Authorial Attribution and Computational Stylistics: If You Can Tell Authors Apart, Have You Learned Anything About Them?

Hugh Craig
Department of English, The University of Newcastle, Australia

Abstract

Within stylometrics, the disciplines of authorial attribution and descriptive stylistics hitherto have been pursued separately. The first has achieved mainstream status within literary studies, while the second is little recognized. A study of the plays of Thomas Middleton compared with a large control sample shows that it is possible to achieve a good classification of his work in 2,000 word segments as against those of his contemporaries, using frequencies of very common words. Such a result can serve as the basis both for testing Middleton's hand in some disputed plays and for a description of his style. Most of *The Revenger's Tragedy*, part of *The Yorkshire Tragedy*, and all of *The Second Maiden's Tragedy* prove to be in a style similar to that of Middleton's uncontested plays. This style, judging by an examination of instances in context of the ten word types most strongly correlated with the Middleton–other discriminant function, is (among other things) rich in deictics and poor in conjunctions, features readily accommodated to previous descriptions including Middleton's own. It is concluded that classification and description can be mutually supportive: the first confirms the validity of the second, while the second helps to establish the stylistic mechanism underlying a successful classification.

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

There is still a large methodological blank between statistics and style. Precious few theoretical models bridge the gap between, on one hand, the counting and analysing of linguistic features, and, on the other, the eclectic, holistic, impressionistic, and yet indispensable business of capturing the impact and flavour of a group of texts. Partly this is a question of some real difficulties. The leap from frequencies to meanings must always be a risky one. The interpreter who is tempted to speculate about the world-view or psychology of a writer, based on quantitative findings, presents an easy target for dismissive critique (Fish, 1973). On the other hand, staying within the safe confines of the statistical results themselves means that one is soon left with only banalities or tautologies. Lower-

Correspondence:
Hugh Craig, Department of English, The University of Newcastle, Callaghan, NSW 2308, Australia.
E-mail: eldhc@cc.newcastle.edu.au

© Oxford University Press 1999

level features are easy to count but almost impossible to interpret in terms of style; counting images or other high-level structures brings problems of excessive intervention at the categorization stage, and thus unreliability and circularity (Van Peer, 1989). If the data come from a mid-level linguistic feature such as a word, then the interpreter must still establish what, and at what level, a statistically significant high or low frequency might mean. How does one take account of the various different senses of a word? Is the word a true variable or really only a proxy, dependent on some phrasal or even larger structure? What relationship do high and low frequencies have to the reading experience anyway?

For the classification of writing samples, the situation is a little different. Here there are some obvious links between the numbers and the coherence and distinctiveness of a group of texts. In the case of authorial attribution, for instance, mannerisms or quirks of expression can explain abnormal frequencies of words or phrases. Where there are alternative equivalent forms, such as *while* and *whilst*, it is easy to accept that individuals might develop idiosyncratic but consistent preferences and habits, leading to the over-representation of one or other form in their writing. Historical linguistics can account for different rates of usage of incoming and recessive forms between authors by reference to their date of birth, their city or rural origins, or their socio-economic class (Hope, 1994).

It is a fact of life, moreover, that it is in attribution studies, and only there, that computational stylistics has achieved mainstream status. Arguments based on quantitative evidence that a text is or is not by a certain author find an audience in the most read scholarly journals, and (when it is a matter of a new Shakespeare ascription) even in the mass media (Holmes, 1998). Descriptive uses of quantitative measures in literary study, on the other hand, have made little impact beyond a small circle of practitioners.

Yet there is an odd asymmetry in the notion that frequencies of linguistic features can classify a style and yet cannot play a part in describing it. Common sense suggests that if quantitative measures are reliable in telling authors apart, and if they offer access to new internal evidence genuinely independent of impressionistic criticism, then some among them ought also to be of use in the main business of literary study, the interpretation of texts. The descriptive aspect may well be an important ally of classification, in any case. After all, how much confidence can one have in an ascription, if the linguistic mechanism behind the results remains a mystery?

