Chapter 6

VERSIFICATION

FROM THE EARLIEST attribution of texts to Jehan, the primary motivating factor for doing so appears to have been a sense amongst scholars that the texts could be connected "wegen der ihnen eignen Einfalt des Vortrags und des unvollkommenen Sprach- und Reimvermögens" (Gröber 1888–1902, 2: 929).[1] Scholars have taken issue with Jehan's apparent "médiocrité technique" when it comes to versification (Munk Olsen 1978, cxxxvi),[2] using the apparent ubiquity of such failings across the *J-corpus* as grounds to suggest that the texts share a single author.

1 "because of the inherent simplicity of their delivery and their imperfect command of language and rhyme."

2 "technical mediocrity."

ABSTRACT This chapter examines Jehan's versificatory practices in the context of the traditional scholarly view of his apparently mediocre poetic capabilities. Rather than engaging in unhelpful debates around value, I instead take the key feature which has previously been presented to critique Jehan's work, namely the frequency of so-called "easy" rhymes, alongside other features such as the rates of use of masculine and feminine rhymes. By comparing the *J-corpus* to the generically comparable *A-* and *R-corpora*, I show that although the *J-corpus* is internally variable for both features, there is a coherence to this variability which differentiates it from the reference corpora. The most significant methodological contributions of this chapter are the introduction of the concept of a *rhyme coefficient*, a measure of how likely a word is to appear in rhyming position compared to elsewhere in the line, and our analysis of admixtures, that is pairs of words which frequently appear together in rhyming position. These techniques allow us to extract the clearest evidence so far of an identifiable authorial fingerprint, revealing that there were specific stock words which Jehan particularly favoured at the end of a line, especially when compared to the generically and formally comparable *A-corpus*. The examination of admixtures, meanwhile, particularly reinforces the shared-author hypothesis for *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus*, the four additional texts which we have associated with Jehan's *œuvre*.

DOI: 10.17302/ogkl548611 © Arc Humanities Press

Evidence for this comes, so the argument goes, from Jehan's copious repetition of the same word in more than one rhyming position in a stanza (Morawski 1939, 334), as well as in his use of non-elision of schwa in hemistich-final position, particularly with the particles *je*, *se*, *ne*, and *que* (Munk Olsen 1978, cxxxix), and his overuse of so-called "easy" rhymes such as words ending in *-ment*. Morawski (1939, 335) also brings our attention to the fact that in two cases (*Deux chevaliers*, vv. 205–10; *Chevalier et escuier*, vv. 69–72b), there are stanzas of more than four lines, the latter of which certainly does not appear to be the result of scribal corruption.

And yet, even if we were to indulge in assertions of aesthetic value, we would still be forced to admit, as Morawski does (1939, 335), that "[t]outes ces irrégularités se retrouvent, souvent à un degré plus élevé encore, dans d'autres poèmes de l'époque, écrits dans la même forme métrique."[3] Within the *A-corpus*, for instance, we may note that stanzas of more than four lines are found in *Confesseur* (five-line stanza vv. 358–62), *Vision* (five-line stanza vv. 566–70), *Thibaut* (six-line stanza vv. 1053–58), and *Vieillards* (five-line stanza vv. 24–28). None of these texts have so far stood out as potential candidates for Jehan's *œuvre*, revealing the difficulty of building a case for shared authorship on this feature. Ultimately if, across similar surviving texts from the period, "la plupart des spécimens […] ne se distinguent guère par une technique souveraine" (Munk Olsen 1963b, 157),[4] then the quality of Jehan's versification is ultimately of minimal value to the question of attribution, since it does nothing to distinguish his work from that of other authors.

Indeed, not only would it be questionably subjective to class Jehan's poetry as either "good" or "bad," it should go without saying that an individual author's output is never consistent in quality, as they hone and develop their craft across their career, or fall from earlier creative heights. The avenues pursued by previous scholarship regarding Jehan's versification therefore offer little evidence concerning the collective authorship of the *J-corpus*. By turning towards digital and statistical analyses of versification, however, we can both more comprehensively interrogate the validity of the existing claims concerning Jehan's versificatory tendencies, as well as pursue new avenues. Combined, these can tell us much more than previous scholarship has been able to determine about the extent to which the versificatory practices of the *J-corpus* unite it as a coherent group, distinct from texts found elsewhere.

3 "[a]ll these irregularities are found, often to an even greater degree, in other poems of the period, written in the same metrical form."

4 "most specimens […] are hardly distinguished by their masterful technique."

Masculine and Feminine Rhymes

With this in mind, we begin our application of digital methods by considering the use of masculine and feminine rhymes within Jehan's poetry.[5] The mean rate of use of feminine rhymes across the corpus is 31.5%, almost exactly the 1:2 ratio identified by Morawski (1939, 335) for his smaller corpus of Jehan's poems, which he claims is "la norme" (the norm) for texts of this period. As we shall see, the frequency with which Jehan uses feminine rhymes is not identical to that observed in texts of a comparable type, but it is within the expected range for the fourteenth century, given Lote's (1949–1996, 2:111–24) observation that the rate of use of feminine rhymes increases at a roughly even rate from near to zero in the earliest surviving poems in French, to near-parity with masculine rhymes by the end of the fifteenth century.

However, this rate of use of feminine rhymes is not without variability across the *J-corpus*, as shown in Figure 5. Morawski identifies unusually low rates in *Chevalier et escuier* and *Merlin* (from our analysis, 21.1% and 23.2% respectively), but misses the similarly low rate of 20% in *Enfant rosti*, and the much lower rate displayed in *Mescreant* (13.7%), while also not recognizing the much higher than average uses in *Povre chevalier* (41.5%), *Beguine* (41.2%), *Respon* (39.6%), and *Pain* (38.1%). Since he does not include either text in his corpus, Morawski also does not find the highest rate of use to appear within our *J-corpus*, namely *Abesse* (42.3%), as well as the high rate observed for *Florence* (41.2%).

One may choose to hypothesize that some explanation for the increased use of feminine rhymes in these texts may result from the fact that, except for *Pain*, all of these texts displaying high rates of feminine rhymes centre

5 A feminine rhyme is one ending in atonic *e*. In practice this includes forms such as the third person plural *-ent*, as well as words ending in *-es*. For this analysis, stanzas are said to have a feminine rhyme if they contain at least two forms ending in any of *-e*, *-es*, or *-ent*. To avoid including masculine rhymes ending in *-ent*, stanzas containing a single example of a form ending in *-ment* or *-ant*, or those with a single letter before *-ent* (such as *vent*), or two if the *-ent* is preceded by a vowel (such as *vient*), are excluded on the assumption that *-ent* in these cases is likely to result in a nasal vowel. These requirements are represented through regular expressions, namely that at least two lines of a stanza must satisfy `es?$|[a-zéäëïöüÿ]{2,}ent$`, and no line in the stanza can satisfy `ment$|ant$|^[a-z]i?ent$|é$`. This will lead to the exclusion of rare stanzas containing third-person plural forms with *m* preceding the *-ent*, such as *entament*, but sampling suggests that the impact on the data is minimal. For more details on the use of regular expressions, see p. 19.

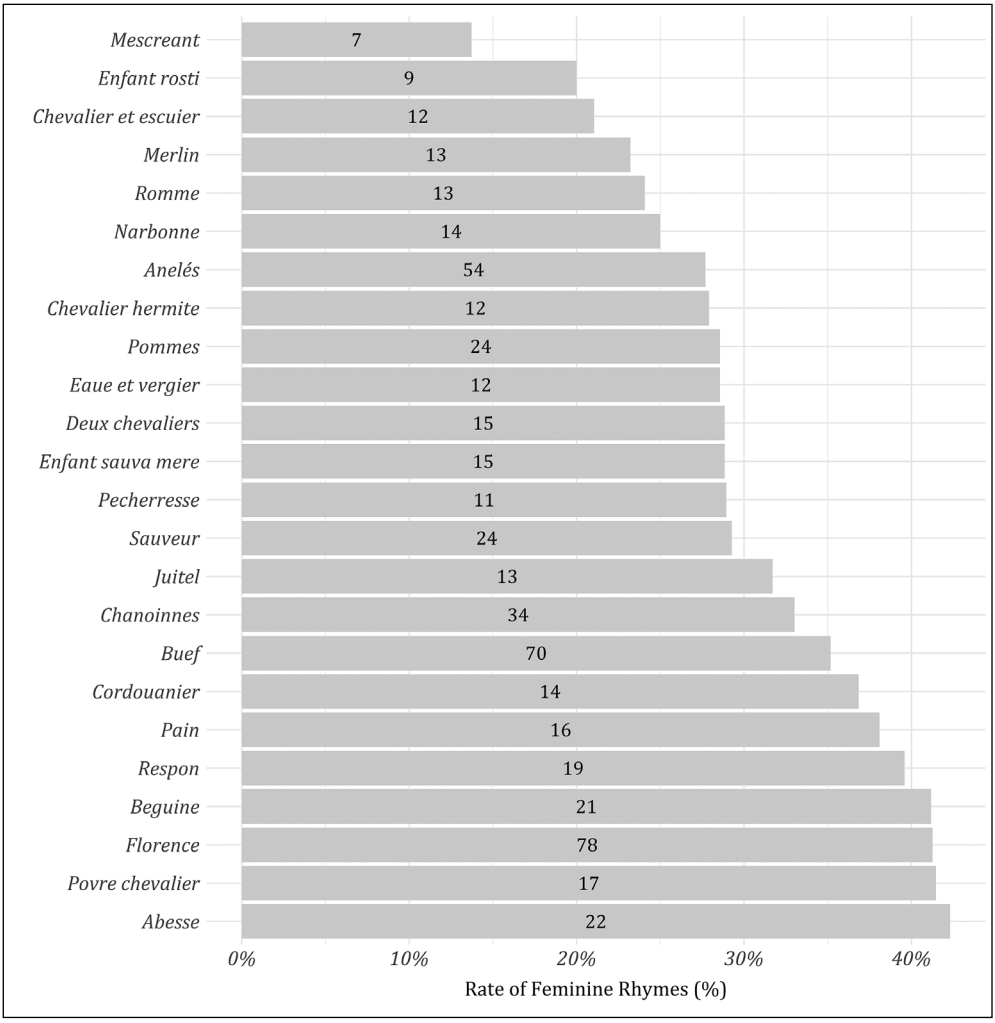

Figure 5. Rates of use (x-axis) and raw numbers (figures) of stanzas containing feminine rhymes for each text of the *J-corpus*.

