Shakespeare by the Numbers: On the Linguistic Texture of the Late Plays[1]

Michael Witmore (Carnegie Mellon University)
and Jonathan Hope (Strathclyde University)

> If I could write the beauty of your eyes
> And in fresh numbers number all your graces,
> The age to come would say 'This poet lies:
> Such heavenly touches ne'er touch'd earthly
> faces'...
> — Sonnet 17
>
> Return, forgetful Muse, and straight redeem
> In gentle numbers time so idly spent...
> — Sonnet 50

Since the late nineteenth century, critics have tried to group some or all of the plays Shakespeare wrote late in his career—*Coriolanus*, *Cymbeline*, *The Winter's Tale, The Tempest, Pericles* (co-authored with George Wilkins), *Henry VIII* and *The Two Noble Kinsmen* (both co-authored with John Fletcher) —into a single critical category, usually on the basis of thematic, dramaturgical, or linguistic similarities among members of the group. While there is no consensus on which category is most appropriate for such a grouping (designations such as 'late plays,' 'romances,' and 'tragicomedies' have been proposed), there is nevertheless a persistent feeling among Shakespeare's readers that something distinguishes several of these plays from the others and that this 'something' ought to be the object of critical analysis. [2] At

---

[2] The first generic identification of some of these plays as 'romances' was made by Edward Dowden in 1877, drawing on the results of metrical analysis by Furnivall and his colleagues in the New Shakspere Society, established in 1874. Furnivall's work was exclusively formal, attempting to establish a chronology of Shakespeare's plays on the basis of metrical patterns. These analyses grouped four of the plays we now know as 'Late Plays' -- *Pericles, The Tempest, The Winter's Tale, and Cymbelene* -- for the first time in critical history, placing them together at the end of Shakespeare's career on the basis of his increasing use of hypermetrical lines. Dowden extended this chronological grouping with content-based criteria, stressing the presence of the sea, and lost children as contributors to the romantic element. In 1901, Thorndike added 'tragicomedy' to the mix, claiming the influence of Beaumont and Fletcher, and this re-casting of the late plays was strengthened by other critics, such as Bentley, who associated the plays with masque, and the purchase of the Blackfriars theatre (see Bentley's article in the first issue of *Shakespeare Survey*).   What started as an association of the plays by chronology, arrived at by a purely formal metrical test, quickly became one of genre and theme, with hitherto unperceived overlaps in

times the perception of such a similarity has been so strong that it has lead to categorical declarations of the sort made by Gerard Eades Bentley in the mid-twentieth century, who asserted that 'no competent critic who has read carefully through the Shakespeare canon has failed to notice that there is something different [about these plays].'[3] Philip Edwards (1958) went even further on this score, arguing that 'the [late plays] seem more closely related than any other group of Shakespeare's plays.'[4] Given that Shakespeare's first editors, Heminges and Condell, found no need for a fourth generic category when dividing the plays in Shakespeare's First Folio (1623) into histories, tragedies and comedies, Edwards' claims about the *definitive* unity of some additional generic grouping seem particularly extravagant. While Shakespeare's first editors were not entirely consistent in their use of early modern genre distinctions to group the 36 plays in their edition (the First Folio omits *Pericles* or *The Two Noble Kinsmen*), they were certainly no worse judges ('no competent critic...') than those who came after them. What, then, have critics been seeing in these plays for over a century that is so vivid, so pervasive, that it confounds the generic distinctions of some of Shakespeare's most intimate seventeenth-century contemporaries?

Genre distinctions are notoriously difficult to express in formulae or abstract terms, even if particular genre labels can be recognized and applied with little experience; genre critics since Aristotle have thus tended to use examples or tokens of a particular genus in order to express one or another feature of a type. Ever the biologist, Aristotle tended to think of genre in genus and species terms, and this particular habit of thought persists through most attempts to specify genre. A tragedy shows a character of high standing being laid low: the rule can be illustrated by any number of examples, of which Aristotle supplies two, Oedipus and Thyestes (*Poetics* 1453a.10-11).[5] Renaissance literary criticism followed the model of genre as genus or formal type, with theorists such as Guarini outlining typical features of a specimen within a genre such as 'tragicomedy' and then going on to provide concrete examples.[6] In some cases a

---

subject matter and concern being readily identified. For a full account of the critical history of this genre and relevant bibliography, see Barbara A. Mowatt, '"What's in a name?" Tragicomedy, Romance, or Late Comedy' in *A Companion to Shakespeare's Works: The Poems, Problem Comedies, Late Plays*, ed. by Richard Dutton and Jean E. Howard, 4 vols (Oxford: Blackwell, 2003), IV, 129-49. See too Russ McDonald, *Shakespeare's Late Style* (Cambridge: Cambridge University Press, 2006).

[3] Gerard Eades Bentley, 'Shakespeare and the Blackfriars Theatre,' *Shakespeare Survey*, 1 (1948), 47-48.

[4] Philip Edwards, 'Shakespeare's Romances,' *Shakespeare Survey*, 11 (1958), p. 1.

[5] Aristotle, *The Rhetoric and Poetics of Aristotle*, trans. by Ingram Bywater (New York: McGraw Hill, 1984), pp. 238-39.

[6] Giambattista Guarini, *Compendio della poesia tragicomica* (Venice, 1601).

particular 'token' might be taken to exemplify the genre in all of its attributes, as was perhaps the case with John Fletcher's *The Faithfull Shepherdess,* which was designed to introduce English audiences to Guarini's tragicomic ideal in all of its aspects. For Fletcher, this play and others of its kind possessed certain attributes and lacked others: 'A tragi-comedie is not so called in respect of mirth and killing, but in respect it wants deaths, which is enough to make it no tragedy, yet brings some neere it, which is inough to make it no comedie'.[7] Some of these features have been attributed to Shakespeare's later plays, leading critics to identify them as 'tragicomedies' as well. We will return to this argument later, but it is enough at this point to notice that Fletcher's definition of genre not only specifies what must be in a play to qualify it for membership in a genre, but also what it must lack. The notion that genre is a set of co-ordinated presences and absences of features is one we will take up in the conclusion of our paper below.

