., *Philomena* or *Guillaume d'Angleterre*, which are found to be outside the attributable grouping.

Certainly the danger of circularity inheres in any attempt to group texts of uncertain attribution. But is that danger always expressed? Or is it avoidable? And does James-Raoul's study avoid it? In the commentary below, we point the way to a third path. By correcting some of the methodological limitations of James-Raoul's study, we respond to such charges of vicious circularity. There are well-known, objective feedback loops of a more *virtuous circularity* that can update our beliefs about authorship attribution without falling prey to the very real dangers noted by Stahuljak et al. In short, what is needed are quantitative, scientific analyses to accompany quantitative, stylistic data. James-Raoul's otherwise exhaustive account of Chrétien's style usually marshals evidence of a more traditional philological and stylistic register; where it does, it shines. At times, however, this stylistic analysis includes quantitative data, yet fails itself to be genuinely quantitative. In what follows we provide new analyses for a quantitative argument made by James-Raoul about a particular attribution of authorship: that of Godefroy de Lagny's authorship of the ending of the *Lancelot* (also known as the *Chevalier de la Charrette*)—a late twelfth-century romance which otherwise finds itself securely within the body of texts attributed to Chrétien de Troyes.[6]

Quantitative results may, at times, confirm something we already know. So be it: sometimes it is nice to know why we know something. While a future study is planned to perform a fuller stylometric analysis that should complement the traditional methods found in James-Raoul's book,[7] we argue here that this worry over circularity in authorship attribution is trivial: a third way exists and promises new insights into medieval authors and their works. This third way prefers probabilities to certainties, attributing authorship while at the same time "keeping these terms ['authorship' and 'style'] open and potentially plural, admitting the possibility of multiple styles and taking into account the agency of copyists, compilers, and patrons."[8] For when the author's signature (*la griffe*) reappears, it does so through the shifting haze of probability—subject to updated evidence but visible nonetheless.

The *Lancelot* is one of five Arthurian romances (*romans*) usually attributed to Chrétien de Troyes. The others are: *Erec et Enide*; *Cligès*; *Yvain* (also known as the *Chevalier au Lion*); and *Perceval* (also known as the *Conte du Graal*). The reason that one might suspect that a second author wrote the ending to the *Lancelot* is simple, for at the very end of the text, a second author seems to tell us just that: "Godefroiz de Leigni, li clers, / a parfinee LA CHARRETE" ("The clerk Godefroy de Leigni / Has put the final touches on the Knight of the Cart"; lines 7102–03).[9] Critics have long wondered why Chrétien passed the plume to Godefroy. Was he too preoccupied writing another of his romances, the *Yvain*? Was he unhappy with the commission given him by his "dame de Chanpaigne" 'Lady of Champagne' (line 1)? Our entry into the debate here is not over the question *why* but, more basically, *whether*. Did a second author write the ending to Chrétien's *Lancelot*? Can it be shown that he did? Or was this all a ruse? Some critics, following Hult,[10] have argued that Godefroy of Lagny was not an actual, historical person,[11] but merely a *nom de plume* for Chrétien de Troyes.[12]

Stahuljak et al. pay little attention to Godefroy; we could only find one parenthetical reference to him in their study. For James-Raoul, however, the distinction between authors is treated as a given, and the division of labor is said to occur somewhere at lines 6140–53.[13] Assuming Chrétien's stylistic signature can be found, differences in style should be detectable if indeed a second author wrote the last 966 or so lines—even, perhaps, if Godefroy was a skilled imitator, or if Chrétien's *boen gré* ("approval"; line 7106) was, in fact, heavy-handed supervision. Such are the promises of stylometry.

A full test of this possibility is beyond the scope of this essay, in part because it requires digital resources not yet available for study.[14] Moreover, our response here seeks only to intervene in the debate between James-Raoul and Stahuljak et al. We believe that the results of a quantitative analysis of James-Raoul's data, though in conflict with James-Raoul's own non- or quasi-quantitative analyses, actually support the larger hypothesis that a stylistic signature for Chrétien (and Godefroy, if only by complement) may indeed begin to emerge from the numbers.

In our analyses, we use the following abbreviations and total line counts:[15]

E = *Erec et Enide* (40,910 words; 6,878 lines; 3,439 couplets)
C = *Cligès* (40,638 words; 6,664 lines; 3,332 couplets)
L1 = *Lancelot* to line 6149 (37,860 words;[16] 6,146 lines;[17] 3,073 couplets)

L2[18] = *Lancelot* from 6150 to end (6,011 words; 966 lines; 483 couplets)
L = *Lancelot* (43,871 words; 7,112 lines; 3,556 couplets)
Y = *Yvain* (42,111 words; 6,810 lines; 3,405 couplets)
P = *Perceval* (54,192 words; 8,988 lines; 4,494 couplets)

1. Temporal Lexemes (*Chrétien de Troyes, la griffe d'un style*, 505–12)

The chronological progression of a narrative is often indicated implicitly, as in the simple sequence of actions described in the present tense: we know the first action described happened before the next and so on. But chronology may also be indicated explicitly through the use of certain temporal nouns, adverbs, conjunctions, etc. In analyzing explicit chronological indicators in Chrétien's five romances, James-Raoul finds more the mark of Chrétien's training than a signature of his individual style: "Ce souci [de préciser une chronologie] est présent dès le premier roman et ne varie guère: c'est donc le résultat d'un apprentissage préalable, l'empreinte laissée par une école, la marque d'un style" ("This concern [to indicate chronology] is present in the first romance and hardly varies thereafter: it is thus the result of prior learning, the imprint left by a school, the trademark of a style"; 506–07). In a table listing the "diversity of temporal lexemes" (507–08), James-Raoul finds a "stability" in their use across the corpus, with two exceptions.