around the experience of women,[6] making words ending in *e caduc* more likely, through for example the increased prevalence of nouns such as *fame* (e.g., *Povre chevalier*, vv. 102, 104), as well as variable adjectives and past participles with feminine agreement (e.g., *Povre chevalier*, vv. 114–15). The inverse argument could be made for the low rate of use found in *Mescreant*, a text containing only male characters which is therefore unlikely to require

6 In *Povre chevalier*, it is the actions of the knight's wife and of Mary which form the crux of the story.

as many words ending in *e caduc* as some others. This offers a somewhat convincing argument for these texts, but does not appear to be the sole explanation for the variation across the *J-corpus*, given the absence of female characters in *Pain* (38.1%), which nevertheless has a much higher rate of feminine rhymes than *Romme* (24.1%), *Narbonne* (25%), and *Pecherresse* (28.9%), which centre around a single female character, and yet in turn have lower rates of feminine rhyme use than the corpus average.

On the whole, then, Jehan's texts fit the overall frequency of use of feminine rhymes observed more widely by Lote and suggested for the corpus by Morawski, but they are certainly more variable than the latter acknowledges. While in some cases this variability can be explained with reference to the contents of a particular text, this is not always the case.

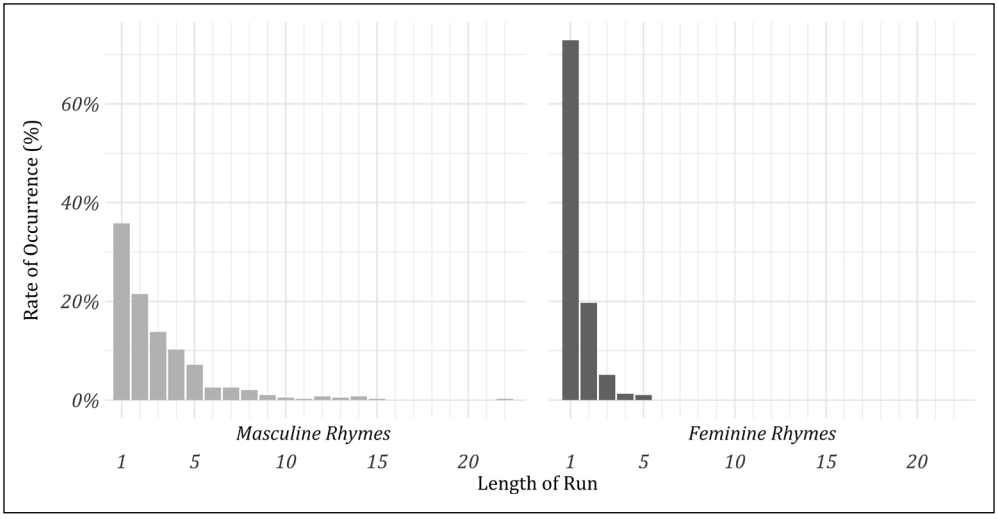

Figure 6. Rates of occurrence for runs of stanzas containing masculine and feminine rhymes across the *J-corpus*.

Munk Olsen (1978, cxxxix) argues that within Jehan's texts, monorhyme stanzas containing masculine rhymes appear alongside those containing feminine rhymes "sans principe fixe."[7] Overall, this is true, and there are no texts in which a clear pattern of usage is maintained the whole way through. Nevertheless, there are some tendencies to be observed in the *J-corpus*. Firstly, on a purely statistical level, by comparing the lengths of sequences, or "runs," of stanzas containing masculine versus feminine rhymes, as represented in Figure 6, we note that the appearance of stanzas

7 "without any fixed principle."

containing rhymes of a given gender flanked on either side by rhymes of the opposite gender, rather than in a run of more than one such stanza, is far more common for feminine rhymes (72.9%) than it is for masculine rhymes (35.8%). That is, stanzas containing feminine rhymes are about twice as likely to appear on their own, flanked by two stanzas of masculine rhymes, than the inverse. This is perhaps at least somewhat due to the simple fact that there are fewer stanzas with feminine rhymes than those with masculine rhymes across the corpus. Were this to be the only explanation, however, we might expect runs of two stanzas to be the most frequent type for masculine rhymes, with masculine and feminine rhymes alternating on a two-to-one basis.

This is not the case for the *J-corpus*, however, since in fact patterns of one-to-one alternation between stanzas of masculine and feminine rhymes are more common. The greater difference between the use of masculine and feminine rhymes can instead be accounted for by the fact that the longest run of feminine rhymes is five (twice in *Florence*, once in *Buef*, and once in *Abesse*), while for masculine rhymes it is twenty-two (*Romme*): in his texts, Jehan appears far more willing (or perhaps able) to use long runs of stanzas containing masculine rhymes than he is to make use of long runs of stanzas involving feminine rhymes.

Indeed, if we exclude those runs of stanzas containing masculine rhymes which have a length greater than the longest run of feminine rhymes (five), we find that only 770 stanzas of a total of 1,172 fit these criteria, putting the stanzas with masculine rhymes at a little under 50% greater than the number of stanzas containing feminine rhymes (539). For the most part, then, the predominance of masculine rhymes across the *J-corpus* perhaps results from an aesthetic choice with Jehan preferring, if not sticking rigidly to, either a one-to-one or a two-to-one alternation between stanzas containing masculine and feminine rhymes, but permitting the occasional interruption by longer runs of stanzas containing masculine rhymes.

By comparing the rates of use of feminine rhymes in the *J-corpus* with the *A-corpus* of narrative alexandrine monorhyme poems which have hitherto not been attributed to Jehan, we can gain a clearer understanding of whether the 2:1 ratio of masculine to feminine rhymes is as typical as earlier scholarship has suggested.

From Figure 7, we observe that the median rate of use of feminine rhymes in the *J-corpus* (28.9%) is relatively comparable to that of the *A-corpus* (31.8%), suggesting that there is an expected rate for alexandrine monorhyme quatrains from this period. The range, however, is vastly greater. The *A-corpus* contains one lower outlier (*Troyes*, 0%) and two upper outliers

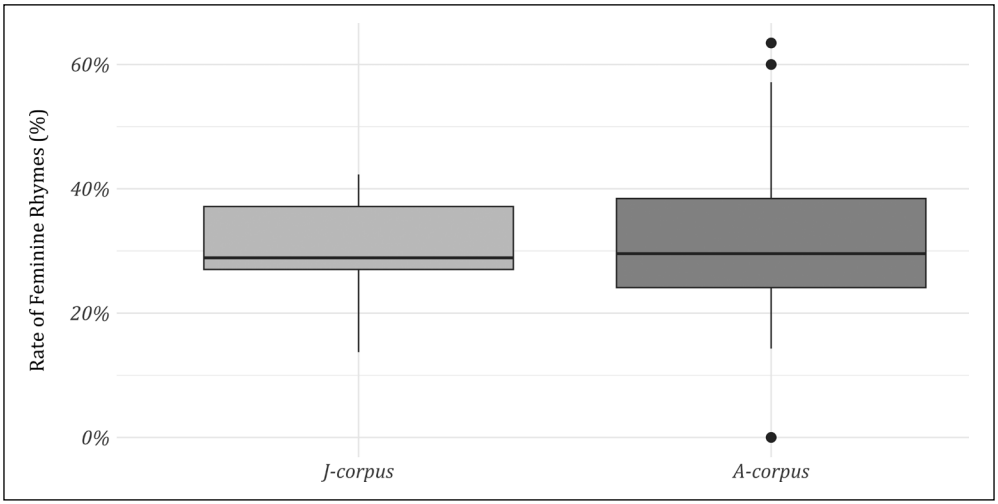

Figure 7. Box and whisker plots for the rate of use of feminine rhymes in the *J-* and *A-corpora*.