Of course, criticism after the Renaissance has developed further critical categories with which to approach the question of genre in Shakespeare's plays, relying variously on biographical, formal, historical and dramaturgical criteria for membership in a generic group. The Victorian critic Edward Dowden, for example, felt that *The Winter's Tale*, *Cymbeline* and *Pericles* were alike in their tendency to retreat to a serene, enchanted world, one that Shakespeare craved after having plumbed the depths of the tragedies earlier in his career.[8] More recent critics have emphasized the seemingly mythical or folkloric atmosphere of these plays, suggesting that the improbabilities of plot found in the later works evidence a desire to enact a messianic quest ending in deliverance or to test the nature of reality by framing it in a deliberately artificial way.[9] Arguments about tone and atmosphere have been accompanied by

---

[7] John Fletcher, *The Faithfull Shepheardesse* (London: Edward Alde, 1610). It is worth pointing out, however, that Fletcher's introduction is an attempt to justify as literature a play that had failed as drama -- and that the plays he went on to write with Beaumont and others, generally taken as exemplars of English tragicomedy, are very different in form and tone from Guarini's model, and Fletcher's own solo attempt at it (see Gordon McMullan and Jonathan Hope, 'Introduction' in *The Politics of Tragicomedy: Shakespeare and After,* ed. by Gordon McMullan and Jonathan Hope (London: Routledge, 1992), pp. 3-7.
[8] Edward Dowden, *Shakspere: A Critical Study of his Mind and Art , 3rd ed. (1899)* (New York: Barnes and Noble, 1967), p. 415.
[9] The interpretation of romance as messianic struggle is advanced in Northrop Frye's *The Anatomy of Criticism* (Princeton: Princeton University Press, 1997). The association of the late plays with folklore appears in Stanley Wells, 'Shakespeare and Romance,' in *Later Shakespeare, Stratford-Upon-Avon Studies 8* (New York: St. Martin's Press, 1967), 49-80; 53. The association of romance with improbable tall tales told by women and gossips is explored in Mary Ellen Lamb, 'Engendering the Narrative Act: Old Wives' Tales in *The Winter's Tale* and *The Tempest*,' *Criticism* 40, no. 4 (1998), 28-43, while the notion that

inquiries into sources and dramaturgical techniques. Barbara Mowat, for example, has argued that Shakespeare's later plays get their distinctiveness from their source (the Greek novel), suggesting moreover that the 'family resemblance' among these plays defies generic labels -- although the plays have a tendency to mix disjunctive presentational and representational modes (essentially, modes of telling and showing).[10] In having recourse to an explicitly Wittgenstinian notion of 'family resemblance,' Mowat is bearing out Fletcher's insight that generic identity is not just a bundle of *sine qua non* features that are 'present' in every token member of a type, but that membership and type only become intelligible within the fullness of a highly variable group.

Whatever it is that critics are reaching for when they see a generic or 'familial' resemblance among different late Shakespeare plays, then, perhaps that 'something' is not a roll call of features (to use the analogy, dimples, dark complexion, curly hair), but a variously distributed set of features common only to the entire group, a set that can further be contrasted in their absence with those that appear in members of another group. The late plays are 'romances' because they bring families together; they are 'tragicomedies' because they bring happiness only after trial. These are real features that can be pulled out of many of the plays listed above, but they do not appear in exactly the same way in each play. Indeed, the basic difficulty Mowat identifies in past attempts to name a fourth genre of Shakespeare's plays may be a difficulty common to all attempts to create stable genre definitions; and yet such distinctions are nevertheless intelligible in practice and can be debated in rich detail, as the critical history rehearsed above suggests. In what follows we would like to take a different approach to the problem of naming a fourth genre in the Shakespeare canon, one that attempts to re-describe the difference critics have sensed between 'romances' or 'tragicomedies' and the more traditional First Folio genres, not by qualitative analysis of tone, atmosphere or plot, but by quantitative analysis of linguistic features. Like Fletcher, we believe that this analysis shows genre to be a fully relative construct in the sense that it involves the strategic collocation of features common to other genres, features that are emphasized or missing in a coordinated

---

these plays seek to question or establish a 'reality principle' can be found in Howard Felperin, *Shakespearean Romance* (Princeton: Princeton University Press, 1972), p. 50.

[10] See Barbara Mowat, *The Dramaturgy of Shakespeare's Romances* (Athens: University of Georgia Press, 1976), p. 36, 69 and '"What's in a Name?": Tragicomedy, Romance, or Late Comedy,' p. 134. Mowat credits the adaptation of Wittgenstinian 'theory of resemblance' to genre theory to Alastair Fowler's *Kinds of Literature: An Introduction to the Theory of Genres and Modes* (Cambridge: Harvard University Press, 1982).

way. But we also believe that such analysis calls attention to a heretofore invisible set of dramaturgical strategies at work in the late plays, strategies that mobilize language so consistently and on such a pervasive verbal level that their effects have gone unnoticed by more traditional literary genre criticism.

<div align="center">

*

*   *

</div>

In order to conduct such analysis, we have made use of a computer text analysis tool called Docuscope. Docuscope is a text analysis and comparison program developed by a team of researchers at Carnegie Mellon University, Pittsburgh, USA. The program was designed for use in writing/rhetoric classes (as understood in the North American model), and aims to allow tutors to make fast, statistically reliable comparisons between texts written by students. To accomplish this function, the program consists of a first stage of textual analysis tools -- essentially, smart dictionaries -- that comb texts for strings of words which are then assigned to a predetermined set of rhetorical categories. (As we will see below, these dictionaries are ultimately an expression of the designers' views on how language works; they are thus, in effect, a working rhetoric.) The frequency results from this combing are then displayed in various graphical formats by the program's sophisticated and user-friendly visual interface. These results can also be exported into statistical analysis packages for more complex statistical analysis.[11]

The rhetorical categories used by Docuscope have been explored in practical and theoretical terms by the designers in a text entitled *The Power of Words: Unveiling the Speaker and Writer's Hidden Craft*.[12] In this text, the primary architect of the dictionaries, David Kaufer, and his colleagues support the choices made in creating Docuscope by advancing an essentially phenomenological view of how different types of language shape or recreate an experientially vivid world for an engaged reader or listener. Here we can provide a brief overview of the categories that the program assumes are working to accomplish this goal; further below we will describe how these categories relate to 'tragicomedy' or 'romance' and to our own sense of what 'genre' must be in the theater. As it scans an electronically formatted text, Docuscope

---

[11] Those interested in examining the precise statistical procedures and data used in this paper are invited to contact the authors.