But before we turn to these two exceptions, we note the striking numbers given by James-Raoul for the lexemes *encore*, *ja*, *jor*, *mais*, and *onque*: these entries, despite occurring with sufficient frequency in L1 to predict occurrences in L2, have no listed occurrences in L2. By collecting our own frequency counts of the temporal lexemes based on the 1989 corrected edition of Ollier's *Lexique et concordance*, we were able to distribute the counts for *encore*, *ja*, *jor*, *mais*, and *onque* across L1 and L2. We also made the following corrections:

ainçois: James-Raoul's distribution is probably based on a miscount. We revise E to 22 and L1 to 32.

derechief: The 19 count given for P is a typo. The count for P is 1, while the total count for all romances is 19, once a misount for Y is corrected.

hui: Ollier's lemma is *ui*.[19] Meaning "today" or, in the expression *hui mes*, 'from now on', one might have expected to find more results than the 4 James-Raoul gives. And indeed, our corrected count yields 86. James-Raoul's count seems to come from using the uncorrected 1986 edition of Ollier's concordance and is doubly wrong. In the 1986 edition, there is a lemma *hui* (adv.) in the concordance (microfiche 1762), but not in the *lexique*. It has 4 occurrences—the same total as James-Raoul's count. Erroneous though it was, that count should have given both of the 2 instances in *Lancelot* to Godefroy, as does the 1986 microfiche concordance: *Lancelot* vv. 6163 and 6198. But the count itself is in error, and the 1989 corrected edition of Ollier's concordance sets things straight. In the 1986 version, the temporal adverb *ui* is split between two lemmas, *hui* and *ui*, which have 4 and 82 entries, for a total of 86. In the 1989 version, those occurrences are all listed under the lemma *ui* (adv.), and the lemma *hui* (sm), absent from the 1986 microfiche concordance, is reinserted in its rightful place.

piece, *pieça*: The lemma *piece* indicates a quantity and can have temporal meanings (e.g., "a long time" or "for a moment") or spatial meanings (e.g., "some distance" or even "pieces"). In the Ollier concordance, there is a lemma *pieça* (adv.; microfiche 2966), all five uses of which have temporal meaning. For the lemma *piece* (sf; microfiches 2967 and, oddly, 2971), there are 71 occurrences divided among four forms given: *pieç* (3), *piece* (44), *piece1* (17), and *pieces* (7). None of the uses of *pieces* are temporal; all of the uses of *pieç* and *piece1* are temporal. Sometimes the translation of *piece* is ambiguous. The occurrences at *E* 1435, *L* 304, 1296, 2283, and *Y* 5790 could mean "a way" or "a while." Even though we prefer a spatial rendering, we have left those cases in the count to better match that of James-Raoul. A line like *L* 5281 specifies "de la voie," and so we do not include it in this group. If our counts still differ, it may be because we reject as nontemporal: *L* 1437, 2275, 4572, and 5281; *Y* 283, 3378, 4219, and 4940.

quant: Ollier's lemma is *cant*. We found a slightly different distribution for L1 (-2 to 174) and L2 (+2 to 31), regardless of the concordance edition used. James-Raoul seems to have only counted those 29 occurrences for L2 on microfiche 463, neglecting to count the 2 on microfiche 465 (v. 6704) and 466 (v. 6509).

From the uncorrected data, James-Raoul makes the following inferences:

(1.1) The number of occurrences of the following eight temporal lexemes is elevated in L compared to the other four romances: *ainçois*, *atant*, *des*, *lors*, *ore*, *orendroit*, *piece*, and *quant*. (509)

(1.2) The use of the lexemes in L2 closely matches that of L1. (509)

The idea here is that Chrétien's project in writing the *Lancelot* involved a particular attention to temporal framings of the narrative, beyond that of his training. If true, we would have an example where a result in quantitative analysis can generate literary criticism: why should the *Lancelot* have an elevated use of temporal lexemes? James-Raoul also claims that this elevated use holds for both Chrétien and Godefroy, generating the speculation that either Chrétien gave Godefroy explicit instructions in this matter or that Godefroy may have intuited the supposedly striking and novel stylistic temporality of the *Lancelot*'s first 6100 plus lines. To come back to our opening discussion, we may translate claim 1.2 as a successful "imitation attack," making us think that Godefroy's temporal-lexeme signature is the same as Chrétien's.

But are those eight lexemes really so elevated when compared to the other romances? There is no consideration of total word counts to tell. And even if there is agreement for these eight lexemes between the two authors, do the other data reveal any differences which might show two signatures present in one text? Let us consider the claims in turn with the requisite quantitative analysis.

(1.1.a) First, to evaluate whether, as a group, the selected eight temporal lexemes were significantly elevated in L as compared to the other romances, we conducted an Independent T-Test with 10,000 bootstrapping samples taken with replacement. Independent T-Tests are employed to compare the means of two groups, calculated using the values obtained at all data points (in our case, the mean percentages of each temporal lexeme). Bootstrapping is used to determine a confidence interval for any test statistic (in our case, the difference between two means) when the distribution of that test statistic is non-normal, i.e., it does not conform to the traditional bell curve. Our null hypothesis states that there is no difference between the mean percentage of this set of temporal lexemes in L and the mean percentage of those same temporal lexemes in all the other romances. If the confidence interval produced by the bootstrapping technique contains zero, then we cannot refute the null hypothesis.

To conduct this first analysis, the frequency of each lexeme was normalized by the total number of words in each romance. The mean normalized frequency of the lexemes in L was compared to their mean normalized frequency in the four other romances. Results revealed that there was no significant difference between the mean normalized frequency of this lexeme subset in L as compared to the other romances (*BCa [-0.2110, 0.0253]*). This analysis addresses whether, overall, this subset of eight lexemes is elevated in L compared to the other romances; however, it does not address the possible discrepancy between the relative frequency of these lexemes in L and the relative frequency of these lexemes in each of the other romances.

(1.1.b) Here we address the relative frequencies of the subset of eight lexemes compared across the romances. We do so, again, by first normalizing the frequency of each lexeme by the total number of words in each romance. Then, Spearman's Rho Correlations (non-parametric) were calculated to indicate the relationship between L and the other romances (E, C, Y, P). Correlations were also calculated for all pairwise combinations of E, C, Y, and P. Because these correlations were computed using the normalized relative frequencies of each of the eight lexemes in the subset, they indicate the amount of similarity between L and the other romances reflected in this lexeme subset. Next, an Independent Samples T-Test was performed, comparing the mean correlation between L and each of the other romances to the mean of the pairwise correlations of the other works. Finally, a bootstrapping procedure with 10,000 replications was carried out to determine the confidence interval for the difference between the means. We found no significant difference between the mean correlation of L to the other works and the mean correlation of all the pairwise combinations of the other works (*BCa [-0.1948, 0.0056]*).

(1.1.c) The methods outlined in 1.1.b were repeated testing L1 and L2 against the other romances, each in a separate analysis. Using this subset of eight lexemes, we found no significant difference between L1 as compared to the other romances (E, C, Y, P; *BCa [-0.2128, 0.0204]*). Further, we found no significant difference between L2 as compared to the other romances (E, C, Y, P; *BCa [-0.0520, 0.3164]*).