(*Dieu d'Amours*, 60%; *Quatre sereurs*, 63.5%), and a further four texts display rates greater than the highest rate of the *J-corpus* (*Abesse*, 42.3%), namely *Confesseur* (44.1%), *Barbe* (44.6%), *Thomas* (47.4%), and *Yves* (57.1%). As a result, while it is possible to dismiss *Yves* and *Dieu d'Amours* on account of their brevity, at just seven and five stanzas respectively, the remaining evidence provides coherence to the *J-corpus* by suggesting that Jehan's work is much less variable than other comparable texts.

The rate of use of feminine rhymes, then, appears to be broadly comparable across texts of this kind, but it is at least in part also a feature of authorial preference, since variability is possible. It is therefore all the more notable, looking at the individual rates of the *A-corpus* in Figure 8, that *Guillaume* and *Robert*, the two texts which we have thus far hypothesized to be the work of Jehan, as well as *Jehan Paulus* and *Richard*, the apparent echoes of earlier works by the same author, should display rates of use of feminine rhymes that place them squarely in the middle of the rates observed for the *J-corpus*, at 31.6%, 28.7%, 29.0%, and 29.1% respectively. This overlap heightens the association between these texts, and reinforces the shared-author hypothesis for these texts alongside the *J-corpus*. Of course, the presence of similar rates for *Basile*, *Alexis*, *Grégoire*, *Sebastien*, *Vieillards*, and *Dieudonnée* means that this evidence on its own is not enough to confirm attribution, and but it does still contribute to the more holistic evidence which we are building over the course of this study.

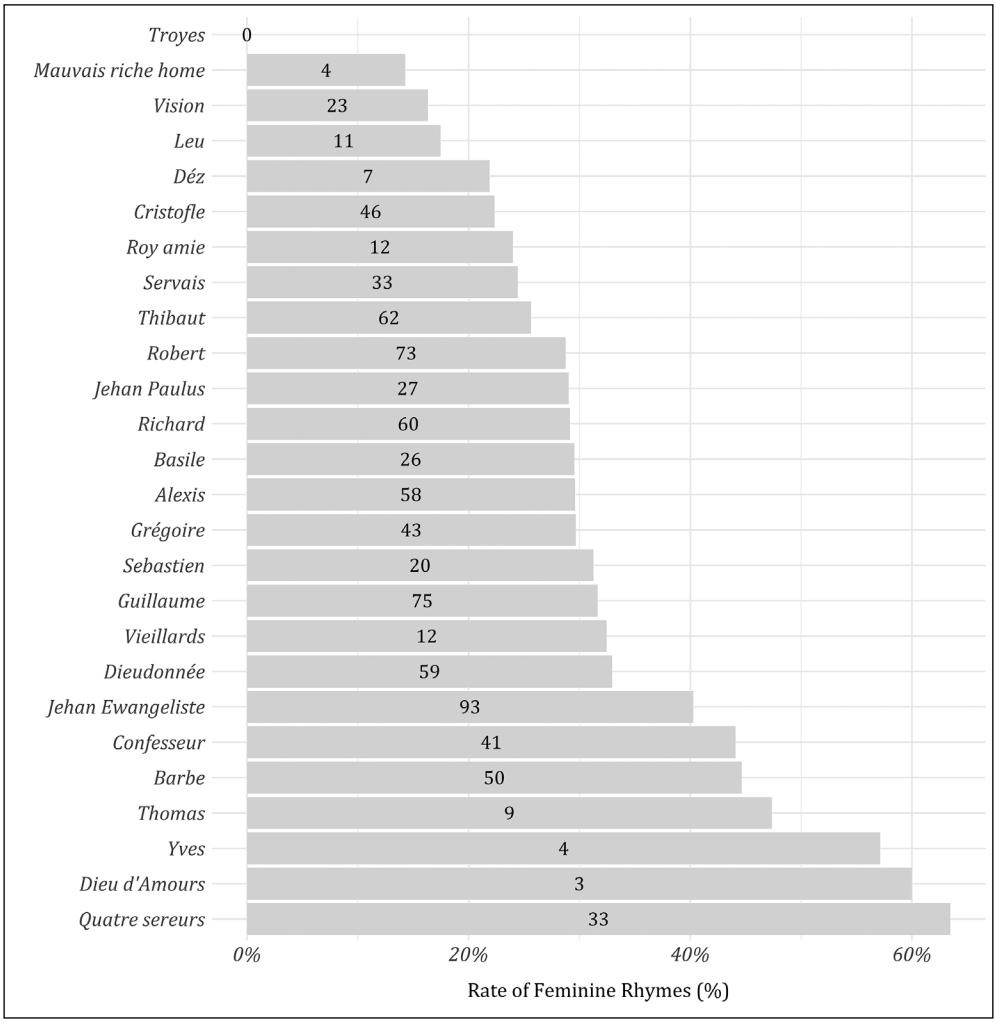

Figure 8. Rates of use (x-axis) and raw numbers (figures) of stanzas containing feminine rhymes for each text of the *A-corpus*.

Returning also to the "runs" of stanzas containing masculine versus feminine rhymes, as represented in Figure 9, we observe a similar pattern in the *A-corpus* to that seen for the *J-corpus*, namely that runs of masculine rhymes are often much longer than runs of feminine rhymes. The longest run of feminine rhymes in the corpus is seven (one each in *Barbe* and *Jehan Ewangeliste*), while the longest run of masculine rhymes is twenty-six (in *Leu*), with runs of twenty-two and twenty-four stanzas also found in *Alexis* and *Vision* respectively. *Guillaume*, *Robert*, *Richard,* and *Jehan Paulus*, meanwhile, align with the broader trends of the *J-corpus*, with one ten-stanza run

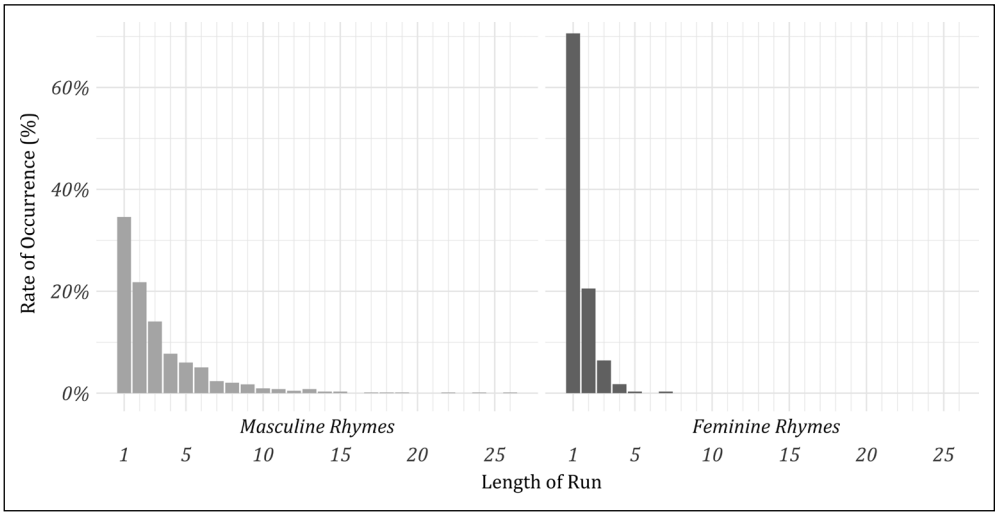

Figure 9. Rates of occurrence for runs of stanzas containing masculine and feminine rhymes in the *A-corpus*.

of masculine rhymes appearing in *Guillaume*, two twelve-stanza runs in *Robert*, one nineteen-stanza run in *Richard*, and one ten-line run in *Jehan Paulus*.

As with the *J-corpus*, by excluding those runs of stanzas containing masculine rhymes which exceed a length of five, the number of masculine stanzas in *Guillaume*, *Robert*, *Richard,* and *Jehan Paulus* drops from 555 to 347, and is approximately 50% greater than the number of feminine stanzas (235), therefore suggesting that for these texts, as in the *J-corpus*, stanzas of masculine and feminine rhymes alternate on a roughly one-to-one or two-to-one basis, with the overall data skewed by the presence of longer runs of masculine stanzas.

To conclude our section on feminine rhymes, we find that the usage observed for the *J-corpus* is arguably close to typical, as previous scholarship has suggested, but still displays some variability within these rates. Some level of variation in the use of feminine rhymes appears to be possible both within the *œuvre* of one single author and between the work of different authors, and while we are occasionally able to explain this variation by considering a text's content, it is frequently impossible to do so. Evidence of far more variability within other narrative texts written in monorhyme alexandrine quatrains, however, supports the collective attribution of the *J-corpus* to a single author, as well as reinforcing the shared-author hypothesis for *Guillaume*, *Robert*, *Richard,* and *Jehan Paulus*, and suggesting that the *remanieur(s)* of the latter two texts did not make especially significant changes to Jehan's original versification.

Rhyme Frequency and "Easy" Rhymes

Many of the value judgements concerning the quality of Jehan's poetry stem from the claim that Jehan makes excessive use of so-called "rimes faciles" (Munk Olsen 1978, cxxxvii), "easy" rhymes, the preponderance of which allows us, so the argument goes, if not to identify the work of an individual author, then at least to "distinguer le bon poète et le versificateur irrémédiablement mauvais" (1963b, 158).[8] Morawski (1939, 334) defines this class of "easy" rhymes as consisting primarily of nouns and adverbs ending in *-ment*, as well as verb forms sharing common endings, in particular the present participle (*-ant*) and infinitives.