[12] David Kaufer, Suguru Ishizaki, Brian Butler, Jeff Collins, *The Power of Words: Unveiling the Speaker and Writer's Hidden Craft* (London: Lawrence Erlbaum, 2004).

organises word strings into three high-level categories (termed 'clusters'). These clusters correspond to a theoretical model of the effects texts seek to have on their readers developed from a Hallidayan theoretical base (51-5). The model groups rhetorical effects as follows:

1   Internal Perspectives: those sequences of words -- 'strings' in the parlance of Docuscope's creators -- that are used to communicate the interior mind of the writer, or a character, to the reader (for example, grammatical first person features, expressive and subjective vocabulary, complex tense/aspect constructions that imply a relation between two different times entertained in a narratorial consciousness).

2   Relational Perspectives: strings used to connect readers to the representations within a text. For example, aspects of the language that recognize or engage shared processes of reasoning between reader and author or their implied social ties; also strings that orient readers to other locations in the text.

3   External Perspectives: strings that refer out of the text, but to the physical world (rather than the metatextual or social values of 'Relational Perspectives'). These strings include types of description of physical objects, strings describing the spatial location of objects, and representations of movement through space and time.

A fundamental assumption of the creators of Docuscope is that texts will vary in the frequency with which they employ string types from each of these clusters depending on the writer's purpose. Variation visible at the highest level is rather crude, but a broad expectation would be that fiction and autobiography ought to be high in cluster 1, while instructional writing (technical manuals for example), ought to be high in clusters 2 and 3. This is hardly an impressive or surprising finding. The real value of Docuscope comes, rather, in the fine-grained analysis that becomes possible when comparisons are made at a far more detailed level of string category. Within Docuscope, the three high-level clusters are further divided into six 'families', and then further into 'dimensions' that contain multiple 'language action types' (LATs) which allow a high degree of interpretive distinction to be made in the analysis of texts. An example, adapted from the authors' exposition of their categories (59-88), is given below:

Cluster 1: Internal Perspectives

    Family 1:   Interior Thinking (strings involved in exposing the audience to the activity of another mind)

        Dimension 1: <u>First person</u>

LAT: *Grammatical first person* e.g. first person pronouns -- 'I', 'me', 'my', 'mine'
LAT: *Self-disclosure*: e.g. first person plus simple past -- 'I went'
LAT: *Autobiographical reference*: e.g. first person plus habitual past verb phrase -- 'I used to go'

Dimension 2: <u>Inner thinking</u>
LAT: *Private thinking* e.g. private cognition or thinking verbs -- 'contemplate', 'decide', 'discover'
LAT: *Disclosures* e.g. verbs of speaking, some adverbs -- 'confessed', 'acknowledged', 'personally', 'frankly', 'tellingly'
LAT: *Confidence* e.g. 'that'-complement, situational 'it', existential 'there': 'I know that the box is upstairs'; 'It's a boy!'; 'There's an apartment down the street that you can afford'
LAT: *Uncertainty* e.g. adverbials -- 'allegedly', 'to the best of my knowledge', 'nearly', 'almost'

Dimension 3: <u>Positive affect</u>
LAT: *Think positive* -- e.g. 'loving', 'succulent'; attitudinally marked prepositions: 'up'

Dimension 4: <u>Negative affect</u>
LAT: *Think negative* -- e.g. 'too many', 'too much'; attitudinally marked prepositions: 'down'

Note how the shift from 'family' to LAT allows the analyst to make some relatively fine distinctions in the stylistic effects produced by texts, particularly when each of these distinctions can be counted and compared with one another across a large corpus of texts. Within the Dimension 'First person', for example, Docuscope distinguishes bare first person pronouns such as 'I', 'me', 'mine', which produce a simple point of view within a text, from 'pronoun + tensed verb' strings ('I went...', 'I'll go...'), which produce a particularised consciousness, self-realised in terms of time. Look at the difference between these examples:

I often use facts about Einstein's laws in my work

I often used facts about Einstein's laws in my work

In the first example, 'I' appears with a simple present tense and establishes that the text is written by a specific individual, but not much else. The second example, however, complicates the point of view presented considerably: the self constructed in the sentence is one who looks back on a past state of selfhood, analyses itself, and discloses something about that analysis. As mentioned above, Docuscope assigns strings to the various dimensions identified by the

development team using a set of dictionaries that seek out word clusters (so for example, 'I' on its own will be assigned to the LAT *Grammatical first person*, but 'I' followed by a past tense verb will instead be assigned to the LAT *Self-disclosure*). Words can only be counted as part of one string, and the program always counts the longest possible string. In practice, Docuscope is capable of parsing or categorizing an enormous amount of strings within the English language. When used to 'read' the Frown Corpus, a multi-genre corpus of contemporary American English of approximately 1 million words, the version of Docuscope used for this paper classed 76% of its contents; when used to 'read' the plays of the First Folio, Docuscope classified approximately 75% of all the words or phrases, which is approximately the average 'find rate' of the tool for most large collections of texts it is used to study.[13]

<div align="center">

*

*   *

</div>

Our initial research with Docuscope involved the analysis of the First Folio of 1623, a promising starting point for our work, we felt, because the Folio presented us with a concrete editorial decision about generic divisions (the editors sorted the plays by Comedy, History, and Tragedy). Furnished with that initial qualitative decision, we could then attempt to re-describe it in statistical terms and see if we learned anything new. Our initial findings were reported in 2004, and we have since embarked on several research projects examining Shakespeare's genres and the historical development of his style. In the course of this research, we have learned that Docuscope, a device created to teach writing four hundred years after Shakespeare wrote his plays—and by designers with no academic interest in Shakespeare or early modern drama—can nevertheless 'see' the First Folio genres, sometimes quite distinctly.[14] It should be said here that if Docuscope were redesigned to count different things (different strings of words, based on a different interpretation of how texts work), we would *still* expect it to find statistically significant patterns of string usage that differentiate the various genres; this is because,

---

[13] Many of the unclassifiable words were place or person names, non-standard English words, hyphenated words, or extremely common words such as 'and' or 'that.' The set of dictionaries employed by Docuscope is under continuous revision, which is why we asked that a state of those dictionaries (May 14, 2006) be 'frozen' so that our future analyses of texts (for example, the Middleton corpus) could be made under identical conditions and the results compared.