At this point, I propose to explore a particular case, in the hopes of illustrating some of the unexplored possibilities in this nexus between attribution and description. There are twelve plays which are generally accepted as the work of the English Renaissance dramatist Thomas Middleton (Lake, 1975). Many others have been associated with Middleton since their appearance. Among these is a group of three tragedies, *The Revenger's Tragedy*, *The Second Maiden's Tragedy*, and *The Yorkshire Tragedy*, which, if added to Middleton's canon, would make him into one

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

of the most considerable exponents of this genre. Arguments for and against these attributions have been pursued over many years, drawing on external as well as internal evidence (Barker, 1945; Schoenbaum, 1951; Foakes, 1953; Lake, 1975; Jackson, 1979; Rasmussen, 1989; Smith, 1991; Holdsworth, 1994). Middleton's style, meanwhile, has proved a little elusive in definition. T. S. Eliot thought he had 'no point of view' (Mulryne, 1979, p. 55); Bradbrook refers to the 'pregnant simplicity' on which Middleton relies (1964, p. 239); White refers to the 'plainness' of his style (1992, p. 141). Ricks is unusual in making a close study of Middleton's verbal structures, in particular the playwright's use of significant repeated 'commonplace words' (1961, p. 249). He explores the way in which these create potent dramatic effects which combine 'fundamental simplicity with local subtlety' (1960, p. 303).

With classificatory and descriptive purposes in mind, a quantitative analysis was designed using a set of ninety-seven plays of the period, including all twelve of Middleton's undisputed authorship. They are listed in Table 1. The texts used for counting were old-spelling ones, early printed or manuscript versions. These were preferred to modern edited texts, which approach contractions and variant spellings in different ways, and are sometimes conflations of various early versions. With old-spelling versions as copytexts, orthography can be standardized consistently across the corpus. To refine the data a little, and to make the results easier to interpret, certain function words were tagged to separate grammatical functions (*that* as demonstrative, relative, and conjunction, for instance).

The data used are counts of 155 very common words in ninety-seven plays broken up into 2,000 word segments.[1] The list represents the commonest words in the corpus, with some exclusions at the lower (less frequent) end, to avoid word variables associated too strongly with local subject matter (e.g. *king*) or hard to disambiguate (e.g. *ere*). For sheer efficiency in classification by the discriminant analysis, the larger the word list the better—the odds of finding markers of one group or another are increased—but it is the commonest words that are most suitable for descriptive stylistic analysis, since they could be expected to represent structural rather than idiosyncratic features. The present list is a compromise designed to accommodate both uses.

Figure 1 shows the results of a discriminant analysis of a group of segments taken from plays by Thomas Middleton, compared with segments from plays by others. Discriminant analysis is a well-established multivariate technique which creates a function designed to separate predefined groups of observations, a function which can then be used to classify cases whose group membership is unknown (Van de Geer, 1971, pp. 243–72). In the figure, the top two bands show the results for 110 Middleton segments and 775 non-Middleton ones. The analysis is mostly, but not entirely, successful in separating the two groups: all but one of the Middleton segments are classified correctly—the vertical dotted line shows the boundary between the two groups that the analysis creates—while just eight of the non-Middleton ones are assigned

1 The word variables are: a, about, all, am, an, and, any, are, as, at, be, been, before, best, better, both, but, by(preposition), can, cannot, canst, come, comes, could, did, do, does, done, doth, first, for(conjunction), for(preposition), from, give, go, had, has, hast, hath, have, he, her(adjective), her(personal pronoun), here, him, his, hope, how, I, if, in(adverb), in(preposition), into, is, it, know, let, like(preposition), list, little, long, may, me, men, might, mightst, mine, much, must, my, nay, need, needst, no(adjective), no(exclamation), noble, nor, not, nothing, now, O, of, on(adverb), on(preposition), once, one, or, ought, our(royal plural), our(true plural), out, see, shall, shalt, she, should, shouldst, since, sir, so(adverb of degree), so(adverb of manner and conjunction), stay, still, such, take, tell, than, that(conjunction), that(demonstrative), that(relative), the, thee, their, theirs, them, then, there, these, they, thine, think, this, those, thou, though, thus, thy, till, time, to(infinitive), to(preposition), too, up(adverb), upon(preposition), us(royal plural), us(true plural), very, was, we(royal plural), we(true plural), well, were, what, when, which(relative), who(interrogative), who(relative), whose, why, will(verb), wilt, with, world, yes, and yet.