While we may wish to dismiss the value judgements with which it is associated, the supposed abundance of these so-called "easy" rhymes means that they are an ideal candidate for a case study in the examination of the frequency of certain kinds of rhymes within the *J-corpus*, which we will perform here using methods of digital textual analysis. What follows is therefore not an attempt to pass aesthetic judgement on Jehan's texts. For the sake of simplicity, we will refer to the forms in question as "easy" rhymes, yet the aim of this section is solely to reach a clearer understanding of the idiosyncratic uses of such rhymes that, so it has been claimed, serve to unite the corpus, and in so doing to further interrogate the hypothesis that the texts share an author.

Our focus here will be on two specific types of "easy" rhyme which are claimed to be particularly common in Jehan's texts, therefore giving us the largest possible dataset, namely words ending in *-ment* and *-ant*. On occasion, while he largely prefers to keep them separate, perhaps on etymological or visual grounds, Jehan occasionally rhymes these forms interchangeably with each other, as in the stanza below from *Romme*, underscoring the fact that for Jehan, by the time he wrote his poems at the beginning of the fourteenth century, and in keeping with broader linguistic trends, *an* and *en* had an identical pronunciation.

> De leur biensfez estoit par tot le renon grant;
> Il orent de .ii. anz .i. moult tres bel enfant:
> Il estoit valeton, si l'amerent forment.
> Li preudon n'en ot plus en trestot son vivant.
>
> (*Romme*, vv. 13–16)

8 "distinguish the good poet from the hopelessly bad versifier."

To reduce the influence of orthographic variation on the data, which could result from scribal rather than authorial preferences, I also include those rhymes where the final *t* of *-ant* is replaced by *s* or *z*, since this variation does not appear to be caused by variation in the underlying sound of the rhyme, as evidenced by the stanzas below, in which *-ant*, *-ans*, and *-anz* rhyme interchangeably.[9] It is nevertheless notable that on the whole these forms are indeed kept separate, again perhaps for visual or etymological reasons rather than auditory ones, with stanzas containing a mixture of forms only occurring relatively rarely.[10]

> Celui qui fu preudons si fu tous merveillans
> De ce que son voisin li aloit devisans.
> Il hauça sa main destre, si s'en ala seignant
> Et puis si li a dit: "Vous estes mescreant,"
>
> (*Mescreant*, vv. 49–52)

> Li preudons respondi, moult liez et moult joians,
> Pour ce que Diex ot fait vertuz suz ses enfans:
> "Sire, a vostre gré faire sui touz obeissanz."
> "Tu as droit," dit le pappe, "ja ni seras perdanz."
>
> (*Buef*, vv. 769–72)

Through examination of the uses of "easy" rhymes in Jehan's corpus of texts, as represented in Figure 10, we find a similar story to that which we observed for feminine rhymes, in that the use of such rhymes in the *J-corpus* is significantly variable. The data have a mean of 11.3% for the usage of "easy" rhymes, but rates for individual texts range from 4.2% (*Florence*) to 22.2% (*Romme*). It is difficult to come to any sort of conclusion as to why this variation would occur in the *J-corpus*, but there does appear to be a weak negative correlation between the number of stanzas contained within a text, and the rate at which that text contains "easy" rhymes, that is to say, the shorter a text is, the more likely it is to make use of "easy" rhymes (-0.37247). A Pearson's product-moment correlation test suggests that the correlation is not quite

9 The same does not appear to be true for *-mens* and *-menz*, which never rhyme with *-ment* in Jehan's texts. The regular expression which at least one line-final word in each stanza must match to be counted is therefore `an[tsz]$|ment$`.

10 This would suggest that while, as we have seen, Jehan's texts do appear to have been destined for performance, there was nevertheless also a preference for some degree of aesthetic uniformity in the written medium, whether that arose from a preference by the author or by subsequent scribes.

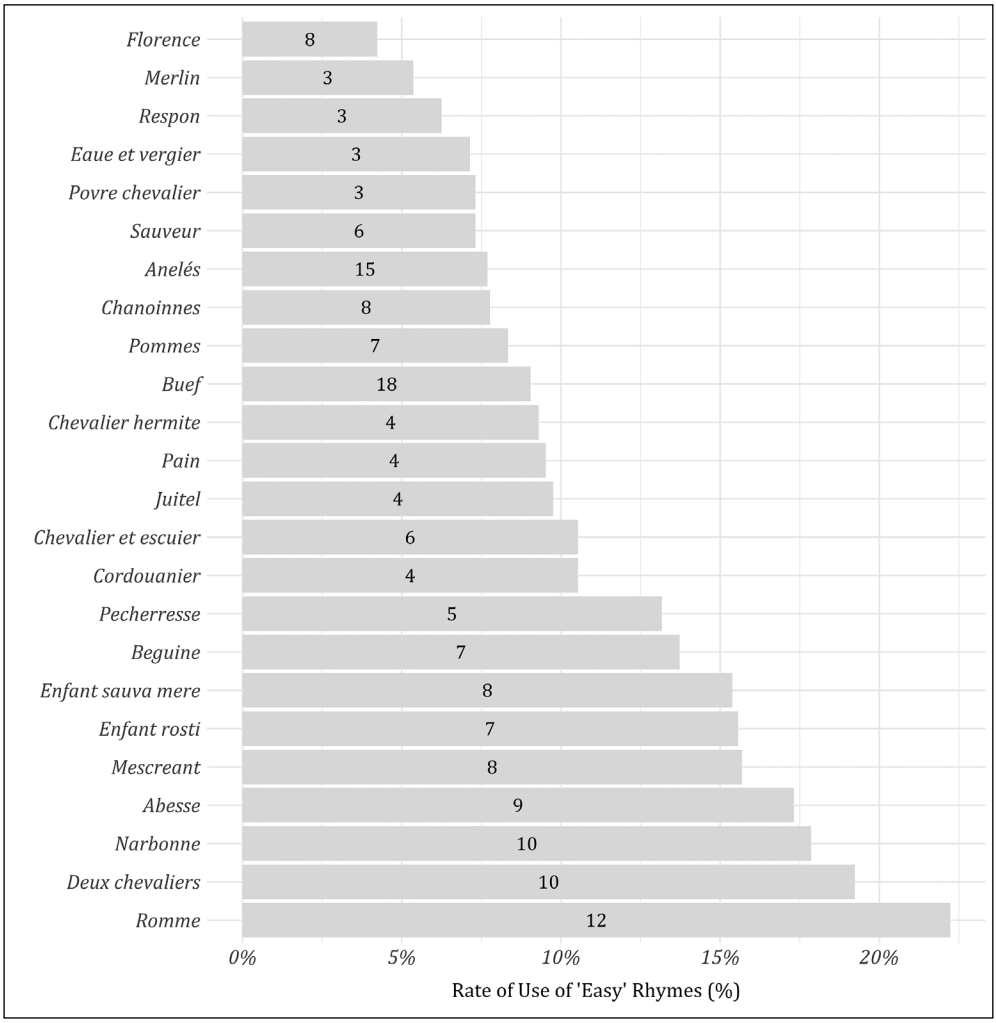

Figure 10. Rates of use (x-axis) and raw numbers (figures) of stanzas containing "easy" rhymes in the *J-corpus*.

strong enough to determine statistical significance by our 95% threshold ($p = 0.07306$), but the very possibility of such a correlation is intriguing, as it could suggest that, if indeed we assume that such "easy" rhymes require less artistic effort than others, then the works which demanded Jehan's greatest commitment to versificatory complexity were his longer texts.[11]

11 No such significant correlation appears between the use of feminine rhymes and text length.

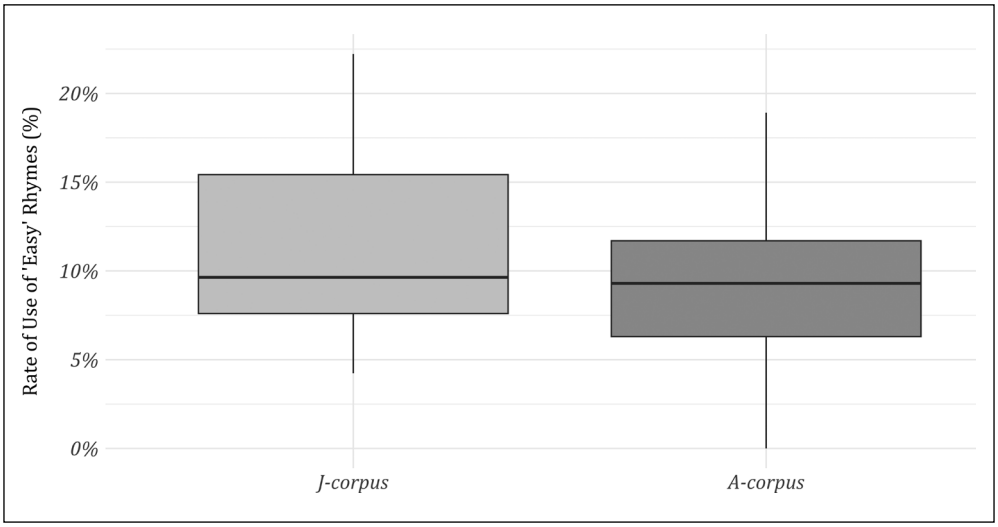

Figure 11. Box and whisker plots for the rate of use of "easy" rhymes in the *J-* and *A-corpora*.