[14] The results of our earliest attempt to discriminate the Folio genres using Docuscope appear in Jonathan Hope and Michael Witmore, 'The Very Large Textual Object: A Prosthetic Reading of Shakespeare'. *Early Modern Literary Studies* 9.3 / Special Issue 12 (January, 2004): 6.1-36. URL: http://purl.oclc.org/emls/09-3/hopewhit.htm.

assuming enough different types of items are being counted, such items ought still to show signs of coordination because (1) theatrical language is saturated, potentially at every level, with dramaturgical effect and (2) the particular type of story that critics recognize as 'history,' 'comedy,' 'romance,' or 'tragicomedy' makes certain material demands on the speakers of a play and thus the language they must use to advance the story. Our conclusions about Shakespeare so far are based on a micro-analysis of LAT frequencies rather than assumptions about how such LATs may group into Families or Clusters. This is a crucial point, since it allows us to remain agnostic on the question of whether or not the architecture of categories used by Docuscope represents the 'best' or most 'functional' way to organize language. Ultimately, the statistical patterns we are interpreting are used as prompts or pointers to particular uses of words that remain invisible in the linear flow of reading, but may nevertheless have some more general rhetorical and dramaturgical function. The art here is not in crafting the 'best' categories, but in seeing how the coordinated presence and absence of various types of words (type being itself an interpretation) might function in a concrete, dramaturgical setting. We call this assumption of pervasively coordinated rhetorical patterning the 'principle of dramaturgical saturation.' In adopting it, we are assuming that the material constraints of the theater and the temporal entailments of certain narrative conventions create a situation in which even the smallest particles of language are regularly pressed into the service of telling a particular type of story.[15]

   With these caveats in mind, we move to the questions that motivate this essay: What could a literary linguistic analysis using Docuscope provide in support of the claim that the late plays constitute a distinct group? We know from our previous work with Docuscope that the three Folio genres have distinct linguistic features: do the late plays, identified as such first of all by formal metrical tests, also share formal linguistic features that characterise them, aside

---

[15] In one of our experiments, we asked Docuscope to count the number of words beginning with the letters 'm' 'o' 'a', and 'i' and then to use only these counts to produce a recipe for the various genres. Unsurprisingly, it did. A factor analysis indicated that comedies are distinguished by a relative lack of words beginning with 'o' combined with a relative abundance of words beginning with 'm' and 'i'. The experiment showed us that with enough observations (counted items), a statistical portrait of a genre can be created out of almost anything; the trick is knowing what to make of what you count -- being able to connect a quantity of something counted to an activity that is meaningful (e.g., a dramaturgical event such as dreaming, narrating, arguing, etc.). Given the principle of total dramaturgical saturation, we would hold that eventually even the lack of 'o' in comedy could be explained in terms of the dramaturgical features of the genre. It would take something like a virtuoso form of Vico's 'maker's knowledge' with respect to writing, performing and staging plays to get at what these 'o' words were *not* doing in comedies, however.

from their shared content concerns with the sea, lost children, reunited parents, and so on? To what extent can a quantitative analysis of different classes of language yield clues about (1) the dramaturgical strategies that may have colored the composition of these plays or (2) the interpretive criteria that have led critics to make such claims as those offered above? What, further, can such analysis reveal about the degree to which these plays are an amalgam of previous genres such as comedies and tragedies, as the term 'tragicomedy' suggests?

To answer these questions, we are going to rely on two types of evidence. The first consists of findings based on a statistical procedure known as 'factor analysis,' a procedure that is blind with respect to genre judgments and instead simply looks for correlations among frequently used types of words or phrases. To the non-statistician, this procedure might be understood through the following example: if you were to create 36 unique decks of playing cards at random from an infinite pile of cards, you would expect the resulting decks to differ from one another in certain ways. A factor analysis would compare the relative frequency with which different 'types' of card (twos, threes, queens, etc.) occur in every deck, working then to characterize any pattern that cuts across all of the decks. For example, it might turn out that decks containing lots of fours and fives have almost no sixes: this 'factor' (the metavariable that coordinates 'having' fours and fives with 'lacking' sixes) could then be used to 'rate' all of the decks in order, from those that are highest in this factor (they have lots of fours and fives but no sixes) to those that are lowest (they have almost no fours and fives but lots of sixes). If you think of each Shakespeare play in the Folio as a deck of cards in this example, and the individual cards (two of hearts, two of spades, two of clubs, two of diamonds) as being like the words (or strings of words) that Docuscope is organizing into types and then aggregating as factors, you begin to understand the nature of the procedure. Sometimes a factor will arrange the plays in a way that make no sense to the observer—it is picking up on a pattern that has nothing to do with genre. [16] But often, a factor ends up 'sorting' the plays in such a way that almost all specimens of a particular Folio genre appear at one end of a particular spectrum, something we discovered was the case with Shakespeare's histories; this is a particularly

---

[16] In our procedure, Docuscope generates factors without any knowledge of the Folio genre divisions and then, once these are obtained, automatically screen the results for those factors that separate out the genres we have identified; this latter procedure is mediated by a statistical algorithm, the Tukey Test for statistical significance, eliminating our role in the selection procedure. Since these algorithms are well known in statistics, our results could be independently reproduced.

powerful statistical finding, precisely because the computer knows nothing about the target genre grouping that is being sought.

The second type of evidence we will be introducing is that resulting from something called analysis of variance (ANOVA) among single variables, which looks at the relative frequency of particular types of words or phrases in groups that we have already sorted into generic types. Unlike factor analysis, which characterizes those broad patterns that emerge from frequent and consistent use of certain types of words or strings of words, single variable analysis can hone in on those strings and words (LATs) that are (1) used less frequently but still occur in very different proportions across a predefined group of genres or (2) on LATs that appear in dense clusters rather than a consistent spread throughout the entire text. Taken together, the two techniques give us a detailed picture of how the words identified by Docuscope's counting procedures are really working to separate out the genres.[17]

<center>*</center>

<center>*   *</center>

We began by dividing the Folio plays into four groups -- the original comedies, histories and tragedies of the 1623 editors Heminges and Condell -- and a fourth called 'late plays' that consists of *Cymbeline*, *The Winter's Tale, The Tempest*, and *Henry VIII*.[18] Because Docuscope looks for and correlates over 90 LATs, we further sub-divided the plays into smaller chunks -- of 2500 and 7500 words   -- before dropping them into our four genre containers (C, T, H, L). This chunking procedure is normal statistical practice: ideally, the number of individuals in a population to be analysed (i.e. the number of plays, or chunks of plays) should exceed the number of categories being searched for. The 36 plays of our sample are thus too small a population to search for over 90 LATs with statistical rigor. By automatically chunking the plays into 1000, 2500 and 7500 word sections, we make the procedure statistically respectable, and also ensure that items identified as statistically significant in their distribution are consistently used throughout the texts of the plays identified. This final

---

[17] The decks are somewhat loaded in this second type of analysis, since in single variable ANOVA Docuscope is only re-describing, as it were quantitatively, a set of generic distinctions that have already been assumed in the initial constitution of the four generic groups.