The results from 1.1.a-c are difficult to interpret for two reasons. The first reason is the small sample size of the subset of temporal lexemes ($n = 8$). Selecting a small number of lexemes from the entire set of lexemes greatly reduces the statistical power of the analysis and the conclusions that can be drawn from its results. A more serious problem with this particular analysis is that in choosing 8 out of the 36 temporal lexemes, James-Raoul introduces a selection bias. If temporal lexemes, as a set, are descriptive of authorship style (as James-Raoul proposes), then the *entire* data set of temporal lexemes should be subject to analysis, not simply those lexemes in the set that seem to support a given hypothesis. If it had been shown independently that only the lexemes *ainçois*, *atant*, *des*, *lors*, *ore*, *orendroit*, *piece*, and *quant* were valid measures of authorship style, then it would, in fact, make sense to address only these lexemes in the analysis. If, however, there was no reason to believe that any given temporal lexeme was more telling of authorship style than any other such lexeme, then all examples should be considered. James-Raoul gives us no independent reason to believe that the use of this subset of lexemes is a more valid measure of authorship style. Instead, in only offering the insight that these lexemes seem to support the hypothesis and the others do not, James-Raoul has introduced a serious selection bias.

(1.2.a) There were discrepancies between our frequency counts and those of James-Raoul for the entire set of temporal lexemes (as described above). However, these differences did not affect the significance values for any of the following statistical tests, which were performed on the corrected counts.

To determine whether the use of lexemes in L2 closely resembles the use of lexemes in L1, we conducted an analysis similar to the analysis used in 1.1.b-c. Specifically, we first tested the difference in the relative distribution of temporal lexemes between L1 and the other romances (E, C, Y, P). Then we tested the difference in the relative distribution of temporal lexemes between L2 and the other romances (E, C, Y, P). Finally, motivated by our results from these first two steps, we examined the difference in the distributions of temporal lexemes between L2 and E, C, L1, Y, P. Evaluating the entire subset of temporal lexemes provides a larger sample size ($n = 36$) and avoids selection bias.

For each of the tests mentioned above, the frequency of each lexeme was normalized by dividing by the total number of words in each romance (see Fig. 1). Then, Spearman's Rho Correlations (non-parametric) were calculated between the text in question (L1 or L2) and each of the other texts (as indicated above). Correlations were also calculated for all pairwise combinations of the other texts (see Fig. 2). Next, an Independent Samples T-Test was conducted comparing the value of the mean correlation between the text in question and each of the other works to the value of the mean of the pairwise correlations of the other works. Finally, a bootstrapping procedure with 10,000 replications was performed to determine the confidence interval for the difference

	E	C	L1	L2	L	Y	P
ainz	0.00032	0.00027	0.00037	0.00017	0.00034	0.00045	0.00031
ainçois	0.00054	0.00081	0.00085	0.00116	0.00089	0.00055	0.00052
anuit	0.00024	0.00010	0.00021	0.00000	0.00018	0.00012	0.00024
atant	0.00049	0.00025	0.00071	0.00067	0.00071	0.00036	0.00054
aube	0.00010	0.00002	0.00003	0.00000	0.00002	0.00005	0.00007
demain	0.00037	0.00017	0.00018	0.00000	0.00016	0.00057	0.00020
dementre(s) que	0.00000	0.00000	0.00003	0.00000	0.00002	0.00005	0.00002
des (que)	0.00073	0.00030	0.00085	0.00067	0.00082	0.00071	0.00031
derechief	0.00007	0.00002	0.00032	0.00000	0.00027	0.00005	0.00002
encore	0.00095	0.00054	0.00069	0.00067	0.00068	0.00069	0.00074
endemain	0.00015	0.00025	0.00011	0.00000	0.00009	0.00009	0.00013
enjusque	0.00000	0.00000	0.00003	0.00000	0.00002	0.00007	0.00002
enquenuit	0.00005	0.00000	0.00003	0.00000	0.00002	0.00009	0.00004
entretant	0.00000	0.00000	0.00008	0.00000	0.00007	0.00002	0.00000
hier	0.00020	0.00012	0.00011	0.00000	0.00009	0.00002	0.00004
hui	0.00020	0.00007	0.00034	0.00050	0.00036	0.00031	0.00085
ja	0.00235	0.00468	0.00357	0.00416	0.00365	0.00413	0.00345
jor	0.00117	0.00116	0.00092	0.00116	0.00096	0.00121	0.00125
lors (que)	0.00142	0.00130	0.00211	0.00316	0.00226	0.00126	0.00113
mais	0.00565	0.00817	0.00674	0.01248	0.00752	0.00693	0.00491
mienuit	0.00002	0.00000	0.00003	0.00000	0.00002	0.00000	0.00000
none	0.00005	0.00010	0.00013	0.00000	0.00011	0.00002	0.00000
nuit	0.00071	0.00071	0.00050	0.00000	0.00043	0.00069	0.00057
onque	0.00178	0.00153	0.00188	0.00166	0.00185	0.00233	0.00192
oan	0.00002	0.00002	0.00008	0.00000	0.00007	0.00000	0.00000
or	0.00242	0.00249	0.00285	0.00466	0.00310	0.00316	0.00292
ore	0.00027	0.00032	0.00055	0.00100	0.00062	0.00028	0.00024
orendroit	0.00020	0.00005	0.00040	0.00067	0.00043	0.00017	0.00026
piece, pieça	0.00027	0.00020	0.00042	0.00083	0.00048	0.00024	0.00015
prime	0.00005	0.00005	0.00008	0.00000	0.00007	0.00002	0.00002
primes	0.00002	0.00017	0.00016	0.00000	0.00014	0.00007	0.00004
puis (que)	0.00164	0.00057	0.00156	0.00150	0.00155	0.00128	0.00120
quant	0.00381	0.00440	0.00460	0.00516	0.00467	0.00347	0.00345
que que	0.00007	0.00005	0.00008	0.00000	0.00007	0.00019	0.00017
soir	0.00000	0.00005	0.00005	0.00000	0.00005	0.00012	0.00004
uitaves	0.00000	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000

Fig. 1: Normalized frequencies of all temporal lexemes.

	E	C	L1	L2	Y	P
E	1	0.929	0.937	0.854	0.912	0.933
C	0.929	1	0.933	0.827	0.880	0.885
L1	0.937	0.933	1	0.904	0.885	0.894
L2	0.854	0.827	0.904	1	0.844	0.856
Y	0.912	0.880	0.885	0.844	1	0.959
P	0.933	0.885	0.894	0.856	0.959	1

Fig. 2: Spearman Rho correlations between L2 and other works based on all temporal lexemes

between the means. We found no significant difference between the mean correlation of L1 to the other works and the mean correlation of all the pairwise combinations of the other works (*BCa [-0.0421, 0.0448]*). However, we did find a significant difference between the mean correlation of L2 to the other works and the mean correlation of all the pairwise combinations of the other works (*BCa [0.0362, 0.1009]*).