When we compare the rates of use of "easy" rhymes in the *J-corpus* to the *A-corpus* containing narrative poems in alexandrine monorhyme quatrains which are not usually attributed to Jehan, as represented in Figure 11, we observe that the median rate of the *A-corpus* (9.30%) is very comparable to that of the *J-corpus* (9.64%), but also that the range of the data is lower than that of the *J-corpus*, while the interquartile range is wider (*J*: 7.8%; *A*: 5.4%). All of this serves to reinforce the suggestion that in many cases, if not on average, Jehan's texts display a greater rate of use of "easy" rhymes than other, comparable poems, and across a broader spread.

Looking at the individual rates for each of the texts in Figure 12, we again observe similarities with Figure 10, and note that *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus*, the four texts to which we have extended the shared-author hypothesis, all fit comfortably within the range of the *J-corpus*. Beyond this, there are also differences between the corpora in that, even excluding the fragmentary *Dieu d'Amours*, there are three texts in the *A-corpus* which display lower rates of "easy" rhymes than the lowest rate in the *J-corpus* (namely *Florence*, at 4.2%), and only two which exceed a rate of 15% (*Troyes* and *Vieillards*), compared to seven in the comparably-sized *J-corpus*. In the *A-corpus*, there is also no correlation between stanza count and rate of use of "easy" rhymes which even approaches statistical significance ($p = 0.8718$), such that there are clearly some differences between the preferences for "easy" rhymes expressed in the *J-corpus* and those of the *A-corpus*.

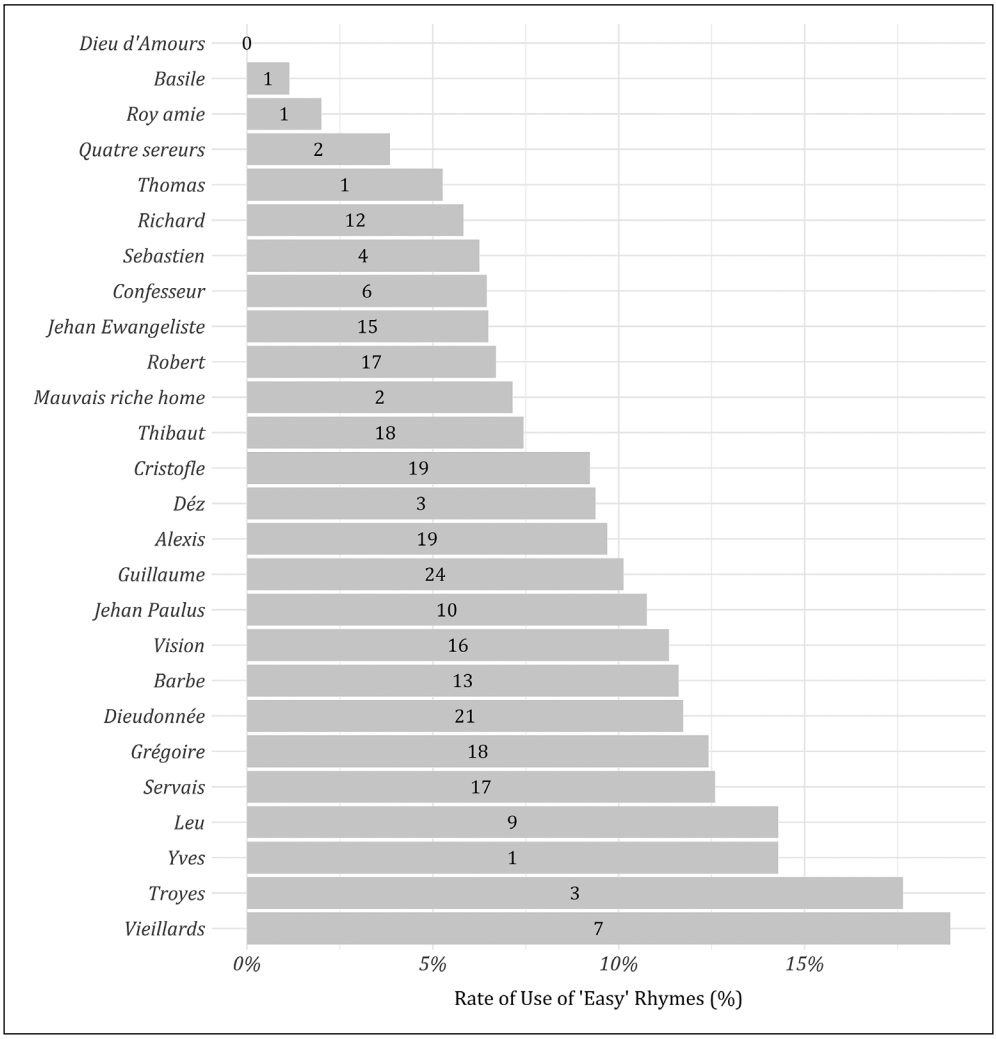

Figure 12. Rates of use (x-axis) and raw numbers (figures) of stanzas containing "easy" rhymes in the *A-corpus*.

Overall, then, we are able identify certain features of Jehan's versificatory preferences through examination of "easy" rhymes, in particular his willingness to allow higher rates in some of his texts, especially the shorter ones. However, an additional statistical test here reveals that it is not straightforwardly true that Jehan simply makes greater use of "easy" rhymes than those writing comparable works, as the existing scholarship claims. A Welch Two Sample t-test here produces a result ($p = 0.09249$) that is not quite sufficient to suggest that there is a statistically significant difference

between the two corpora for this feature, and this remains true even if we exclude *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus* (p = 0.1369) from the *A-corpus*, or include them alongside the *J-corpus* (p = 0.2032). That is, while such a suggestion appears plausible, the two ranges are sufficiently similar that we are unable to conclude with certainty that their difference results from anything other than random variation, and thus attributed authorship alone cannot account for variation in this feature.

Following these observations in the *A-corpus*, and as a final point of analysis in this section, it is worthwhile comparing the *J-* and *A-corpora* to the *R-corpus*, containing the complete surviving poetic works of Rutebeuf. In doing so, we can determine whether the rate of use of "easy" rhymes is an apparently standard feature of narrative poems written in alexandrine monorhyme quatrains, when compared to Rutebeuf's *œuvre* which is more diverse in form.[12]

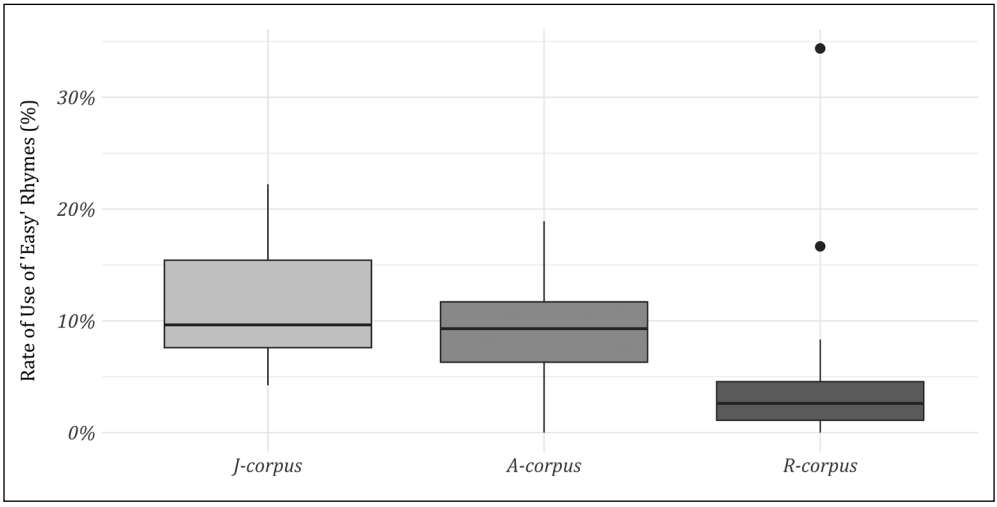

Figure 13. Box and whisker plots for the rate of use of "easy" rhymes in the *J-*, *A-*, and *R-corpora*.