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

Table 1 Plays

Author	Title	Copytext	Publication date
Anonymous	*Arden of Feversham*	STC733	1592
Anonymous	*A Warning for Fair Women*	STC25089	1599
Anonymous	*The Revenger's Tragedy*	STC24150	1608
Anonymous	*A Yorkshire Tragedy*	STC22340	1608
Anonymous	*The Second Maiden's Tragedy*	Malone Soc Repr.	1909
Beaumont, Francis	*The Knight of the Burning Pestle*	STC1674	1613
Beaumont and Fletcher	*The Maid's Tragedy*	STC1676	1619
Chapman, George	*The Revenge of Bussy d'Ambois*	STC4989	1613
Davenant, William	*The Unfortunate Lovers*	Wing D348	1643
Dekker, Thomas	*A Shoemaker's Holiday*	STC6523	1600
Fletcher, John	*Valentinian*	Wing B1581	1647
Fletcher, John	*The Bloody Brother*	STC11064	1639
Fletcher and Massinger	*The Double Marriage*	Wing B1581	1647
Ford, John	*The Broken Heart*	STC11156	1633
Ford, John	*Love's Sacrifice*	STC11164	1633
Ford, John	*'Tis Pity She's a Whore*	STC11165	1633
Fulke, Greville	*Mustapha*	STC12362	1609
Goffe, Thomas	*The Courageous Turk*	STC11977	1632
Greene, Robert	*Friar Bacon and Friar Bungay*	STC12267	1594
Greene, Robert	*I Selimus*	STC12310a	1594
Heywood, Thomas	*A Woman Killed with Kindness*	STC13371	1607
Jonson, Ben	*The Case is Altered*	STC14757	1609
Jonson, Ben	*Every Man in his Humour*	STC14766	1601
Jonson, Ben	*Every Man out of his Humour*	STC14767	1600
Jonson, Ben	*Cynthia's Revels*	STC14773	1601
Jonson, Ben	*Poetaster*	STC14781	1602
Jonson, Ben	*Sejanus his Fall*	STC14782	1605
Jonson, Ben	*Volpone*	STC14783	1607
Jonson, Ben	*Epicoene*	STC14751	1616
Jonson, Ben	*The Alchemist*	STC14755	1612
Jonson, Ben	*Catiline his Conspiracy*	STC14759	1611
Jonson, Ben	*Batholomew Fair*	STC14753 5	1631
Jonson, Ben	*The Devil is an Ass*	STC14754	1640
Jonson, Ben	*The Staple of News*	STC14753.5	1631
Jonson, Ben	*The New Inn*	STC14780	1631
Jonson, Ben	*The Magnetic Lady*	STC14754	1640
Jonson, Ben	*A Tale of a Tub*	STC14754	1640
Jonson, Ben	*The Sad Shepherd*	STC14754	1640
Kyd, Thomas	*The Spanish Tragedy*	STC15086	1592
Markham and Sampson	*Herod and Antipater*	STC17401	1622
Marlowe, Christopher	*Dido, Queen of Carthage*	STC17441	1594
Marlowe, Christopher	*I Tamburlaine the Great*	STC17425	1590
Marlowe, Christopher	*The Jew of Malta*	STC17412	1633
Marlowe, Christopher	*Doctor Faustus*	STC17432	1616
Marmion, Shakerley	*The Antiquary*	Wing M703	1641
Middleton, Thomas	*The Phoenix*	STC17892	1607
Middleton, Thomas	*Your Five Gallants*	STC17907	1608
Middleton, Thomas	*A Trick to Catch the Old One*	STC17896	1608
Middleton, Thomas	*A Mad World, my Masters*	STC17888	1608
Middleton, Thomas	*Michaelmas Term*	STC17890	1607

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

Table 1 Plays (*Continued*)