[18] The electronic text used in the analysis was that of the Complete Moby(tm) Shakespeare. Speech prefixes, stage directions, and act/scene divisions were 'stripped' from the plays for the purpose of analysis, since some of these -- particularly speech prefixes such as 'King' -- were dead giveaways to Docuscope when it went to produce factors for genre. We plan to write about this decision to exclude non-spoken text and its implications for our theory of genre and dramaturgy in a future article.

point is important: 'chunking' means that we only identify as significant LATs which are consistently used across the whole text of a play (they appear in a high number of the chunks of a play). LATs that appear in only a few chunks of a play, even if they have a very high frequency there, are excluded.

Using the results of Docuscope's counting procedure, we employed multivariate statistical analysis to identify factors – again, metavariables that capture coordinated variations in smaller variables (high amounts of A appear in plays with low amounts of B and C) -- that differentiated the genres from each other, particularly those that distinguished the late plays from *all* of the other genres contained in the original Folio genre divisions.   We then examined the LATs that Docuscope relied upon in making such discriminations, returning to read the texts of the plays in order to see what these particular types of words or phrases were doing rhetorically, dramaturgically, thematically, and the like. In the process of re-reading the plays with the significant LATs highlighted for easy recognition, we consulted the results of the single variable analysis (ANOVA) to see which individual LATs were reliably present or absent in the late plays in comparison with all the other genres. Finally, we returned to the factor analysis to see what aspects, if any, the late plays shared, heightened or excluded from the previous three genres -- in effect, looking for the 'family resemblance' within and across the late plays in the larger context of the Folio genres.

Our most significant finding was that multivariate analysis could indeed identify factors which separate the late plays as a distinct linguistic group from the other Folio genres. In our analysis, two factors emerged as statistically significant, producing a clear separation of the late plays from all of the other genres -- that is, the ranges of variation for these factors showed no overlap between the late plays and those of the other genres. This is a striking finding: Docuscope, and multivariate analysis, have identified a familial linguistic resemblance between plays which were not generically associated until the late nineteenth century. Further, the single variable analysis (ANOVA) supported the initial findings by factor analysis (which is initially "blind" with respect to genre), suggesting that on both the pervasive and local levels, certain types of language were being used in ways that created a statistical footprint for the late plays different from that of any of the earlier Folio genres. In the following discussion, we first describe the two comparative factors which distinguish the linguistic texture of the late plays from that of the rest of Shakespeare's work. We follow this with a brief examination of those aspects of comedy, tragedy and history that are only partially present in the late plays,

concluding with a discussion of how multivariate analysis of linguistic structures might compliment traditional understandings of genre in future studies of early modern texts.

CF7: 'Focalized Recollection'

The first comparative factor (CF7/2500 in our analysis) is closely involved with the communication of the past. Two LATs are highly significant in constituting this factor: *Narrative Verb (Narrating)* and *Think Back (Retrospect)*. *Narrative Verb* strings consist of simple past tense verbs. Note the following example, which like others we will be discussing, contains the LATs under discussion underlined in the passage:

> I was by at the opening of the fardel, <u>heard</u> the old shepherd deliver the manner how he <u>found</u> it [...] only this methought I <u>heard</u> the shepherd say, he <u>found</u> the child
> (*The Winter's Tale* 5.2.2-6/TLN3013-7)[19]

As it does in hundreds of instances across the late plays, Docuscope is tracking the past tense narration of events through simple past constructions like 'heard', 'found', and 'heard'. The linguistic effect of this type of past narration is to focus on the event itself, which is usually presented as a discrete, completed action. The narrator, though clearly present, is not made a vivid part of the events described or the context in which they take place: to a large extent, the past events are presented directly, without narratorial comment.

The second LAT on CF7, *Think Back*, involves a more complex presentation of past events. *Think Back* is triggered by past forms of *to be* and some auxiliary constructions, and under single variable ANOVA analysis, proved to be a regular feature of the late plays, but relatively absent in the other three genres:[20]

> but the changes I perceived in the king and Camillo <u>were </u>very notes of admiration: they seemed almost, with staring on one another, to tear the cases of their eyes; <u>there was</u> speech in their dumbness, language in their very gesture; they looked as <u>they had</u> heard of a world ransomed, or one destroyed: a notable passion of wonder appeared in them;

---

[19] Quotations are from the Moby Text Shakespeare, as analysed by Docuscope. As this is an electronic text, line numbers are unreliable, so we have also included Folio through-line numbering (TLN).
[20] For histories, the median and mean frequencies of *Think Back* for 2500 word chunks were 0.57 and 0.58; for comedies, 0.52, 0.54; tragedies: 0.52, 0.54; late plays 0.72, 0.74.

but the wisest beholder, that knew no more but seeing, could not say if the importance <u>were</u> joy or sorrow

<div align="right">(The Winter's Tale 5.2.8-18/TLN 3020-8)</div>

Here continuous states rather than discrete events are communicated and, significantly, the past state is presented via an explicit narratorial present. That is, the past is filtered through the consciousness of the narrator who is, quite frequently, making some sort of implied or explicit emotional evaluation of the events described. The distance of one time from another (narrative present to historical past) becomes less important than the emotional effects created by that distance. Note the *Think Back* verbs tagged in the following:

First Gentleman The dignity of this act <u>was</u> worth the audience of kings and princes; for by such <u>was</u> it acted

<div align="right">(The Winter's Tale 5.2.72-74/TLN 3088-89)</div>

and now note the effect of shifting them to *Narrative Verb*:

The dignity of this act <u>justified</u> the audience of kings and princes; for such <u>acted</u> it

Here, a shift away from past forms of *to be* into simple past forms of main verbs strips away the focalising presence of the First Gentleman: the actual events (the justification, the acting) are arguably more vivid, but we lose the sense that we experience them through his consciousness and judgement.[21]

This comparative factor then, shows us that one of the most striking linguistic differences between the late plays and the other genres is that the late plays show an increase in the direct representation of the past (the LAT *Narrative Verb* is more frequent), *but also, at the same time*, an increase in the focalised representation of the past, where the representation is directed through the explicitly realised consciousness of the narrator (via the LAT *Think Back*). Past events could therefore be argued to be more important in the late plays than in the early work -- but this increase in concern with the past is also accompanied by a shifting relation to the past: the past is important in as much as it is relevant to, and contextualised via, the present. The increased use of language designed to focalize past events in the late plays

---

[21] On focalization, see Mieke Bal, *Narratology*, 2nd edition (Toronto: University of Toronto Press, 1994), pp. 142-60.

appears, in our view, to be connected with these plays' thematic preoccupation with family re-unions and plots of wandering, both of which place a dramaturgical premium on the emotionally charged narration of past events with an eye toward a redemptive, corrected present.