The data for this comparison are presented in Figure 13. In this case, for the *R-corpus*, we note that the median use of "easy" rhymes is much lower than that observed for the *J-* and *A-corpora*, with thirteen texts of the fifty-six in the *R-corpus* not containing even a single instance of an "easy"

12 Given this formal variety, when calculating the rate of "easy" rhymes for the *R-corpus*, it is necessary to consider the texts on a line-by-line basis, since we are often not able to divide the text into monorhyme stanzas.

rhyme. Notably, this includes two of Rutebeuf's poems written in alexandrine monorhyme quatrains, namely *Li diz des Cordeliers* and *Des Jacobins* (neither of which are strictly narrative in their content). The presence of the two upper outliers adds somewhat to the sense that the work of a single author need not be consistent in its use of "easy" rhymes, and in this case adds to the sense of aesthetic variety that is present within Rutebeuf's *œuvre*. The highest rate anywhere in the *R-corpus*, displayed in *La chanson de Puille* (34.4%), appears to result from the aesthetic concept of the poem, with "easy" rhymes appearing in the fourth and eighth lines of each of the poem's eight-line stanzas. The relatively high rate in *De Brichemer* (16.7%), meanwhile, appears more straightforwardly to be a function of the brevity of this poem, with "easy" rhymes appearing in just four of the poem's twenty-four lines, all rhyming with each other as one of the two rhymes in an eight-line alternating stanza.

Overall, the *R-corpus* contains significantly lower rates than the *J-corpus* ($p = 0.00000$) and the *A-corpus* ($p = 0.00004$). This observation allows us to reframe Munk Olsen's claim that, for poems written in alexandrine monorhyme quatrains, "la plupart des spécimens [...] ne se distinguent guère par une technique souveraine" (1963b, 157).[13] Setting aside his value judgement, it does nevertheless appear to be true that Jehan was in good company alongside the other authors of narrative poems written in alexandrine monorhyme quatrains, in that he displays a notable preference for rhymes ending in *-ment* and *-ant*. What is equally clear, however, is that this feature alone, while allowing us to identify outliers like *Basile* which are unlikely to be the work of Jehan, cannot enable us to claim that Jehan's use of easy rhymes somehow marks him out as a "versificateur irrémédiablement mauvais" (Munk Olsen 1963b, 158)[14] relative to his peers.

Word Frequency in Rhyming Position

Alongside his claims about the quality of Jehan's poetry, Munk Olsen (1963b, 160) suggests a further element of versification which may prove effective in considering the shared-author hypothesis for Jehan's poems. It may be fruitful, Munk Olsen argues, to consider *associations des rimes* or *rhyme associations*, that is, which words appear most frequently in rhyming position when a certain rhyme is required. While he does not perform a full analysis, Munk

13 "most specimens [...] are hardly distinguished by their masterful technique."

14 "hopelessly bad versifier."

Olsen's method appears promising, since it allows him to compare his corpus of Jehan's texts to *Alexis* (included in our *A-corpus*) through analysis of the rhymes *-on(s)* and *-is*.[15] In doing so, Munk Olsen observes differences in the words most frequently found for each rhyme, and reinforces his claim that the text cannot be said to be another work by Jehan.

Digital textual analysis, however, allows us to go one step further, looking not simply at which words appear most frequently for a given rhyme, but at which words are more likely to appear in rhyming position than at other moments during the text. To do so, we calculate the rate at which a given word appears in rhyming position, as well as the rate at which that same word appears across all remaining tokens of a given text.[16] We then divide the former rate by the latter to determine what I term the *rhyme coefficient*, that is, how much more or less likely a given word is to appear in rhyming position than elsewhere in the line. This method allows us to make a critical distinction between those words which appear frequently due to the simple fact that they are often used across the corpus, because of authorial preference or narrative necessity, and those words which could be described as "stock" rhyming words, used to complete a rhyme, but rare elsewhere.

In this section of our analysis, we do not focus on the word forms that appear at the end of lines, rather their associated lemmata, that is, the arbitrarily determined "dictionary" form of a lexeme, as given by RNNTagger. This is an important distinction since it allows us to ignore the influence of both orthographic and inflectional variation and focus on the underlying lexeme choice. Rates of appearance are therefore calculated based on the appearance of lemmata, regardless of the exact word form to which they are associated. To avoid cases in which the data would be skewed by the appearance of only a few forms of a word, we limit our examinations here to lemmata which appear on at least ten occasions anywhere in the corpus, and at least once in rhyming position.

As an initial proof of concept, let us apply this method to the *R-corpus*, for which single authorship is more assured. Across the fifty-six texts of the *R-corpus*, we observe some forty-nine lemmata which match our selection

15 The corpus Munk Olsen uses for this analysis is shorter than the 1978 corpus on which he eventually settles.

16 The line-final rate is calculated by dividing the number of observed instances of a given lemma in final position by the total number of line-final tokens in the respective corpus (6,829 in the *J-corpus*), while the non-final rate is calculated using the total number of non-final tokens in the corpus (52,636).

criteria, and which appear with a rhyme coefficient greater than 1, meaning that they are more commonly used in rhyming position than elsewhere in a line. Of these, *longuement* (681.3), *jugement* (264.9), *mentir* (100.9), *errer* (75.6), *durement* (63.1), *semblant* (56.8), *neant* (53.4), and *vent* (50.5) are at least fifty times more likely to be included at the end of a line, suggesting that these are words which Rutebeuf significantly favoured in rhyming position, and indicating that our approach has significant validity as a means of identifying the individual versificatory "fingerprint" of a given author.

Applying this analysis across the *J-corpus*, we find several trends of note. Table 32 shows the top twenty lemmata across the corpus, when ordered by descending rhyme coefficient. First, we discover that forms of *certainement*, *envie*, and *demeuree* only ever appear in Jehan's texts in rhyming position, marking these forms out as words which may have been retained by Jehan for use solely in such a position (their rhyme coefficient is therefore infinite). We also note that many lemmata with a high rhyme coefficient, including *certainement*, *vraiment*, *commandement*, and *sauvement*, are examples of the so-called "easy" rhymes discussed above.[17]

Based on these examples of words with high rhyme coefficients, Jehan therefore appears to have had a stock of nouns and adverbs which he used largely in rhyming position when required to complete a stanza.[18] This is in addition, of course, to the many verbal forms satisfying the criteria for "easy" rhymes which contribute to the overall trends observed above, and yet notably, none of the verbs with the highest rhyme coefficients appear

17 Across all lemmata which appear on more than ten occasions, we find a further twenty such lemmata with a rhyme coefficient greater than 1: *largement* (23.1), *jugement* (17.3), *isnellement* (14.3), *joiant* (12.3), *tourment* (12.3), *talent* (12.0), *malement* (7.7), *longuement* (6.9), *couvent* (6.4), *errammment* (5.8), *avant* (4.2), *briément* (3.5), *argent* (3.3), *maintenant* (3.1), *gent* (1.9), *dolent* (1.9), *souvent* (1.9), *doucement* (1.7), *enfant* (1.4), and *fortement* (1.3). Given that this is also the case in the texts of the *R-corpus*, with fifteen of the forty-nine forms with a rhyme coefficient above 1 counting as "easy" rhymes, we can further resist the claims of previous scholarship that an abundance of such forms has anything to do with poor versificatory practice.

18 At the other end of the scale, words with rhyme coefficients close to zero are usually function words which occur with significant frequency across the whole text, but which are rare in rhyming position by dint of being forms which must usually be followed immediately by a more contentful word. A rhyme coefficient of less than one therefore most likely results from linguistic factors, rather than being the result of an authorial choice to include the word less frequently in rhyming position. Our attention here should thus be squarely focused on the forms with the highest rhyme coefficients.

Table 32. Counts and rates of lemmata which appear on ten or more occasions in the *J-corpus*, limited to the top twenty results.

Lemma	Line-final count	Line-final rate	Non-final count	Non-final rate	Rhyme coefficient
certainement	13	0.19%	0	0.00%	—
envie	11	0.16%	0	0.00%	—
demeuree	10	0.15%	0	0.00%	—
contree	12	0.18%	1	0.00%	92.5
plevir	12	0.18%	1	0.00%	92.5
vraiment	12	0.18%	1	0.00%	92.5
arreter	11	0.16%	1	0.00%	84.8
partie	21	0.31%	2	0.00%	80.9
commandement	10	0.15%	1	0.00%	77.1
appareiller	9	0.13%	1	0.00%	69.4
usage	9	0.13%	1	0.00%	69.4
tenter	13	0.19%	2	0.00%	50.1
pensee	19	0.28%	3	0.01%	48.8
endurer	11	0.16%	2	0.00%	42.4
chiere	10	0.15%	2	0.00%	38.5
corage	20	0.29%	4	0.01%	38.5
esbahir	9	0.13%	2	0.00%	34.7
debonnaire	8	0.12%	2	0.00%	30.8
matiere	8	0.12%	2	0.00%	30.8
sauvement	8	0.12%	2	0.00%	30.8

with particular frequency in the present participle form, for example, suggesting that in forming "easy" rhymes, Jehan may have extracted his forms from whatever verb was necessary to the narrative, rather than having a stock of verbs which he preferred to use in such situations.

The lack of much significant overlap between the forms with high coefficients in the *J-* and *R-corpora* therefore suggests that the rates at which certain words appear in rhyming position is as a direct result of authorial preference. By also comparing such forms with the highest rhyme coefficients to the *A-corpus*, it is possible to identify trends which assist us in attributing the authorship of the *J-corpus* to a single author. Table 33 displays the top twenty lemmata for the *A-corpus*, when ordered by rhyme coefficient, allowing us to make direct comparisons to the texts of the *J-corpus*.

Table 33. Counts and rates of lemmata which appear on ten or more occasions in the *A-corpus*, limited to the top twenty results.