Motivated by this result, we evaluated the normalized relative frequencies of temporal lexemes of L2 against all of the other texts, including L1. Here too we found a significant difference between the mean correlation of L2 to the other works and the mean correlation of all pairwise combinations of the other works (see Fig. 3; *BCa [0.0123, 0.0847]*). This result indicates that L1 might group with the other texts in terms of relative frequencies of temporal lexemes while L2 might not.

(1.2.b) To further evaluate the idea that, in the use of temporal lexemes, E, C, L1, Y, and P form one coherent group of texts, while L2 seems best classified outside that group, we performed a reliability analysis. A reliability analysis provides a measure of the overall parallelism, homogeneity, and consistency among a group of components in a composite.

If we form a composite of E, C, L1, L2, Y, and P based on the assumption that they are homogenous in their use of temporal lexemes, then we can determine a numerical value, which describes how internally consistent that composite is. We used Cronbach's Alpha as our numerical indicator, under the framework of Domain Sampling Theory. Cronbach's Alpha can range from 0 to 1, where a value of 0 indicates that the components of the composite are not at all parallel or homog-

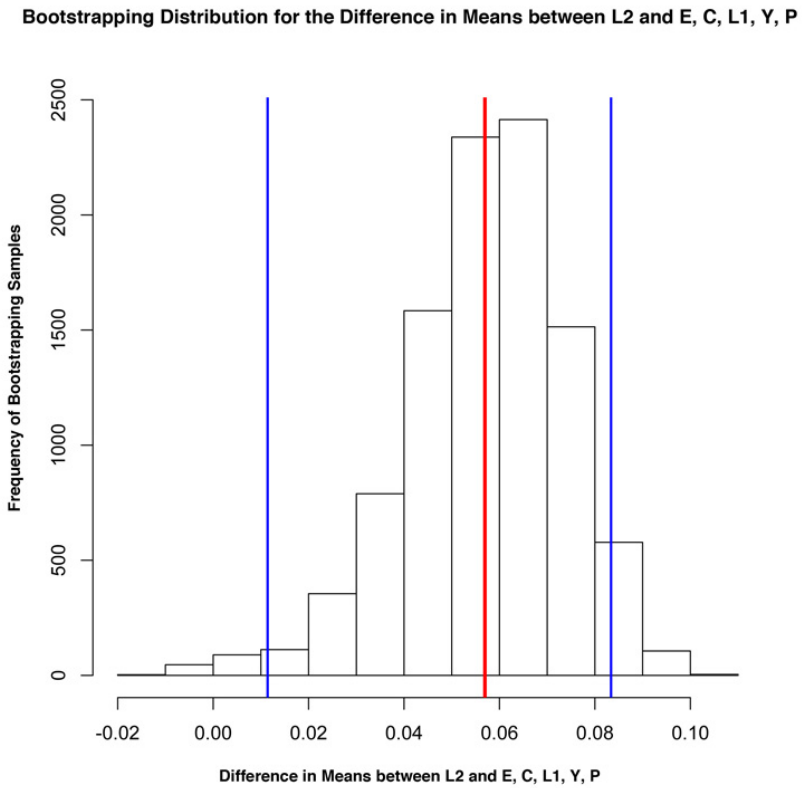

Fig. 3: The bootstrapping distribution for the difference in the mean correlation of L2 to the other texts compared to the mean correlation of all the pairwise combinations of the other texts (E, C, L1, Y, P). The center line indicates the observed value of the difference. The outer lines represent the upper and lower bounds of the BCa confidence interval. Zero lies outside this confidence interval, indicating a significant difference between mean correlation of L2 to the other texts and the mean correlation of all the pairwise combinations of the other texts.

enous and a value of 1 indicates that they are exactly parallel and homogenous. Convention predicates that any value of Cronbach's Alpha above 0.7 is acceptably homogenous. Describing Cronbach's Alpha in the framework of Domain Sampling Theory allows us to infer that the value achieved translates to the degree to which our composite, thus grouped, correlates with any randomly generated composite from the entire domain of composites definable by the components in this domain. The domain being considered here is the domain of temporal lexeme usage in all of the Arthurian romances usually attributed to Chrétien de Troyes. We therefore evaluate how well this composite of works (E, C, L1, L2, Y, and P) would correlate with any random selection of Chrétien's works, whatever they might be, on the basis of temporal lexeme usage.

We obtained a value of 0.980 for Cronbach's Alpha, indicating excellent internal consistency: based on temporal lexeme usage, E, C, Y, L1, L2, and P form a highly homogenous group. However, the question remains whether this group would be more or less internally consistent with the deletion of one of the six texts. We can address this question by calculating Cronbach's Alpha for a new composite when one item is deleted. If the value of Cronbach's Alpha decreases with the deletion of any one item, then the remaining group has become less internally consistent. If, however, the value of Cronbach's alpha increases with the deletion of any one item, then the remaining group has become more internally consistent. Figure 4 shows that the internal consistency of the remaining composite decreases if E, C, L1, Y, or P is deleted; however, the internal consistency of the remaining composite increases if L2 is deleted. These results further point to the possibility that E, C, L1, Y, and P form a cohesive group, while L2 might belong outside this group.

Further work to determine the best grouping scheme for these texts might incorporate factor analyses, which would provide dimensions along which certain works would cluster. Effective factor analysis would require more data collection, including other measures of authorship style. That being said, we have shown on the basis of temporal lexeme usage that E, C, L1, Y, and P might be a more internally consistent group when L2 is excluded.

We still have not shown, however, that temporal lexeme usage in general is a valid measure of style to begin with. Fortunately, James-Raoul, on the basis of expert stylistic analysis, has already made this claim. As a result, our statistical analyses, paired with James-Raoul's argument about the measure's validity, begin to indicate how the grouping of these texts might reflect the signatures of two different authors. Still, only after many critical stylistic markers, independently and theoretically determined, are submitted to statistical analysis, can we begin to build the framework for text classification and author attribution with any confidence.

The other six quantitative arguments concerning Godefroy de Lagny we found in James-Raoul's study had less data to consider. We give a brief description of their content here since, with more data, the analyses could prove informative.

	Cronbach's Alpha if Item Deleted
E	0.977
C	0.972
L1	0.972
L2	0.990
Y	0.972
P	0.978

Fig. 4: Cronbach's Alpha if item deleted

2. Toponyms (*Chrétien de Troyes, la griffe d'un style*, 371–77)

James-Raoul suggests that toponyms provide a useful insight into Chrétien's unique style, for he uses them "d'une manière relativement nouvelle et personnelle" ("in a relatively new and individual manner"; 372). Moreover, if a toponym shows a certain degree of stability across the manuscript tradition, one could assume that the place, whether real or imaginary, was recognizable by the scribe; instability might instead indicate an invented toponym (373).