CF5: Subjective Particularization

The second comparative factor that distinguishes the late plays (CF5/2500 in our analysis) correlates the presence of the two LATs *Asides* and *Verb State*. The Language Action Type known as *Asides* identifies strings that introduce or end digressive comments: 'by the way...', 'anyway', 'as an aside', 'let me digress', 'incidentally', 'to return', 'at any rate'. A significant marker of asides for the developers of Docuscope is 'which' as a non-restrictive relative pronoun -- that is, a relative introducing information which is non-essential, hence an aside. Indeed, when Docuscope analyses Shakespeare, *Asides* strings are almost wholly made up of 'which' forms, and there is a significant increase in these strings over his career, and particularly in the late plays.[22]

We have already identified (on CF7) a trend in the language of the late plays that sees Shakespeare increase the frequency of two types of past narration, and in particular *Think Back*, which presents the past filtered through the explicit consciousness of a narrator. It is possible to identify a similar effect in this increase in *Asides* over the course of Shakespeare's career. We will take as an example *The Winter's Tale,* act 5, scene 2, where the discovery of Perdita is narrated by one 'gentleman' to two others, employing a linguistic style which relies heavily on *which*-relativisation. Note in the following how the relative clause always introduces some form of subjective judgment or evaluation on the part of the speaker:

Then have you lost a sight, <u>which was to be seen, cannot be spoken of</u>

(42-3/TLN 3052-3)

now he thanks the old shepherd, <u>which stands by, like a weather-bitten conduit of many kings' reigns</u>

(54-6/TLN 3064-5)

---

[22] Here too single variable ANOVA for 2500 word chunks showed a significant increase of Aside words in the late plays. For histories the median and mean frequencies of Aside were 0.17, 0.18; comedies, 0.14, 0.15; tragedies, 0.16, 0.18; late plays, 0.27, 0.29.

I never heard of such another encounter, <u>which lames report to follow it, and undoes description to do it</u>

(56-8/TLN 3065-7)

Like an old tale still, <u>which will have matter to rehearse, though credit be asleep, and not an ear open</u>

(61-3/TLN 3070-2)

this avouches the shepherd's son; who has not only his innocence, <u>which seems much</u>
(63-5/TLN 3072-3)

One of the prettiest touches of all, and that <u>which angled for mine eyes, caught the water, though not the fish,</u> was when[...]

(81-3/TLN 3090-91)

Compare these subjectively inflected descriptions with those associated with 'that' relatives, which tend to introduce material more likely to be objectively true of the antecedent:

this avouches the shepherd's son; who has not only his innocence, which seems much, to justify him, but a handkerchief and rings of his <u>that Paulina knows</u>
(63-6/TLN 3072-5)

What, pray you, became of Antigonus, <u>that carried hence the child?</u>
(59-60/TLN 3068-9)

Neither relative here gives us access to any speaker's subjectivity. Rather, each supplies us with an objective fact about the head noun (Paulina recognises the handkerchief and rings; Antigonus took the child away).

Remember: although we have concentrated on one scene, Docuscope has shown that this type of effect increases significantly right across the late plays. Asides play an important role, for example, in Prospero's recounting of his prior life as Duke of Milan to Miranda in <u>The Tempest</u> (2.1); in Giacomo's confession of his feigned seduction of Innogen at the conclusion of *Cymbelene* (5.6); and in Henry's declaration that it was his conscience that prompted him to abandon his marriage to Katherine in *Henry VIII* (2.4) The very choice to have the identification of Perdita narrated at the end of *The Winter's Tale* rather than enacted on stage seems characteristic of Shakespeare's dramatic style at this point: by this stage in his career, Shakespeare is exploring, and exploiting, linguistic resources that allow him to track events through an individual's consciousness rather than to depict them impersonally -- a

prefiguration of Henry James, perhaps. It is hard to imagine such a crucial scene being narrated in one of the early plays.

We suspect that what Docuscope calls an Aside was, for Shakespeare and his actors, a concrete opportunity for communicating a subjective impression of things in the mind of the speaker (rather than rendering their supposed material reality more 'immediately' in the physical world of the play). The dramaturgical effect of this linguistic multiplication of 'internal' perspectives on the action would have been atmospheric, contributing—as it still does—a certain prismatic quality to the late plays that distances spectators from events rendered in the play in a way that a more naturalistic theatrical practice does not. Folded into the words and minds of the characters, events in the late plays can only, we might say, be accessed from a particular linguistic angle. We have the option to go further and make connections with other impressions that have been taken from these plays -- the sense that they are dreamlike, convoluted, difficult, spatially impossible. These interpretations would be supported, but not dictated, by the results of linguistic analysis, whose translation into dramaturgical strategy and (thus) theatrical effects is never one way or automatic.

We can broaden this analysis of *Asides* in CF5 by looking to the second LAT with which it is correlated, *Verb State*. This LAT consists mainly of present-tense usages of the verb *to be* (*is, are, be*) often taking the form: 'X <u>is</u> Y':

> A fair one <u>are</u> you
>
> > (*The Winter's Tale* 4.4.82/TLN 1885)

> > the fairest flowers of the season
> <u>Are</u> our carnations and streak'd gillyvors,
> Which some call nature's bastards
>
> > (*The Winter's Tale* 4.4.86-8/TLN 1889-91)

> > > over that art
> Which you say adds to nature <u>is an</u> art
> That nature makes
>
> > (*The Winter's Tale* 4.4.96-8/TLN 1901-3)

This LAT is also triggered by '*to be* + Determiner' combinations (as with 'is an' in the third example above), and *have* combining with forms such as *being* and *bearing on*. The *-- ed* forms of verbs also play a major role here, in combinations with *of* (for example 'composed of'), and when they occur clause-finally:

As you are certainly a gentleman, thereto
Clerk-like <u>experienced</u>

<div align="right">(The Winter's Tale 1.02.391-2/TLN 499-500)</div>

the need I have of thee thine own goodness hath <u>made</u>
<div align="right">(The Winter's Tale 4.02.11-2/TLN 1624-5)</div>

Non-finite forms such as 'being' and 'given', when used in a clause initially ('Being the mother of two sons...'; 'Given the state of the roads...') are also coded for this LAT.