Author	Title	Copytext	Publication date
Middleton, Thomas	*A Chaste Maid in Cheapside*	STC17877	1630
Middleton, Thomas	*No Wit, no Help Like a Woman's*	Wing M1985	1657
Middleton, Thomas	*More Dissemblers Besides Women*	Wing M1989	1657
Middleton, Thomas	*The Witch*	Malone Soc. Repr.	1948
Middleton, Thomas	*Hengist, King of Kent*	ed. R. C Bald	1938
Middleton, Thomas	*Women Beware Women*	Wing M1989	1657
Middleton, Thomas	*A Game at Chess*	Trinity, Camb. ms	1624
Middleton and Rowley	*The Changeling*	Wing M1980	1653
Marston, John	*Antonio's Revenge*	STC17474	1602
Marston, John	*Sophonisba*	STC17488	1606
Massinger, Philip	*The Unnatural Combat*	STC17643	1639
Massinger, Philip	*The Roman Actor*	STC17642	1629
Peele, George	*King David and Fair Bethsabe*	STC19540	1599
Porter, Henry	*Two Angry Women of Abingdon*	STC20122	1599
Rowley, William	*All's Lost by Lust*	STC21425	1633
Rowley, William	*A New Wonder, a Woman Never Vexed*	STC21423	1632
Shakespeare, William	*The Comedy of Errors*	STC22273	1623
Shakespeare, William	*Richard III*	STC22314	1597
Shakespeare, William	*The Taming of the Shrew*	STC22273	1623
Shakespeare, William	*Titus Andronicus*	STC22328	1594
Shakespeare, William	*Love's Labours Lost*	STC22294	1598
Shakespeare, William	*Romeo and Juliet*	STC22323	1599
Shakespeare, William	*The Merchant of Venice*	STC22296	1600
Shakespeare, William	*I Henry IV*	STC22280	1598
Shakespeare, William	*II Henry IV*	STC22288	1600
Shakespeare, William	*Henry V*	STC22273	1623
Shakespeare, William	*Julius Caesar*	STC22273	1623
Shakespeare, William	*Twelfth Night*	STC22273	1623
Shakespeare, William	*Hamlet*	STC22276	1604
Shakespeare, William	*Troilus and Cressida*	STC22331	1609
Shakespeare, William	*Measure for Measure*	STC22273	1623
Shakespeare, William	*Othello*	STC22305	1622
Shakespeare, William	*King Lear*	STC22292	1608
Shakespeare, William	*Macbeth*	STC22273	1623
Shakespeare, William	*Antony and Cleopatra*	STC22273	1623
Shakespeare, William	*Timon of Athens*	STC22273	1623
Shakespeare, William	*Coriolanus*	STC22273	1623
Shakespeare, William	*The Winter's Tale*	STC22273	1623
Shakespeare, William	*The Tempest*	STC22273	1623
Shirley, James	*Love's Cruelty*	STC22449	1640
Shirley, James	*The Traitor*	STC22458	1635
Shirley, James	*The Cardinal*	Wing S3461	1652
Suckling, John	*Aglaura*	STC23420	1638
Tourneur, Cyril	*The Atheist's Tragedy*	STC24146	1611
Webster, John	*The White Devil*	STC25178	1612
Webster, John	*The Duchess of Malfi*	STC25176	1623
Wilmot, Robert *et al.*	*Tancred and Gismund*	STC25764	1591