3. Rhyme Assonances (*Chrétien de Troyes, la griffe d'un style*, 496–500)

Rhyme assonances should be fairly resilient to scribal error, getting us as close as possible to an authorial signature:

> C'est le stylème d'un écrivain et non pas celui d'un simple copiste, parce qu'une rapide comparaison de nos édition de références [CFMA] avec celles qui ont été réalisées sous la direction de M. Zink [*Lettres gothiques*], par exemple, souligne que les trains d'assonances considérés ne subissent pas de modifications dans d'autres copies manuscrites […]

> It is the *styleme* of a writer and not that of a simple amanuensis, since a rapid comparison of our reference editions [CFMA] to those under the direction of M. Zink [*Lettres gothiques*], for example, highlights that the series of assonances considered do not undergo change in other manuscript copies. (497)

It would be useful to do more than a "rapid comparison" to test this point. Moreover, while we agree that James-Raoul may have identified an important mark of style, the data provided only cover the *first* 500 lines of each romance, plus the *last* 500 lines for the end of the *Lancelot*, and only give groupings of rhyme assonances that persist over at least two couplets. A more comprehensive data set should include all assonances for the final 500 lines of the other four romances to ensure that normal intra-textual variation is not mistaken for a distinction between L1 and L2 (although the incomplete state of the *Perceval* makes the utility of its inclusion questionable).[20]

4. *Couplets brisés* (*Chrétien de Troyes, la griffe d'un style*, 518–23)

Although not the inventor of the *couplet brisé* (broken couplet), Chrétien used it extensively. James-Raoul gives us the percentage of broken couplets in the first 1500 lines of E, C, L1, Y, P, and *Guillaume d'Angleterre*, as well as the lines from L2 and *Philomena*. Perhaps because they are "difficult to interpret" (521), no interpretation is given for them. Three interesting arguments are advanced concerning the use of broken couplets, but again the necessary data are not given. James-Raoul considers the instances in which a switch between speakers occurs within a broken couplet and finds it to be a mark of Chrétien's growing maturity as an author. And from that inference from data mentioned but not shown, a case is put forward that: *Philomena* is more like Chrétien's earlier works; *Guillaume d'Angleterre* is more like the later ones; and Godefroy de Lagny uses the broken couplet technique more often (522).

5. *Ainsi* (*Chrétien de Troyes, la griffe d'un style*, 529–35)

James-Raoul's discussion of the connective *ainsi* shows a true mastery of traditional stylistic analyses. The paradoxical nature of this connective reveals a technique that blurs narrative sequences in a process of *tuilage*, "tiling" (529). James-Raoul provides counts for the total occurrences of *ainsi*, as well as for the number of lines begun by *ainsi* (534). From these data, James-Raoul makes the following claims (534–35):

(5.1) Chrétien's debt to the author of the *Roman de Thèbes* is great. That is, the works of Chrétien best approximate, in their use of *ainsi*, the *Roman de Thèbes*.
(5.2) In the use of *ainsi*, E fits with the *Roman de Thèbes*, *Éracle*, and *La prise d'Orange*.
(5.3) From C on (which is to say, in C, L1, Y, and P), Chrétien uses *ainsi* "two to three times" more than his contemporaries (here represented only by *Roman de Thèbes*, *Éracle*, and *La prise d'Orange*).
(5.4) L2 behaves more like E, and so distinguishes itself from the expected use of *ainsi* from 5.3.
(5.5) *Philomena* can be situated between E and C in its use of *ainsi*. (This presumably is an argument for dating *Philomena* within Chrétien's corpus.)
(5.6) *Guillaume d'Angleterre* is most like L1. (Again, presumably an argument for the dating of this romance should it be attributed to Chrétien.)

6. Equivocal Rhymes (*Chrétien de Troyes, la griffe d'un style*, 666–67)

It is claimed that Chrétien uses equivocal rhymes rarely, whereas Godefroy uses them much more often (666–7). A total of 14 such rhymes are said to be found in L2, presumably the number of couplets, which would mean that 3% of Godefroy's couplets employ this type of rhyme. Whether that is "much more" than Chrétien's use cannot be determined since the data for E, C, L1, Y, and P are not given.

7. Homonym Rhymes (*Chrétien de Troyes, la griffe d'un style*, 669–75)

Homonym rhymes are considered to be particularly stable and informative, allowing us to "voir se dessiner les constantes proprement stylistiques d'une écriture, d'un auteur ou d'une œuvre, d'une école ou d'un creuset littéraire" ("see being sketched out the specifically stylistic constants of a writing style, of an author or a work, of a school or literary crucible"; 670). Ignoring ambiguous cases like *or en droit* / *orendroit*, James–Raoul makes the following arguments:

(7.1) The percentages of homonym rhymes are stable across manuscripts. (670)
(7.2) The five Arthurian romances attributed to Chrétien are stable in their use of homonym rhymes, which is quite restrained. (670)
(7.3) L is different from E, C, Y, and P, but only because of L2. (Thus Godefroy's use of homonym rhymes is markedly different from Chrétien's.) (670–1)
(7.4) *Philomena* is like E, C, L1, Y, and P. Indeed, it falls within "the [one?] standard deviation" (*écart-type*) of those five romances. (671)
(7.5) *Guillaume d'Angleterre* is unlike E, C, L1, Y, and P. (Does this then mean beyond one standard deviation from the mean? Or perhaps two?)
(7.6) The differences between manuscripts, though they are stable (7.1), allow us to see the taste of particular scribes for these rhymes. For example, the Berne 134 scribe for P shows the growth in taste for this rhyme by the fourteenth century.

Summary of Statistical Analyses Regarding Authorship Style and Attribution

When a subset of temporal lexemes was selected by James-Raoul without independent theoretical reason, no difference was found between L2 and the other texts. However, in creating this subset, a selection bias was introduced, rendering the analysis inert to any meaningful conclusions. The analysis of temporal lexemes as an entire set has the advantage of (1) having a suitable sample size for good inferential statistics and (2) having theoretical validity as indicative of authorship style. Our statistical analysis of all given temporal lexemes revealed that there was a significant difference between the relative distributions of these lexemes in L2 compared to all other texts. A test of reliability lent further support to the idea that, in the realm of temporal lexemes, L2 might be more appropriately defined as *distinct from* the texts E, C, L1, Y, and P.

Conclusion: Virtuous Circles

Did two different authors write the *Lancelot*? Before we ponder the complexities of this belyingly simple question, let us be clear that our study provides no certain answer to it. But this is because, when faced with such a question, we must move away from certainties and toward probabilities. Our analysis of the use of temporal lexemes in the five Arthurian romances usually attributed to Chrétien de Troyes does provide some evidence that a second author composed the end of the *Lancelot*. While that evidence is too exiguous to command assent, it is also sufficiently suggestive to encourage further stylometric analyses. We note that our quantitative analyses are driven by so-called traditional methods of authorship attribution, which include non-stylistic data internal to the text. It is after all the text itself which tells us that it had two authors!