Lemma	Line-final count	Line-final rate	Non-final count	Non-final rate	Rhyme coefficient
agreer	13	0.11%	0	0.00%	—
embraser	12	0.10%	0	0.00%	—
demeuree	11	0.09%	0	0.00%	—
droiturier	10	0.09%	0	0.00%	—
passion	11	0.09%	1	0.00%	84.9
debonnairement	10	0.09%	1	0.00%	77.2
lumiere	10	0.09%	1	0.00%	77.2
renommee	10	0.09%	1	0.00%	77.2
targer	10	0.09%	1	0.00%	77.2
oublier	9	0.08%	1	0.00%	69.5
honorer	17	0.14%	2	0.00%	65.6
contraire	10	0.09%	2	0.00%	38.6
droiture	10	0.09%	2	0.00%	38.6
vis	10	0.09%	2	0.00%	38.6
folie	19	0.16%	4	0.00%	36.7
retraire	19	0.16%	4	0.00%	36.7
saisir	14	0.12%	3	0.00%	36.0
prouver	9	0.08%	2	0.00%	34.7
remüer	9	0.08%	2	0.00%	34.7
courage	16	0.14%	4	0.00%	30.9

The first main observation of note is that only *demeuree* appears with an infinite rhyme coefficient across both the *J-* and *A-corpora*, suggesting that this lemma is particularly likely to appear in rhyming position across the work of multiple authors. The lack of overlap on other words, meanwhile, suggests that authorial preference may be more at play in these cases. Broader comparison reinforces this suggestion further, since several lemmata have much higher rhyme coefficients in the *J-corpus* than in the *A-corpus*, for example *commandement* (*J-corpus*: 77.1; *A-corpus*: 7.7), *endurer* (42.4 to 4.9), *contree* (92.5 to 11.0), and *tenter* (50.1 to 7.7). Looking in the other direction is similarly fruitful, with *mari* (1.9 to 11.6), *aimer* (1.9 to 9.9), *defendre* (2.8 to 14.2), and *tendre* (5.1 to 25.7) all appearing with markedly greater rhyme coefficients in the *A-corpus* than in those texts previously attributed to Jehan.

It is therefore clear that, for certain words, the rhyme coefficient differs a great deal between the *J-* and *A-corpora*, suggesting a role played by authorial preferences and therefore supporting the shared-author hypothesis for the *J-corpus*.[19] This does not appear to be the case for every word that appears at the rhyme, however, since despite the examples we have seen, there is still a correlation between the rhyme coefficients observed in the *J-corpus* and those observed for poems in the *A-corpus*. Examining those lemmata within the *J-corpus* which appear at the rhyme in at least one instance and at least ten times anywhere in the line across each of the *J-* and *A-corpora*, we find a statistically significant, weakly positive correlation (0.42489, p = 0) between the rhyme coefficients for individual lemmata. We can say with some certainty, then, that apart from a few examples, the lemmata which are more likely to appear at the rhyme in the *J-corpus* are also those which are more likely to do so in the *A-corpus*.

Producing a linear regression model for these data, however, reveals that the link between these two corpora is not straightforwardly one-to-one, in that it is not simply the case that lemmata tend to have the same rhyme coefficient across both corpora. Instead, a calculated regression line (intercept = 3.0553, gradient = 0.2910) suggests that a lemma in the *J-corpus* with a rhyme coefficient of 100 would be expected to appear with a rhyme coefficient of just 32.16 in the *A-corpus*. While we can therefore expect certain lemmata to have a greater rhyme coefficient across both the *J-* and *A-corpora*, Jehan reserves certain words for use solely, or primarily, at the ends of lines to a much greater extent than is observed in the more heterogeneous *A-corpus* where multiple competing preferences are in play from the corpus' various authors. This observation thus reinforces the suggestion that the *J-corpus* is a single-author corpus, whereas the *A-corpus* is not.

Figure 14 maps rhyme coefficients in both corpora as a scatter plot, along with the calculated regression line (dashed line), and a line indicating

19 We note, of course, at this stage, that the *A-corpus* contains *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus*, to which we have extended the shared-author hypothesis. The methods discussed in this section lend themselves best to examination of large-scale corpora, rather than individual texts, since the latter are likely to be influenced in their word choices by their content, and so it has therefore not been possible to apply this analysis to this smaller sub-group of texts. The analysis here therefore concerns the extent to which the *J-corpus* displays a coherence which the *A-corpus* does not, regardless of the identity of the authors of the *A-corpus* texts, and as such it is not problematic that the *A-corpus* contains texts which appear to also be the work of Jehan.

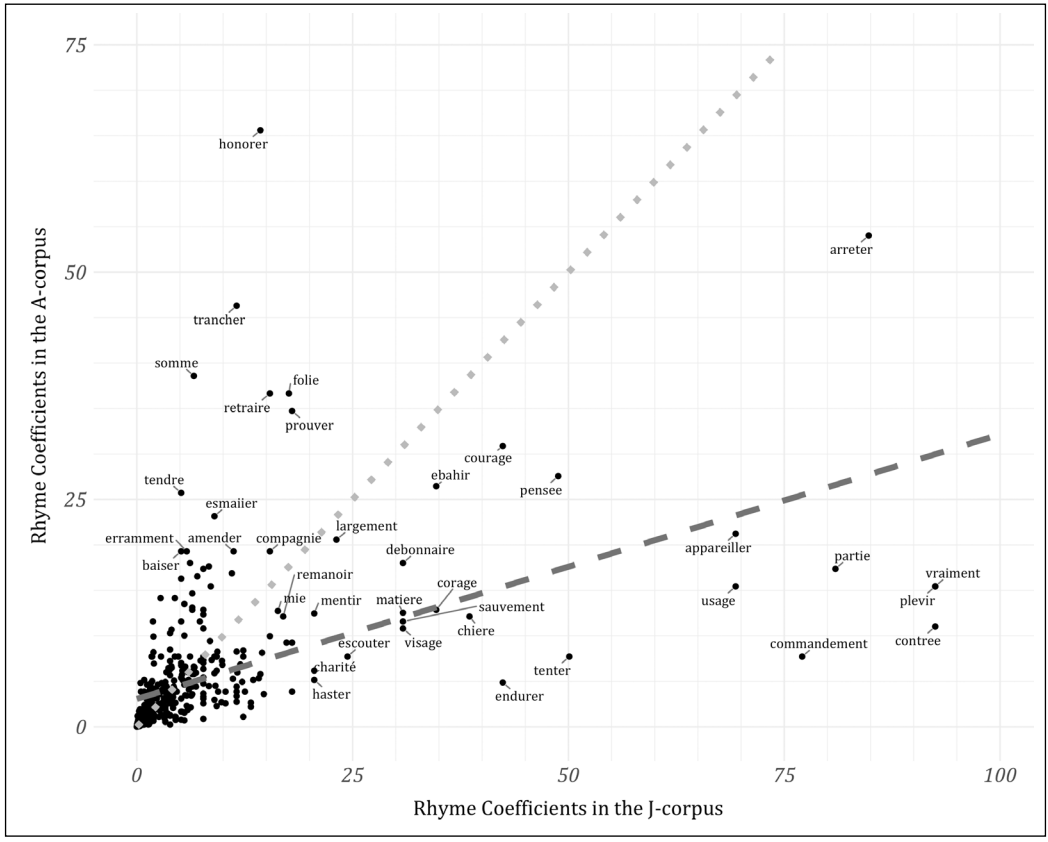

Figure 14. Scatter plot for the rhyme coefficients of lemmata found in rhyming position in the *J-* and *A-corpora*.

a putative one-to-one correlation (dotted line).[20] From it, we observe that there are indeed certain words which do share something resembling a one-to-one connection between their rhyme coefficients in the *J-* and *A-corpora*, including *largement* and *ebahir*. For these words, it seems, the rhyme coefficient may be relatively predetermined regardless of authorial identity, perhaps resulting from structural, linguistic factors determining where the term tends to fall, or indeed also from broader contemporary expectations of style for poems written in alexandrine monorhyme quatrains. Alongside this, we also get a clearer picture of the outlier status of lemmata such as

20 Several points have been labelled with their corresponding lemmata, but most have not, to ease readability. Lemmata have been excluded where the rhyme coefficient is infinite in either corpus, that is, where the only examples of a lemma are in the rhyming position and no examples are to be found elsewhere.

endurer, *tenter*, *commandement*, and *contree*, which have much higher rhyme coefficients in the *J-corpus* than in the *A-corpus*, and which Jehan therefore uses more frequently at the rhyme than comparable poets, as well as *tendre*, *somme*, *trancher*, and *honorer*, for which the opposite is true. Lemmata such as *matiere*, *corage*, and *appareiller*, meanwhile, stick closely to the calculated regression line, indicating that while they are common at the rhyme in both the *J-* and *A-corpora*, they are still much more likely to appear in rhyming position in Jehan's poems.

Thus, while the rhyme coefficients for some lemmata appear to be consistent across the work of different poets, or at least roughly in line with a consistent ratio of usage, authorial identity is a significant factor in determining rhyme coefficients. The examination of this feature has therefore added to the overall cohesiveness of the *J-corpus*, strengthening the shared-author hypothesis.