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

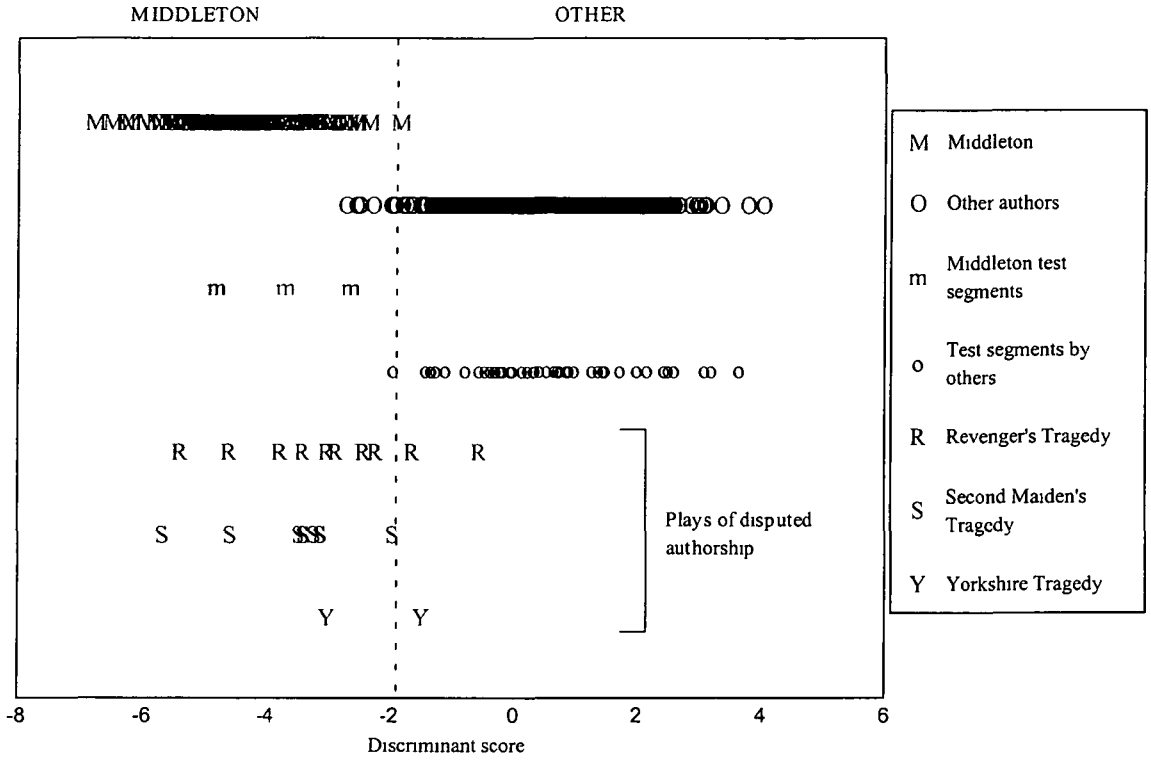

Fig. 1 Discriminant analysis of play segments: Middleton versus the rest.

wrongly to Middleton. There is always a suspicion, however, that while a tailor-made function may perform well in separating these particular segments, it is attuned so specifically to the training set that it will not provide reliable results with new segments, such as ones by unknown authors. It is common, therefore, to reserve a test set of observations from the two groups and to calculate scores for them for the same function. The middle part of the chart shows the results for test groups consisting of four Middleton segments (two of them have such similar scores that they appear on the chart as a single symbol) and forty-four segments by others, all chosen at random from the original set. All four of the Middleton test segments are classified correctly; so are forty-three of the forty-four segments by others. It would seem that the differentiation between Middleton and the others may indeed be generalized beyond the original set. *The Changeling*, a collaboration between Middleton and William Rowley, was also reserved for testing, but not included in the chart; eight of the nine segments were classified with the Middleton group, another result which gives confidence in the reliability of the discriminant function.

With this serviceable tool for authorial attribution, we can try out some of the anonymous texts which have been attributed to Middleton. How well do these other texts fit the Middleton style profile in the present analysis? The lowest part of the chart shows discriminant scores for the three disputed plays. Eight of ten of the *Revenger's Tragedy* segments are

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

classified as Middleton, as are all nine segments of *The Second Maiden's Tragedy*, and one of two segments of *The Yorkshire Tragedy*. On this basis, Middleton would seem to have had a large but not exclusive part in *The Revenger's Tragedy*, to have written the whole of *The Second Maiden's Tragedy*, and to have had a share at least in *The Yorkshire Tragedy*.

The discriminant analysis has defined a weighted combination of the original word variables which effectively separates Middleton segments from the others. The function works by giving the heaviest weights to the variables with the largest and most consistent differences between Middleton counts and the others. These word variables can claim to be the foundation of the Middleton style as defined by the analysis. Table 2 shows the first ten of these word variables, with means and standard deviations for Middleton segments and the others.[2] The means are also converted into an expected number of instances for a 2,000 word segment.

At this point, we are back at the moment of transition from frequencies to description with which we began. For the student seeking an insight into Middleton's style, what is there here? E. H. C. Oliphant, without the benefit of mechanized counting, saw that the demonstrative *that* (which appears at the top of the list) was exceptionally frequent in Middleton, as a consequence of what Oliphant called one of Middleton's 'tricks of construction' (Oliphant, 1970, p. 84). Yet there is more to be said. The demonstrative *that* is a strongly deictic word, directing the attention and point of view of an interlocutor. In Middleton's dialogue. it often punctuates a more or less marked turning of the tables: 'And when a thing's done, 'tis done,/You taught me that Mother' (*A Chaste Maid*, II.ii)[3]. Judging from a sampling of instances, anaphoric uses are especially common. Follywit in *A Mad World, My Masters* remarks, 'And thats worth all sir'; Sir Bounteous Progress replies, 'And thats worth all indeed my L[ord]' (II.1). In both pronoun and adjective form, it is frequently indignant or disdainful in Middleton. In *A Chaste Maid*, we hear of 'that foule whore his Wife' (V.i); in *A Mad World* 'Gag that gaping raskall'