Although we allow the text to motivate the question, there is no reason to trust the text itself for an answer. And here we come to the profundities of the nature of authorship. What is an author? Stylometry does not force us to abandon the many nuances we can bring to our answers. Medieval authorship is particularly fraught with subtlety, for these reasons among others: medieval *inventio*, while not slavish repetition, created, perhaps, a higher degree of imitation than we come to expect today; medieval textual transmission almost ensures a degree of collaboration. When confronted with a medieval text, stylometry has its work cut out for it.[21] But the corpus considered here has the advantages of being uniform in genre and scribal hand.[22] It shows much promise as a subject for further study.

The "key unresolved problem" of stylometry noted by Banks is that of "variable selection." For example, Kenny's *Stylometric Analysis of the New Testament* used ninety-nine measures of style to study whether Luke wrote *Acts*, John wrote *Apocalypse*, and Paul wrote the Pauline letters (123–24). But, as Rudman reminds us in a different context, "that only scratches the surface—there are thousands of quantifiable style markers" (28c). The usual ways through this problem involve either data-mining or a selective approach. The former can lead to cherry-picking, as Banks explains: "Between any two bodies of text, some of the enormously many measurable variables will be consistently similar solely by chance, even if both texts are written by different people; conversely, some features will be consistently different, even though the works have common authorship." And since the number of total possible measures is so large, we might be picking cherries without even knowing it. Still, there are statistical and editorial ways to guard against cherry picking. One can validly eliminate style markers that are intra-corporeally inconsistent or those that are inter-corporeally indistinguishable (Rudman 31a). Moreover, one must verify the accuracy of the base text since inaccurate data will invalidate any subsequent analysis; we did so above in our corrective use of Ollier's revised concordance.[23]

The selective approach is more obviously problematic: "Many researchers proceed unautomatically; they use human intuition to notice unusual stylistic features in the uncertain text or the candidate authors, and then base their analyses solely on those. This is close to what human experts do, but it courts the danger of bias; people tend to notice things that conform to their prior hopes" (Banks). But here, we have perhaps found a third way to navigate between a cherry-picking Scylla and a Charybdian circularity.[24] For we did not choose the measure of temporal lexemes; it was chosen as a measure for us by a specialist of Chrétien's style. Had James-Raoul argued from the temporal lexeme data that there was a marked difference between Chrétien's style and Godefroy's style, then we would indeed be "court[ing] the danger of bias," even if we were correct in our statistics. But our analyses do not "conform to [the] prior hopes" of James-Raoul, and yet they still show that, if the use of temporal lexemes is a valid marker of style for Chrétien de Troyes, it may be an etch (a *griffe*) of his—or Godefroy's—signature, the trace of which is outlined by the numbers above. Not only did we not choose this style marker, but the scholar who chose it drew conclusions from it contradicted by our results. We note here a gentle irony: any valid marker of style for Chrétien's authorial signature supports James-Raoul's overarching thesis that such markers exist and point to an historical author with a recognizable style.

It is said throughout the literature on stylometry that statistical approaches must mingle with traditional analyses. The latter not only make sure that the statistics are being brought to bear on validly edited objects, but also give meaning to the results. While statistically significant differences between texts might suggest different authors, we still have other interpretive escapes. For example, by coming to interpret Chrétien as having performed a successful "obfuscation attack," we might cling to Hult's ingenious suggestion that Chrétien invented Godefroy: Chrétien deviously provided us with a false co-signature for his fictitious co-author! On the other hand, if the results of further stylometric analyses show that the beginning and ending of the *Lancelot* are most likely by the same author, then we would be more justified in believing the Godefroy-as-fiction thesis. And yet, we might instead use the concept of an "imitation attack" as a metaphor for a medieval continuator having successfully made his text stylistically indistinguishable from his model: thus, although Godefroy signs his name at the end, he may have wished the transition to be as anonymous as possible, leaving us no trace in the text but his name. The differences among successful / unsuccessful imitations or successful / unsuccessful obfuscations will remain open to critical interpretation. Still, there is every reason to include the mathematics behind such interpretations when we seek to answer the question: just who let Lancelot out of the tower?

Notes

1. See Labbé and Labbé, or Forestier's web-site devoted to the controversy.

2. More recently, Fish argues that the data-mining of texts ("text-mining") is a method ("if it can be called that") that "proceed[s] randomly or on a whim." Liberman responds by calling attention to the "paradoxically semi-quantitative nature" of Fish's own method of reading.

3. Consider M. Gardner's *Visitors from Oz*. Gardner sought "to come close to [L. F.] Baum's simple style, and to be faithful to the whimsical land and characters he created" (xv). But a stylometric analysis by Binongo found that Gardner's "imitation attack," as it were, failed. The later text still differentiates itself from Baum's original texts. Indeed, it better imitates the style of Baum's continuator, R. P. Thompson! (15)

4. "This *senhal* or *nom de plume* ["Chrétien Girls"] accompanied us throughout our collaboration, creating a disposable, collective, authorial persona that made it possible to agree on the meanings, effects and affects of a collaboration" (Stahuljak et al. vii).

5. The "Chrétien Girls" appropriately treat their *nom de plume* as homologous to that of "Chrétien de Troyes".

6. We counted seven quantitative arguments advanced by James-Raoul in support of this thesis; in all but one the data are insufficient to support the necessary analysis.

7. We say "complement" to emphasize the interanimating roles of both traditional and non-traditional attribution measures, as well as of both internal and external evidence. All must cohere, and so any and all should be used to test any one argument for attribution. See Vickers for more helpful terminology (111).

8. We thus incorporate Stahuljak et al.: "In our view there is more to be gained by keeping these terms ['authorship' and 'style'] open and potentially plural, admitting the possibility of multiple styles and taking into account the agency of copyists, compilers, and patrons" (5).

9. We quote from the reference editions used by James-Raoul: those in the *Classiques français du Moyen Âge* (*CFMA*) series, all of which are based on BnF 794, the Guiot manuscript. The translation of the *Lancelot* is from the edition by Kibler.