The common thread here is that all of these different *Verb State* strings are associated with the communication of information that is taken to be (or at least presented as being) universally true, or states of affairs which continue from the past into the present. *Verb State* strings present continuous states rather than discrete actions ('He is kind' vs 'He acted kindly'). Interestingly, modern rhetorical analysts associate its use with non-fiction texts: texts that have a high information density, and texts that communicate information by telling rather than showing. The tell/show distinction is useful in this dramaturgical context because it helps us see that *Verb State* locutions are inevitably distanced from the thick of unfolding action: a judgment of some sort is being made in many of the examples above, the stative verb serving to provide a snapshot of a person or thing in time so that the judgment or predication can take place. Even if the thing discussed -- streak'd gyllivors, for example -- is a subject of heated debate, the 'stilling' or 'fermata' effect of the stative verb seems more attuned to the ether of consciousness than the swells of passion. A speaker who says 'I love you' is much more convincing than one who states, 'you are my love.'

The dramaturgical effects of this last feature is related, we believe, to the subjective effects we have already identified in uses of its factor correlate above, *Asides*. (Note the frequent co-location of both in the examples above). As we have said, *Verb State* is involved with the communication of states rather than actions, and we would argue that just as *Asides* involves a foregrounding of the subjectivity of the narrator, so *Verb State* tends to foreground the narrator in as much as it is more likely to involve the assertion of an opinion rather than the description of an action. In terms of overall effect, this factor again suggests a shift in dramatugical technique from the enactment of actual events in the early genres (showing) to the depiction of the mind's reaction to events in the late plays (telling), what might be called a technique of subjective particularization. Note how in the following passage a simple

statement of a past event is elaborated with increasing subjectivity involving both *Asides* and *Verb State*:

> the penitent king, my master, hath sent for me; to whose feeling sorrows I might be some allay, or I o'erween to think so, which is another spur to my departure
>
> (*The Winter's Tale* 4.02.6-9/TLN 1619-22)

The simple past event ('the penitent king...hath sent for me') becomes a vehicle for reflection on the role of the speaker himself ('to whose feeling sorrows I might be some allay'), an explicit acknowledgement of self-absorption ('or I o'erween to think so') and an assertion using *Verb State* and *Asides,* which turns a subjective opinion into something resembling a generally accepted truism ('which is another spur to my departure'). The end result: *Asides* and *Verb State* work hand in hand to render—not the truth of action declared on the stage—but the truth of *reaction* as it unfolds in the mind.

<p style="text-align:center">*</p>

<p style="text-align:center">*   *</p>

Focalized retrospection and subjective particularization are only shorthand ways of describing the larger pattern of linguistic features that make the late plays distinct. These descriptions were in some sense prompted by statistical analyses, but were certainly not dictated by them. Nor can we say that the factors above in any way exhaust the linguistic patterns that can be unearthed through close analysis of the plays.[23] Docuscope's identification of *Asides*, for example, as one of the key features of the style of late plays focused our attention on relativisation (and thus subordination) as a means of stamping the subjective impressions of the speaker on what is said. A fuller analysis of Shakespeare's language in the late plays than we have space for here would show that the effects we have associated with *Asides* are produced by a much wider range of features than simply relatives introduced by 'which'. Parallel adverbial clauses and other forms of clausal subordination are also used to specify or elaborate nouns with reference to prior events or actions, rather than the attributive, adjectival

---

[23]Single variable ANOVA, for example, showed the late plays to be high in a LAT called *Comparison,* which includes words like 'greater' and 'lesser,' 'more,' 'equal,' and 'best.' This LAT was not attached to any factor—it was not distributed evenly throughout the plays bur rather was clustered in limited areas. It may be related to the tendency to stamp narratorial judgment on the description of events, but then again, it may not. Linguistic features or patterns do not *all* have to serve or respond to a single dramaturgical demand.

modification we find in the earlier plays. These subordinating strategies cannot be 'counted' by Docuscope, but must be turned up through careful parsing and comparisons of possible alternative 'routes' of expression. Where there is smoke, however, there is often fire, and so the tokens of subordination counted by Docuscope may be more prominent precisely because *all* of the strategies being used in the late plays to produce internal perspective are related. Our interpretive leap is to assert this relatedness. The 'which' becomes worth counting, then, because it is a linguistic footprint indexing complementary linguistic and dramaturgical strategies; these, in turn, register or express a vision of the world and its order that gets ascribed to tragicomedy by critics and is perhaps experienced by audiences in the theater.

The factor analysis also suggests the ways in which the late plays are partially like and unlike the other three genres, in effect mapping out the features of family resemblance across a very large array of possible members. Although we can only gesture toward these results here -- no factor can really be understood without an investigation and discussion of examples -- it is clear that, given what Docuscope can count, the late plays exhibit linguistic features that are present in certain other genres but absent in others (or conversely, they lack certain features that some cognate genre *also* lacks in relation to others). For example, the late plays share features that distinguish both tragedies and comedies, but not histories. At the level of 1000 word chunks, factor analysis showed that the late plays were like both comedies and tragedies (and unlike histories) in their comparatively high frequencies of the LATs named *Denial/Disclaim* and *Resistance* (CF2/1000) and *Person Pronouns* (CF10/1000).[24] The late plays also lack something that both comedies and tragedies lack in relation to histories -- a LAT called *Time Duration*, which is made up of words and phrases that indicate time spans, such as 'in his eleventh year,' 'summer,' 'hour,' 'when,' 'while,' 'day,' and 'night.' While the resemblance with comedies and tragedies is attested in one way or another by at least six factors at this level, late plays do share some similarity with the history plays in that both possess a comparatively high frequency of a LAT called *Generalization*, which includes words and phrases such as 'every,' 'all,' 'of all,' 'for all,' 'with all,' and 'every thing.' An analysis of criss-crossing similarities and differences on this scale is far beyond the means of this paper, but even a cursory glance at the results points out the degree to which Docuscope's understanding of 'pattern' and

---

[24] *Denial/Disclaim* LATs include words and phrases such as 'cannot,' 'not,' 'is not,' 'nothing,' 'no,' and 'never.' Its partner in CF2/1000, *Resistance*, includes words and phrases such as 'yet,' 'but,' 'against him,' 'against her,' 'even so,' 'rather,' and 'would not.' The LAT *Person Pronoun*, which constitutes CF10/1000, includes the following: 'he,' 'his,' 'him,' 'himself,' 'she,' 'her,' 'hers,' 'herself,' 'whose.'