Table 2 Ten Middleton markers

	Percentage of total dialogue				Expected instances per segment	
	Mean		Standard deviation			
	Middleton	Other	Middleton	Other	Middleton	Other
that (demonstrative)	0.74	0.41	0.24	0.18	15	8
has	0.25	0.10	0 13	0 12	5	2
hath	0.02	0.17	0 04	0 13	0	3
now	0.62	0 40	0 21	0 18	12	8
a	2.38	1 81	0 44	0 61	48	36
and	2.22	2 94	0.42	0.72	44	59
sir	0 98	0 47	0 58	0.51	20	9
there	0.41	0.27	0 16	0.15	8	5
that (conjunction)	0.18	0.34	0.11	0 18	4	7
doth	0.01	0.08	0.02	0.09	0	2

Listed in order of correlation with the Middleton–other discriminant function. All have *t*-test *P*-values below 10^{-14} for Middleton segments versus the rest

2 All Middleton segments have been included, and only the anonymous plays and the Middleton collaboration, *The Changeling*, excluded from the other group.

3 Quotations are from the early versions listed in Table 1. Act and scene references are given in the text.

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

(II.i). In either form, it tends to be emphatic. It is associated with familiarity, ellipsis, and a sort of verbal nudging.

Has and *hath*, next on the list, are neatly complementary: Middleton has as many of the two forms together as his contemporaries, and is only unusual in preferring *has* to *hath*. There is a similar pattern, though not quite so neat, with *doth* and *does*: *doth* (at the bottom of the table) is markedly scarce in Middleton, but *does* is not significantly different in frequency between his and the others' plays (and does not, therefore, appear in the list of the ten most important Middleton markers). With these two pairs, the analysis offers evidence more useful for identification than for style, though Middleton's neglect of the more traditional, perhaps fustian-sounding variants is of interest, and has been noted previously (Stein, 1987). *Sir* also belongs most properly with the markers which are something like neutral in terms of style, a writer's tic which adds only to an impression of characters with a particularly insistent address to each other.

Now (fourth on the list) is more promising as a stylistic marker. It indicates that Middleton's characters are unusually attuned to the present moment. In comic dialogue, this is often associated with a sense of the moment-to-moment fluidity of time: 'Ile try these fellowes now,—a word sir, what will you carry me to that widdow now?' (*A Trick to Catch the Old One*, IV.i). In the tragedies, it suggests an ominous fixity, implying to the audience, if not always to the character, that all can change irrevocably, and catastrophically, in a moment. Beatrice-Joanna, in a section of the main plot of *The Changeling* almost certainly by Middleton (Bawcutt, 1958, p. xxxix), says to her murderous lover Deflores, 'I'me forc'd to love thee now,/Cause thou provid'st so carefully for my honor' (V.i). Middleton can comment ironically by having characters in the same scene use the word in divergent senses. In Act III, scene i of *Women Beware Women*, Leantio, recently married to Bianca, is returning from a business trip, full of complacent anticipation of accumulated affection from his new wife. However, Bianca has been corrupted by the Duke in his absence. Leantio's *now* expresses impatient expectation: 'Now for a welcome/Able to draw mens envies upon man', he says. For Bianca, the same word is inflected with the instability of court fashion, the changeability of things and especially of affections. Rather than exchange sweet nothings with her husband, she asks pointedly, 'What news now of the Pirats, any stirring?' After all, as she reminds him, 'We have been married a whole fortnight now'.

