Now I am alone:
A corpus stylistic approach to Shakespearian soliloquies

Sean Murphy

Lancaster University

Abstract

A popular interest in Shakespeare has been matched in recent years by an increasing number of computer-assisted analyses of the plays. Although not without their critics, corpus stylistic studies have offered scope and reliability in the study of literary texts, particularly through key word analyses. In this paper, I show how *Wmatrix*, a web-based corpus processing environment (Rayson, 2003, 2007), in conjunction with other corpus tools, can systematically extend such key analyses from words, to parts of speech and semantic fields. By so doing, a greater understanding of linguistic aspects of an author's literary output may be achieved. This study is based on a key word, grammatical category and semantic field analysis of soliloquies and asides in 12 Shakespeare plays. An investigation of the linguistic characteristics of soliloquies/asides as opposed to dialogic speech reveals the overuse of the interjection *O* and words related to the body. Comparisons of soliloquies across genres tend to match intuitive assumptions. Finally, soliloquies written in the later period (1596-1606) tend to have a far greater proportion of 'the (noun) of (noun phrase)' structures. The paper ends by suggesting that more empirical work of this nature is needed to underpin qualitative literary judgements.

1. Introduction

Shakespeare, it seems, is as popular as ever. In 2006, the Royal Shakespeare Company produced a cycle of all his plays. The historian and broadcaster Michael Wood has documented his life in a TV series entitled *In Search of Shakespeare* (Wallace, 2003), while modern adaptations of his plays bring them to new audiences in *Shakespeare Retold* (Percival *et al.*, 2006). New books abound on almost every aspect of his life and work (e.g. Shapiro, 2005), and Rough Guides, better known for travel guides, has even produced *The Rough Guide to Shakespeare* (Dickson, 2005). Shakespeare's celebrated use of language continues to receive attention (Alexander, 2004; Kermode, 2000), while an anthology of soliloquies displays "Shakespeare's poetic genius in all its richness" (Kerrigan, 2002). At the same time, computers and the Internet are having a considerable influence on approaches to Shakespeare with online databases such as WordHoard (Mueller *et al.*, 2006), Shakespeare's Words (Crystal & Crystal, 2004) and Hyperhamlet (Engler *et al.*, 2003) offering users multiple opportunities to investigate all aspects of the Bard's output.

Parallel to this popular interest in Shakespeare, there has been a rise in the use of computer-assisted textual analysis in the field of literary stylistics, although not as much as one might expect (Wynne, 2005) and met with resistance by some literary critics (see Louw, 1997). In its defence, McEnery & Wilson (2001) point out that the whole concept of style rests on the notion that authors choose to express their ideas using certain linguistic resources in preference to others, which logically must be measurable to some degree. While complete objectivity in the analysis of style may be an unattainable goal, "without quantitative confirmation, statements on style lack the support of concrete evidence" (Leech & Short, 1981). The underlying characteristics of a corpus-based approach to language study are that it is empirical, it uses a large corpus of natural texts collected on a principled basis, computer and manual analyses of the corpus are carried out, and both quantitative and qualitative techniques of analysis are employed. The consequent strengths of such an analysis lie both in its scope and reliability (Biber *et al.*, 1998).

Another advantage of using computer software that has been pointed out is its ability to identify potentially significant textual features which have gone unnoticed by literary critics (Stubbs, 2005). Researchers (e.g. Barnbrook, 1996) often advocate using a larger general corpus as a norm for comparison against which smaller corpora can be measured to enable us to "marshal hosts of instances too numerous for our unassisted powers" (Burrows, 2002). Indeed, it has been with the aid of computers that some of the most interesting stylistic work on Shakespeare has taken place in recent years. Recent corpus-based studies have examined aspects such as differences between female and male language (Sobhan Raj Hota & Moshe Koppel, 2006), key semantic domains and metaphor in love tragedies and love comedies (Archer *et al.*, 2005), characterisation in *Romeo and Juliet* (Culpeper, 2001), and imagery in *Macbeth* (Zyngier, 1999).

In presenting the arguments above, I do not, of course, intend to suggest that quantitative techniques in corpus-based approaches to the analysis of literary style can or should supplant qualitative analysis. Semino & Short (2004) make the valuable point

that one need not preclude the other, and indeed they should be interdependent. It is my firm belief that quantitative treatments of corpora are preliminary steps to qualitative assessments of the resulting output. Like Semino & Short, I do not see corpus-based studies as competing with other forms of stylistics, but rather strengthening the analytical rigour of stylistic research.

Within corpus linguistics, the notion of key words, words whose frequency in a corpus is unusually high when compared with a norm,[1] has become a useful way of characterising the 'aboutness' of a text or corpus. Two word-lists, from the study and reference corpora, are compared. By doing this, it is possible to find out which words characterize the former, the latter providing data for reference comparison. Software such as *WordSmith Tools* (Scott, 1999) and *Wmatrix* (Rayson, 2003, 2007) use the log-likelihood statistical test to calculate keyness. Key words which are statistically significantly more frequent in the study corpus than in the reference corpus are called 'positive key words', and statistically significantly infrequent ones 'negative key words'. The basic premise behind this study is that key words, key parts of speech or key semantic fields as defined by software can provide the stylistician with a valuable tool in textual analysis.