'significance' implies multiple forms of relation, often predicated on the simultaneous presence and absence of key features in groups of texts in some apparent pattern of coordination.

We must remain content here to conclude that, on a truly pervasive linguistic level, the late plays seem to do what many auditors and readers have experienced them doing on stage and page: they make way for inner life and revelation through memory and recognition; they pivot theatrical and readerly attention on the movements of a mind engaged in thought; they accommodate complexity of plot and long stretches of wandering by allowing the contingencies of romance wandering to be glossed with ruminating digressions; and they subordinate the declaration of actions present and past to the stillness of judgment. These dramaturgical possibilities are not dictated by the 'types' of language being used, just as the 'types' of language that are spoken on stage are not an expression of some universal set of theatrical scenarios or archetypal "scenes." Rather, theatrical utterance and the action we associate with a particular genre must be part of a single structure —drama—which in the end is a story told by speaking beings on a stage in a limited amount of time.

If approaching Shakespeare 'by the numbers' opens up another window onto the playwright's late plays and genre, it should also tell us something new about the ways we might choose to read Shakespeare and other early modern writers in the coming years. Initiatives such as the Text Creation Partnership of Early English Books Online, for example, are creating a corpus of 25,000 digitized early modern texts that could be adapted to far reaching forms of quantitative analysis.[25] As this corpus of early English Books becomes available for statistical analysis, critics will have an unprecedented set of resources to employ in thinking about traditional 'literary' questions such as genre, style, influence, and perhaps authorship. Having worked through a limited number of critical questions with one particularly rich corpus (Shakespeare's plays) and the institution of its criticism, we believe our work with Docuscope may prove instructive to future scholars who want to understand the usefulness of 'counting things' in humanistic inquiry -- quantity being perhaps one of the last concepts in the humanities which has not come in for rigorous theorization.[26] How is it, for example, that

---

[25] According to Early English Books Online, the Text Creation Partnership (TCP) is in the process of creating SGML coding for the full text of 25,000 EEBO works, allowing users to search the full ASCII text of the documents. Because Docuscope can analyze vast numbers of texts once they have been rendered into ASCII form, we find the potential for future quantitative work with such a corpus tremendously exciting.

[26] For a recent attempt to think the value of quantity in literary studies, see Franco Moretti's polemically charged *Graphs, Maps, Trees: Abstract Models for a Literary Theory* (New York: Verso, 2005).

something like a 'factor' from multivariate statistical analysis might capture something as intimate as genre -- 'lateness' in Shakespeare's style being, admittedly, a difficult concept to defend on even the most subjective grounds?[27] How can classifying vast numbers or 'strings' of words into their 'functions' and then looking for patterns in their underlying use/absence provide a sense of the singular experiential richness of a literary text? The answer, we believe, depends upon what one decides to count and how one decides to understand the significance of what has been counted. Both of *these* decisions are interpretive and contextual, even if the things counted are either objectively 'there' or 'not there' to be counted in an editorially stabilized text.

Genre, it seems to us, is just as much a coordinated pattern of various types of dramaturgical and linguistic effects as it is some kind of 'defining set of properties': if it weren't, genre would be completely inaccessible to multivariate analysis. Indeed, the difficulty of saying exactly what makes a genre a genre, a difficulty that leads Mowat to invoke Wittgenstein's family resemblance analogy, may be precisely what makes genre amenable to quantitative statistical analysis: only multivariate analysis can produce a 'thing' as abstract as a 'factor' that coordinates relative frequencies across large numbers of groups. Only a factor, that is, can call the critic's attention, in some organized way, to the pervasive things that happen and don't happen at the same time in a *set* of literary texts. The fact that readers may actually register such quantitative factors as qualities within a particular work suggests that the proverbial quantitative/qualitative divide—or Dilthey's contrast between the *geisteswissenschaften* and the *naturwissenschaften*—hides certain areas of overlap that ought to be explored more carefully. As Henri Bergson would say, every change in quantity is ultimately a change in quality, although he probably would never have dreamed of counting non-restrictive clauses in the late Shakespeare.[28]

Just as important, the use of language in the theater, because it is constrained by material factors such as the number of actors, the size of the stage, the various 'technologies' for simulating experience (music, noise, machines, etc), is deeply marked by these constraints; when one is 'counting' the various types of language that get used in the theater, then, one

---

[27] See the essay by Gordon McMullan in this collection.

[28] Henri Bergson, *Time and Free Will: An Essay on the Immediate Data of Consciousness*, trans. F. L. Pogson (New York: Dover, 2001), ch. 1. Bergson's notion of 'duration', it should be noted, is specifically designed to foreclose attempts at 'converting' quantitative differences into qualitative ones. We have, in our own analyses, tried to avoid such facile conversions.

must not underestimate the degree to which the texts of plays are saturated with dramaturgical exigencies—the need to do something with language in a particular way in a particular set of circumstances. If you want to tell a story about wandering siblings and long lost children and do it without the cinema technology of flashback, expect to do a lot of verbalized retrospection. So too, you may have to rely more on the emotional charge of subjective immersions, where a speaker narrates his or her way around a memory to encounter joy or despair, than on the narrated misfires of action in comic mishap (with its props and circumstantial judgments), tragic plotting and counterplotting (with its scenographically inclined onstage contrivers), or the concrete detail of historical conflict and battle. To understand what a particular kind of play *does* with its verbal resources, which is really what we mean by the word dramaturgy here, it makes sense to ask a blindfolded critic like Docuscope to feel its way around the corpus. You could, similarly, learn a great deal about how your house or apartment works by asking someone else to live in it for a month and then tell you how it works. Certain things are connected that you may have never noticed -- the tap water pressure, say, seems to fluctuate whenever the neighbours turn on their sprinkler. So too, rather prosaically, we have discovered that Shakespeare's penchant for certain types of subordination and verb forms in the late plays seems to compliment a relative abundance of focalized retrospection: these things seem appropriate given the kind of story he and his company are trying to tell. One can try to grasp such interrelations through trial and error, but the odds of doing so are probably only slightly better than that of a monkey writing *The Tempest*. As literary critics attempting to read Shakespeare by the numbers, we are much more lucky than the proverbial monkey in the statistical parable. The deck is loaded in our favor -- there are patterns, no end of patterns, for us to pore over while reading Shakespeare by the numbers. It is only because he was trying to *do* something in the theater, however, that even one of those patterns makes sense to us today.