Admixtures

A final avenue of enquiry regarding Jehan's versificatory practices is that of admixtures, that is, cases where we observe an increased co-occurrence of one word in rhyming position when another word is also present in another rhyming position in the same stanza. Again, we look here at lemmata rather than word forms, to eliminate the possible impacts of orthographic and inflectional variation. Seeking those examples which are likely to be most assured in the trends they represent, our threshold for this analysis is to look only at lemmata which appear at the rhyme on five or more occasions in the *J-corpus*, since the certainty with which we can say that two lemmata are connected at the rhyme decreases the lower the frequency of those words.

Having isolated only those lemmata which are deemed to exist in sufficient numbers to reliably suggest the presence of an admixture, we can then calculate the ratio between the number of forms of a lemma appearing in the context of another lemma to those which do not. Thus, a lemma which appears five times alongside a context lemma and twice elsewhere has a ratio of 2.5:1 (given simply as 2.5). A lemma is said to exist in a prominent admixture with another lemma if this ratio is greater than 1, meaning that the lemma appears alongside its associated context lemma more frequently than it does elsewhere.

There are thirty-four pairs of lemmata which satisfy these criteria within the *J-corpus*, suggesting that they are pairings which the Jehan used together on a particularly frequent basis. The lemma *faire* has four associated lemmata which appear in prominent admixture with it (7 with *debonnaire*,

7 with *retraire*, 5 with *contraire*, and 5 with *exemplaire*),[21] all of which, we note, are used when *faire* is in its infinitive form. The same is not true of *avoir*, whose two associated lemmata (*voloir*, 1.4; *remanoir*, 1.2) can be used in a number of forms, whereas in the case of *dire* (*martire*, —; *despit*, 5; *Jesus-Christ*, 5; *respit*, 2.5), the necessary form at the rhyme is the infinitive for the first form, or the third person singular in the present tense or the past participle for the last three forms.

The case of *martire* is particularly notable in that this lemma does not appear in any rhyming position other than alongside *dire*, meaning that the ratio is infinite, and suggesting a strong association between these two words. The same can also be said of *Rome*, which only ever appears alongside forms of *somme*. In this case, we also find that the same connection exists in the opposite direction, with *somme* (5) appearing almost exclusively in the presence of *Rome*. Such bivalent pairs can be found elsewhere for *debonnaire–retraire* (both 1.67), *chemin–fin* (2 and 1.33), *manière–chiere* (2 and 1.5), and *mere–pere* (1.75 and 7.5). In some cases, as between *mere* and *pere*, there is a clear directionality to this bivalence, with the higher ratio associated with *pere* suggesting that *pere* is triggered by the presence of *mere*, rather than the other way around, whereas in other cases the pairs appear to be more mutually reinforcing.

In the case of *mere* and *pere*, as well as forming a bivalent pair, the two are also both context lemmata for *frere* (6 with *mere*, 2.5 with *pere*), while *mere* is also a context lemma for *clair* (5) and takes *amer* (1.75) as a context lemma. Of the remaining examples, we find four lemmata with *vie* as a context lemma (*compaignie*, 1.66; *folie*, 1.29; *envie*, 1.2; *Marie*, 1.13), four with *enfant* (*grant*, 2.2; *joiant*, 1.67; *avant*, 1.17; *maintenant*, 1.17), two with *certain* (*plein*, 6; *vilain*, 2) and *corage* (*usage*, 8; *image*, 2), and one each with *Marie* (*vilenie*, 4), *mie* (*compaignie*, 1.18), and *raison* (*maison*, 1.13). Interestingly, none of these lemmata are those which we identified as having a particularly high rhyme coefficient, thus reinforcing the suggestion that Jehan chose these forms for the rhyming position on the bases of the surrounding rhyming lemmata, rather than simply because he favoured the word in rhyming position.

To identify how unusual, or otherwise, these admixtures are, we must apply the same approach to the *A-corpus*. This time, examination at the level of a single text is more possible than it was for rhyme coefficients, such that here we exclude *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus*, to which we

21 *faire* is therefore the context lemma for these lemmata.

will return after discussion of the corpus' other texts. In this subset of the *A-corpus*, then, we observe thirty-two lemmata pairs which appear at the rhyme on greater than five occasions and which display a ratio of greater than 1. Five of these are also found for the *J-corpus*, namely *dire* twice (3.5 with *respit*, 2.3 with *martire*), *faire* twice (*debonnaire*, 2; *contraire*, 2), and the bivalent pair *mere–pere* (2.2 and 3). With the exception of *dire–respit*, we note that in these cases the ratios for the *A-corpus* are much lower than the associated ratios for the *J-corpus*, aligning with the assumption that texts in a single-author corpus are more likely to display consistent use of admixtures compared to a corpus of more diverse texts, written by authors with different preferences. This, along with the fact that the remaining twenty-seven lemmata pairs identified for the *A-corpus* (including a further six bivalent pairs) do not appear above our threshold in the *J-corpus*, strengthens the sense of cohesion for the *J-corpus*, and thus reinforces the shared-author hypothesis.

Turning our attention finally to *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus*, these texts are too short on their own to permit a fine-grain analysis of the admixtures that appear within them through the calculation of ratios, but taking them individually, it is nevertheless notable that all four texts contain a number of the same admixture pairs as are also found in the *J-corpus*. *Guillaume*, *Robert*, and *Richard* share two pairs (*compaignie–mie* and *dire–respit*), while *Guillaume* and *Robert* share a further three (*faire–debonnaire*, *maniere–chiere*, and *mere–pere*), *Robert* and *Richard* a further two (*avant–enfant* and *grant–enfant*), and *Robert* and *Jehan Paulus* a further four (*compaignie–vie*, *folie–vie*, *maison–raison*, and *Marie–vie*). *Guillaume* displays three pairs (*dire–martire*, *dire–Jesus-Christ*, and *enfant–joiant*) which do not appear in the other texts, *Robert* contains seven (*clair–mere*, *dire–despit*, *frere–mere*, *mere–amer*, *plein–certain*, *Rome–somme*, and *vilain–certain*), *Richard* contains three (*avoir–remanoir*, *avoir–voloir*, and *faire–contraire*), and *Jehan Paulus* contains just one (*envie–vie*). There are thus only nine lemmata pairs of the thirty-four that meet our thresholds in the *J-corpus* which do not appear in any of these four texts.

In total, *Robert* displays the greatest alignment with the *J-corpus*, aligning for eighteen pairs compared to eight in *Guillaume*, seven in *Richard*, and five in *Jehan Paulus*.[22] The utility of this method is ultimately limited for these short texts, but it nevertheless still allows us to strengthen the growing body

22 Explanation for the latter case in particular likely stems from the fact that *Jehan Paulus* is simply much shorter than *Guillaume*, *Robert*, and *Richard*.

of evidence that they ought to be attributed to Jehan, particularly for *Robert*. From a methodological perspective, meanwhile, there is an indication that this approach may be more usefully applicable for questions of authorship concerning longer poetic texts.

To conclude this chapter, we have seen that while historical claims in the scholarship as to the quality of Jehan's poetry are unhelpful, they nevertheless point to useful avenues through which his versification can be considered in a digital poetic analysis. Often, Jehan's tendencies come close to those observed in the comparable poems of the *A-corpus*, and yet for each of the metrics considered, the texts attributed to Jehan have been shown to be qualitatively and quantitatively distinct from them such that, despite their individual diversity, we may nevertheless still assert that a digital analysis of the versificatory practices of the *J-corpus* contributes to an overarching sense of unity, itself supporting the shared-author hypothesis.

Building the Case for the Shared-Author Hypothesis

Through the digital analysis of versificatory features in this chapter, including the rates of masculine and feminine rhyme use and the frequency of so-called "easy" rhymes, we have reinforced the internal coherence of the *J-corpus* when compared to the metrically and generically aligned texts of the *A-corpus*. We have also observed that, for these features, *Guillaume*, *Robert*, *Richard*, and *Jehan Paulus* sit squarely within the distribution of the *J-corpus*, reinforcing the single-author hypothesis. The examination of rhyme coefficients, while not suitable as a means of examining individual short texts, has nevertheless also added to the internal coherence of the *J-corpus*, and the distinctiveness of Jehan's apparent preferences when selecting a word to appear at the rhyme. The most concrete evidence to support our new attributions, meanwhile, comes from the analysis of admixtures. *Robert* displays the strongest overlap with the *J-corpus* for such pairs, while *Guillaume*, *Richard*, and *Jehan Paulus* also display several admixtures which are hallmarks of the *J-corpus*, reinforcing the case for their inclusion, and suggesting that the *remanieur(s)* of *Richard* and *Jehan Paulus* left the ends of lines largely intact.

Methodological Contribution

This chapter introduces a series of bespoke metrics, namely the concept of a *rhyme coefficient* and the analysis of admixtures, which have been developed to align with the specificities of poetry written in monorhyme alexandrine quatrains but have applicability in other contexts. By calculating how much more likely a word is to appear in rhyming position than elsewhere in the line, as well as by identifying pairs of words which are more likely to appear together, we can identify authorial habits that would remain obscure on a linear reading, and which nevertheless contribute to an individual versificatory fingerprint.