10. It should be noted that Hult's fictionality thesis is made *given* the possibility that "statistical studies" may show an unnoticed stylistic difference between the two parts (79). Still, those studies seem to be limited to the evidence from rhyme patterns presented by Shirt in "How Much of the Lion Can We Put

Before the Cart?" Although "various features" of Chrétien's style were examined (7), Shirt only gives data for the rhymes, a type of "cherry picking" warned against more recently by Rudman. Moreover, Shirt's analysis is limited to a subjective evaluation of percentages: while one percentage may be objectively greater than another, the difference may not be statistically significant. Shirt's earlier quantitative arguments were also given without rigorous analyses of variance ("Godefroi de Lagny et la composition de la 'Charrete'" 37–39).

11. Walters argues against Kay over the historical reality of Godefroy: "The notion that Godefroi de Laigni was a pseudonym assumed by Chrétien (or vice versa) [...] is placed in question by the presence in the Champagne library of texts by a certain 'Magister Godefroi'" (41, n. 23). In the inventory from 1319 of the Collegiate Church of Saint-Étienne in Troyes, there is an "Item magistri Godefridi" (Lalore 2: 270 item 2271). Stirnemann suggests that this text is by Geoffrey of Monmouth: "[N]ous espérons pouvoir identifier quelques derniers manuscrits de l'inventaire, notamment le « *Magister Godefroi* », sans doute un témoin de l'*Historia regum Britannie* de Geoffroy de Monmouth ..." ("[W]e hope to be able to identify the few remaining manuscripts of the inventory, notably the 'Magister Godefroi', without doubt a copy of the *Historia regum Britannie* by Geoffrey of Monmouth"; 38).

12. Indeed, it may be *another* pseudonym for Chrétien de Troyes if even this name is one too, the "Christian from Troy," a pun calling our attention, perhaps, to his embodiment of the *translatio studii* and Christian fulfillment of pagan letters. But collaborative works can have a collective intentionality as stunningly complex as that of a strong poet, so why only see a single author as capable of "a highly nuanced, unlocalizable intentionality" (Hult 88)?

13. Where the division occurs is a little vague in James-Raoul's study. The editor of James-Raoul's reference edition, Roques, had put the division "sans doute à partir du v. 6150" ("undoubtedly from v. 6150 on"; ix). In the introduction, v. 6150 (presumably of Roques's and not W. Foerster's edition) is given by James-Raoul for the division, although vaguely as "aux environs du vers 6150" ("around line 6150"; 19). The distinction might not be so exact, as James-Raoul later gives a less precise division: "[...] Godefroy de Lagny a repris la main après le vers 6139 qui narre la séquestration de Lancelot" ("Godefroy of Lagny took up the relay after line 6139 which narrates the imprisonment of Lancelot"; 247, n. 186). Yet elsewhere, a sharp division is given at vv. 6146 and 6147 (507, n. 149). Ollier's *Lexique et concordance* divides the text "à partir du vers 6150" ("from line 6150"; *1989 ed*. xxviii). When quantitative data are given by James-Raoul, it is sometimes clear that the 6149/6150 division is used, since the *or* at v. 6147 is given to Chrétien by Ollier in the *Lexique et concordance*, as well as by James-Raoul in the table at 507–08. But tables at 521 and 534 give v. 6147 as the beginning. Others have argued for still different divisions. See Shirt, "Godefroi de Lagny et la composition de la *Charrete*," esp. 34–37.

14. The on-line *Dictionnaire électronique de Chrétien de Troyes* is a work in progress that promises to offer such resources.

15. The total word and line counts for the five romances come from Ollier (1989 ed. 234). The couplet count is simply the line count divided by two. We provide all of our data at the following website: <http://www.brianjreilly.com> .

16. Unfortunately, Ollier only gives a word count for the whole of the *Lancelot*. Our count here for L1 is simply Ollier's number for L minus a count for L2 (vv. 6150–7112) which we based on the *DECT*, which uses "semi-diplomatic" transcriptions of the Guiot manuscript by P. Kunstmann. The total word count of the transcription is given as 43,637, a difference of 234 words, or 0.5%, from the *CFMA* edition as reported by Ollier, also based on the Guiot manuscript.

17. Since James-Raoul's data involving line counts considers L2 to begin at v. 6147, we give that line count here and not the one for L2 beginning at v. 6150. To be clear, James-Raoul employs two different divisions between L1 and L2: for word counts, following Ollier, L2 begins at v. 6150; for line counts, L2 begins at v. 6147.

18. We use the neutral L2 to refer to that section of the *Lancelot* possibly by a different author. Ollier uses G for that same section. James-Raoul's data seem, with a few exceptions, to come from this concordance, but an error in that data suggests the use of the earlier, uncorrected edition. See below our discussion of the lexeme *hui*.

19. The *DECT* returns for occurrences of the lemma *hui* seven instances of *huimés* and two of *hui* where it means "cry" (*L* 407 and *P* 5836).

20. Kestemont provides a clear description of stylometry with specific reference to both medieval texts and rhyme in this journal's inaugural edition. With specific reference to the *Lancelot*, Thorington, in two studies, uses statistical analyses of a typology of rhyme to argue for a difference between L1 and L2: "Indeed, the use of rhyme differs so significantly in these two parts of the poem that dual authorship of the *Charrette* is a virtual certainty" ("Rhyme, Reason, and Poetic Technique" 126). Moreover, in the epilogue to the *Lancelot*, Godefroy is said to have successfully mirrored Chrétien's prologue, what we might call a successful imitation attack to maintain the text's *conjointure* ("De conter un conte par rime" 143).

21. Considering the collaborative nature of Shakespeare's dramatic texts, Rudman claims that "[t]here is no Shakespeare dramatic text valid for nontraditional authorship attribution studies" (28a). The texts as we have them are simply too collaborative. While we take Rudman's assertion as a caution, the field continues to advance at remarkable speed, incorporating the latest in statistics, neural networks, etc. While its progress is not inexorable, it has been impressive. For example, Kenny begins *A Stylometric Study of the New Testament* with the claim that "[a] necessary preliminary to stylometric comparisons between texts must be a thorough grammatical analysis of every word in the texts to be compared" (3). But some advances in stylometry have come to study aspects of the text unrelated to grammatical or semantic analyses, even to the point of using mere character *n*-grams. *N*-grams, a selection of characters, words, etc., of

length *n*, might reflect some underlying implicit cognitive processes that structures an author's style. For a description of this and other stylometric methods, see Stamatatos.

22. The Ollier concordance is based on editions of BnF fr. 794, the Guiot manuscript. However interventionist Guiot was as a scribe (from orthography to content), he gives us the earliest complete and consistent corpus we have.

23. We could not, however, check those data against the reference editions or the manuscript itself. The *DECT*, with its linking of text, manuscript image, and lemmatization, thereby promises to be a powerful tool for research. On the need to be both editor and statistician (or at least to create such a research team), see the section on "Selecting the Text" in Rudman (27b–28b).