Middleton's tendency towards high counts of the indefinite article (fifth on the list) can be understood as a disposition on the part of his characters to make reference general rather than particular. The usages suggest worldliness, an urbane capacity to see events in a wider context, as in this pair of lines from that same Bianca: 'Tis a sweet recreation for a Gentlewoman,/To stand in a Bay-window, and see gallants' (III.i). In *The Witch*, we hear a character wonder 'What strange Cuning/Sin helpes a woman to' (III.ii). In *The Chaste Maid of Cheapside*, the strong showing of the word reflects a tendency among the characters to treat entities as

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

commodities, as more or less interchangeable members of a class. This is necessarily so in 'a hundred pound' (I.i), but the usage in 'Am I deny'd a Chamber?' (V.i) and 'beat my Braines against a Bed-post'(V.ii) is more a matter of stylistic choice. The idiosyncratic locution 'this will make a poore S[i]r Walter' (V.iii) shows how even a proper name and a human individual can be transformed by the article into a class.

And is the commonest of all the markers in the table overall, and significantly infrequent in the Middleton texts. This is not a matter of Middleton's substituting another conjunction. *Or*, *but*, and *for* all tend to be around or below the average in Middleton dialogue. Another conjunction, *that*, is on the list of the ten best Middleton markers, being (again) unusually scarce in his dialogue. Middleton in general, it seems, uses fewer conjunctions than his contemporaries.[4] His dialogue is a series of forays with reduced coordination, in manner rather casual and impulsive.

As well as *now* and *that* in its demonstrative sense, *there* is in the list of Middleton markers, a third deictic which occurs exceptionally often in his writing. Together, these three markers are evidence of a grammaticalized style strongly dependent on context. Characters in Middleton point instead of tell. They imply, or claim, common ground with each other, exercising an anaphoric economy of communication.

In the preface to the printed version of *The Roaring Girle*, Middleton compared the kind of drama then in vogue with current modes of dress. Where there were once 'huge bombasted plays,' 'quilted with mighty words to lean purposes,' Middleton says, 'Now in the time of spruceness, our plays follow the niceness of our garments: single plots, quaint conceits, lecherous jests, dressed up in hanging sleeves' (White, 1992, p. 141; Middleton, 1611). Frequency data from common words can help define Middleton's particular kind of 'spruceness' and 'niceness'. A substantial passage of Middleton, so the present argument goes, will accumulate meanings of a stylistic kind in the notable abundances and scarcities summarized in the table. The reader or hearer will become aware of an open, casual, commodified reference; a dialogue of emphatic cross-reference and insistent deixis; and a thinning out of expected conjunctions, suggesting improvization and fragmentation.

The same profile of frequencies[5] serves as a test of whether a passage belongs inside or outside the Middleton canon. The verdict on the anonymous *The Second Maiden's Tragedy* is in favour, as I have already noted. The quantitative results make it possible, on systematic grounds, to see the author of *Women Beware Women* and *The Chaste Maid of Cheapside* in the following exchange from the end of the anonymous play. The dying Bellarius here encapsulates the plot for the benefit of Govianus, the rightful, but presently usurped, king. (Positive Middleton markers are printed in bold, negative ones in italics.)

Bellarius . . . **that** lord yor brother, made his frend Votarius
to tempt his ladie, she was wonne to lust,
the Act reveald here by her servingwoman,

4 A composite variable composed of frequencies of *and*, *as*, *but*, *for*(conjunction), *if*, *or*, *so*(conjunction), *that*(conjunction), *thus*, and *yet* has a Middleton mean of 4.86% and a mean for the others of 6.15%. The *t*-value for the difference of means is −13.5 at 931 degrees of freedom, significant at the <0.001 level. Middleton means are lower for all but *so* and *yet*. With old-spelling texts and Early Modern English, parsing and tagging must be done by hand, so a completely parsed version, which would provide more comprehensive evidence on this question, would be prohibitively laborious.

5 It may be argued that all 155 words are used in the classification process for the segments, and not only the ten in the table, still less the six distinguished here as the more useful stylistic markers. It is true that a large set of word variables is needed to achieve a good separation at the level of the 2,000 word segment. Yet the highlighted word variables remain powerful discriminators on their own. A discriminant analysis of whole plays, using only the six stylistic word variables, creates a function providing a perfect separation of the twelve Middleton plays from the eighty-one others in the set known to be by other authors (omitting *The Changeling* and the three tragedies with suspected Middleton involvement from the larger corpus).