24. Scylla, of course, picked sailors, not cherries, which leads us back to that other term for fallacious statistics: fishing.

Works Cited

Banks, David. "Questioning Authority." *Classification Society of North America Newsletter* 44 (1996). Web. 9 Mar. 2012.

Binongo, José Nilo G. "Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution." *Chance* 16.2 (2003): 9–17. Print.

Brennan, Michael, and Rachel Greenstadt. "Practical Attacks Against Authorship Recognition Techniques." *Proceedings of the Twenty-First Innovative Applications of Artificial Intelligence Conference* (2009): 60–65. Web. 9 Mar. 2012.

Chrétien de Troyes. *Lancelot, or The Knight of the Cart (Le Chevalier de la Charrete)*. Ed. and trans. William W. Kibler. New York: Garland, 1981. Print.

———. *Les Romans de Chrétien de Troyes édités d'après la copie de Guiot (Bibl. nat. fr. 794)* 3: *Le Chevalier de la Charrete*. Ed. Mario Roques. Paris: Honoré Champion, 1958. Print.

———. *Romans, suivis des Chansons, avec, en appendice* Philomena. Ed. Michel Zink. Paris: Le Livre de Poche, 1994. Print.

Fish, Stanley. "Mind Your P's and B's: The Digital Humanities and Interpretation." *New York Times* 23 Jan. 2012. Web. 23 Jan. 2012.

Forestier, Georges, et al. *Molière, auteur des œuvres de Molière*. Web. 9 Mar. 2012. <http://www.moliere-corneille.paris-sorbonne.fr>.

Gardner, Martin. *Visitors from Oz: The Wild Adventures of Dorothy, the Scarecrow, and the Tin Woodman*. New York: St. Martin's, 1998. Print.

Hult, David. "Author/Narrator/Speaker: The Voice of Authority in Chrétien's *Charrete*." *Discourses of Authority in Medieval and Renaissance Literature*. Ed. Kevin Brownlee and Walter Stevens. Hanover: UP of New England, 1989. 76–96. Print.

James-Raoul, Danièle. *Chrétien de Troyes, la griffe d'un style*. Paris: Honoré Champion, 2007. Print.

Kay, Sarah. "Who Was Chrétien de Troyes?" *Arthurian Literature* 15 (1997): 1–36. Print.

Kenny, Anthony. *A Stylometric Study of the New Testament*. Oxford: Clarendon, 1986. Print.

Kestemont, Mike. "Stylometry for Medieval Authorship Studies: An Application to Rhyme Words." *Digital Philology: A Journal of Medieval Cultures* 1.1 (2012): 42–72. Web. 9 Mar. 2012. <http://muse.jhu.edu/journals/digital_philology/v001/1.1.kestemont.html>

Kunstmann, Pierre, dir. *Dictionnaire électronique de Chrétien de Troyes*. LFA, U d'Ottawa ATILF, Nancy U. Web. 9 Mar. 2012.

Labbé, Cyril, and Dominique Labbé. "Inter–Textual Distance and Authorship Attribution: Corneille and Molière." *Journal of Quantitative Linguistics* 8.3 (2001): 213–31. Web. 9 Mar. 2012.

Lalore, Charles. *Inventaire des principales églises de Troyes*. 2 vols. Troyes: Société académique de l'Aube, 1893. Web. 9 Mar. 2012.

Liberman, Mark. "The 'dance of the p's and b's': truth or noise?" *Language log*. Web. 26 Jan. 2012.

Love, Harold. *Attributing Authorship: An Introduction*. New York: Cambridge UP, 2002. Print.

Ollier, Marie-Louise. *Lexique et concordance de Chrétien de Troyes d'après la copie Guiot*. Ed. Serge Lusignan; Charles Doutrelepont; and Bernard Derval. Montreal: Institut d'Études Médiévales, 1986. Print.

———. *Lexique et concordance de Chrétien de Troyes d'après la copie Guiot*. Ed. Serge Lusignan; Charles Doutrelepont; and Bernard Derval. Rev. ed. Montreal: Institut d'Études Médiévales, 1989. Print.

Reilly, Brian J. Personal website. Web. 3 Mar. 2013. <brianjreilly.com>.

Rudman, Joseph. "Cherry Picking in Nontraditional Authorship Attribution Studies." *Chance* 16.2 (2003): 26–32. Print.

Shirt, David J. "Godefroi de Lagny et la composition de la 'Charrete.'" *Romania* 96 (1975): 27–52. Print.

———. "How Much of the Lion Can We Put Before the Cart? Further Light on the Chronological Relationship of Chretien De Troyes's *Lancelot* and *Yvain*." *French Studies* 31.1 (1977): 1–17. Print.

Stahuljak, Zrinka, Virginie Greene, Sarah Kay, Sharon Kinoshita and Peggy McCracken. *Thinking Through Chrétien de Troyes*. Cambridge: D. S. Brewer, 2011. Print.

Stamatatos, Efstathios. "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology* 60.3 (2009): 538–56. Web. 9 Mar. 2012.

Stirnemann, Patricia. "Une bibliothèque princière au XII[e] siècle." *Splendeurs de la Cour de Champagne au temps de Chrétien de Troyes*. Spec. issue of *La vie en Champagne* (1999): 36–42. Print.

Thorington, Ellen [M.] "« De conter un conte par rime »: rimes riches dans *Le Chevalier de la Charrette* (*Lancelot*)." *Œuvres & critiques* 27 (2002): 132–54. Print.

———. "Rhyme, Reason, and Poetic Technique in Chrétien's *Chevalier de la Charrette* (*Lancelot*)." *Dame Philology's Charrette: Approaching Medieval Textuality through Chrétien's* Lancelot. *Essays in Memory of Karl D. Uitti*. Eds. Gina L. Greco and Ellen M. Thorington. Tempe: ACMRS, 2012. 113–26. Print.

Vickers, Brian. "Shakespeare and Authorship Studies in the Twenty-First Century." Rev. of *Shakespeare's Additions to Thomas Kyd's "The Spanish Tragedy": A Fresh Look at the Evidence Regarding the 1602 Additions,* by Warren Stevenson; and *Shakespeare, Computers, and the Mystery of Authorship*. Ed. Hugh Craig and Arthur F. Kinney. *Shakespeare Quarterly* 62.1 (2011): 106–42. Print.

Walters, Lori. "Holy Adultery: The *Charrette*, Crusader Queens, and the Guiot Manuscript (Paris, BNF FR. 794)." *Dame Philology's Charrette: Approaching Medieval Textuality through Chrétien's* Lancelot. *Essays in Memory of Karl D. Uitti*. Ed. Gina L. Greco and Ellen M. Thorington. Tempe: ACMRS, 2012. 37–75. Print.