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

but that wise close Adulteress storde wth arte
to praie vpon the weakenes of that lord
dissembled a great rage vpon her loue
and indeed kild him, wch so wonne her husband
he slew this right discouerer in his furye,
whoe being my mistris I was mov'de in harte
to take some paines with him, *and* has paid me for't
As for the cvninge ladie I comend her,
shee performed that which neuer woman tride
she ran upon twoe weopens *and* so died,
now you have all I hope I shall sleep quiet —
[Dies.]
Govianus. Is death so longe a comynge to mankinde
it must be met half waies? Lass the full tyme
Is (to eternitie) but a minute, A —
was that so long to staie? o cruell speed ;
ther's few men paie their debtes before their daie
if they be readie at their tyme, tis well,
and but a few that are so, what straunge hast
was made among thease people; my heart weeps for't
go, beare those bodies to a place more comely . . . (V.ii)[6]

Here, surely, is one version of that understated and ironic style—poor in conjunctions, rich in anaphora and deixis—which can be seen as, at one and the same time, Middleton's creative achievement and his cryptic signature.

Acknowledgements

I would like to acknowledge the generous hospitality of the Director and staff of the Istituto di Linguistica Computazionale, Pisa, Italy, where most of the work for this paper was done.

References

Barker, R. H. (1945). The Authorship of *The Second Maiden's Tragedy* and *The Revenger's Tragedy*. *Shakespeare Association Bulletin*, 20: 51–62.

Bawcutt, N. W. (ed.) (1958). *The Changeling*. By Middleton, T. and Rowley, W. The Revels Plays. London. Methuen.

Bradbrook, M. (1964). (First edn. 1935). *Themes and Conventions of Elizabethan Tragedy*. Cambridge: Cambridge University Press.

Fish, S. (1973). What is stylistics and why are they saying such terrible things about it? In Chatman, S. (ed.), *Approaches to Poetics. Selected Papers From The English Institute*. New York: Columbia University Press, pp. 109–52.

Foakes, R. A. (1953). On the authorship of *The Revenger's Tragedy*. *Modern Language Review*, 48: 129–38.

Holdsworth, R. V. (1994). Middleton's authorship of *A Yorkshire Tragedy*. *Review of English Studies*, 45: 1–25.

6 The editor's square brackets and alternative readings have been omitted.

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015

Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing*, 13: 111–17.

Hope, J. (1994). *The Authorship of Shakespeare's Plays: A Socio-linguistic Study*. Cambridge: Cambridge University Press.

Jackson, MacD. P. (1979). *Studies in Attribution, Middleton and Shakespeare*. Salzburg: Institut fur Anglistik und Amerikanistik, Universitat Salzburg.

Lake, D. J. (1975). *The Canon of Thomas Middleton's Plays: Internal Evidence for the Major Problems of Authorship*. London: Cambridge University Press.

Middleton, T. (1611). Preface. In Middleton, T and Dekker, T. *The Roaring Girle*. London.

Mulryne, J. R. (1979). *Thomas Middleton*. Harlow: Longmans.

Oliphant, E. H. C. (1970). (First edn 1927). *The Plays of Beaumont and Fletcher: An Attempt to Determine their Respective Shares and the Shares of Others*. New York: Phaeton Press.

Rasmussen, E. (1989). Shakespeare's hand in *The Second Maiden's Tragedy*. *Shakespeare Quarterly*, **40**: 1–26.

Ricks, C. (1960). The moral and poetic structure of *The Changeling*. *Essays in Criticism*, **10**: 290–306.

Ricks, C. (1961). Word-play in Middleton's *Women beware Women*. *Review of English Studies*, 12: 238–50.

Schoenbaum, S. (1951). *The Revenger's Tragedy* and Middleton's moral outlook. *Notes and Queries*, **196**: 8–10.

Smith, M. W. A. (1991). The authorship of *The Revenger's Tragedy*. *Notes and Queries*, **38**: 508–13.

Stein, D. (1987). At the cross-roads of philology, linguistics and semiotics: notes on the replacement of th by s in the third person singular in English. *English Studies*, 5: 406–31.

Van de Geer, J. (1971). *Introduction to Multivariate Analysis for the Social Sciences*. San Francisco: W. H. Freeman.

Van Peer, W. (1989). Quantitative studies in literature: a critique and an outlook. *Computers and The Humanities*, 23: 301–7.

White, M. (1992). *Middleton and Tourneur*. London: Macmillan.

Downloaded from http://llc.oxfordjournals.org/ at Yale University on July 10, 2015