The overall aim of this paper then is to show what a corpus stylistic analysis can reveal about the linguistic nature of soliloquies as opposed to dialogue in Shakespeare's plays, and to what extent this methodological approach can highlight distinctions between comedies, histories and tragedies, and early plays as opposed to mid-career works. For the purposes of this study, I define *soliloquy* as a speech spoken by a single actor who does not intend the words to be heard by any other character (Hirsh, 2003). I have also included asides not directed at another character as instances of self-talk. Following an exploration of the nature of soliloquies and self-talk in section 2, I provide the rationale behind my selection of plays for consideration in section 3. In section 4, I demonstrate how I used Paul Rayson's *Wmatrix* web-based corpus processing environment (Rayson, 2003, 2007) to carry out a key word analysis of various corpora of soliloquies. I present the findings and discussion of my analysis in section 5. Section 6 concludes the paper by summarising the main results and suggesting avenues for further research.

2. Soliloquies and self-talk

The fourth-century work *Soliloquies* by Saint Augustine is, in fact, a dialogue between Augustine and the personification of a faculty of his mind called Reason. The term *soliloquy*, from the Latin *solus* (alone) and *loqui* (place), was not used in a theatrical sense until the late seventeenth century and consequently, it is highly unlikely that Shakespeare was familiar with the term (Crystal & Crystal, 2005). As already stated above, Hirsh (2003) defines the word *soliloquy* as a speech spoken by a single actor who does not intend the words to be heard by any other character. He notes, however, that

---

[1] The two corpora are referred to as the 'study corpus' and the 'reference corpus' respectively.

the word has tended to be used indiscriminately to refer to three types of theatrical practice, namely:

- audience-addressed speech, in which the character is aware of and speaks to playgoers;
- self-addressed speech, in which the character is unaware of playgoers and speaks only to herself; and
- interior monologue, in which the words merely represent thoughts passing through the character's mind.

No evidence exists of a soliloquy representing interior monologue before the middle of the seventeenth century, claims Hirsh (2003: 18), so consequently only the first two conventions were employed by Shakespeare. Similarly, Hirsh notes that the term *aside* has been used to describe four types of stage behaviour:

- a speech directed at one character, but guarded from another;
- a speech directed at playgoers, but guarded from others onstage;
- a speech addressed by a character to herself, despite the presence of others; and
- an interior monologue, representing thoughts passing through a character's mind, despite the presence of others onstage.

Of the four, all but the first match Hirsh's definition of soliloquy, and together with the first two types of soliloquy, could also be described as 'self-talk' (Goffman, 1981). Goffman (1981) notes that people can and do make comments aloud when solitary, and by so doing, split themselves in two, both projecting talk and being an appropriate recipient of the same. Thus natural interaction undergoes a transformation. In a theatre, a secondary transformation occurs as an actor plays a character, members of the audience existing merely as "supernatural out-of-frame eavesdroppers". Such ritualized ethological transformations have been called "routine licences of situation" (Leech, 1969: pp.186). In this paper, I will interchange the terms *soliloquies/asides* and *self-talk* for reasons of stylistic variation.

Although aspects of Shakespearean soliloquies have been studied in terms of linguistic theories such as politeness (Brown & Gilman, 1989), speech acts (Porter, 1979; Rudanko, 1993), Gricean maxims (Gilbert, 1995, 1997) and cognitive metaphor theory (Freeman, 1998), no corpus linguistic study of soliloquies has, to the best of my knowledge, thus far been undertaken. Soliloquies have, of course, received extensive critical attention, but only a few studies are devoted solely to the subject (Arnold, 1911; Clemen, 1964, 1987; Hirsh, 2003; Newell, 1991; Skiffington, 1985). The research presented in this paper aims to shed a little more light on what in my view is an aspect of Shakespeare's work which has much to gain from a corpus stylistic methodological treatment.

3. Selection of plays

A comprehensive study of self-talk in all of Shakespeare's plays was beyond the scope of the present study, so a principled selection of plays was necessary to address my research questions. In order to have a representative enough corpus to be reasonably sure of my findings, I decided to base my study on 12 plays: four comedies, four histories and four tragedies, with two of each from an early period (1591-95) and two from a later period (1596-1606), all as far as possible containing a large degree of self-talk.

To determine which plays had most self-talk, I turned to Arnold (1911), who provides a soliloquy line count for each play. By taking the total number of lines per play (Dunton-Downer & Riding, 2004), I was able to calculate Arnold's lines of soliloquy as a percentage of the number of lines per play (see Table 1 below). Although there were obvious problems with this approach, not least in that Arnold provides no details of how he made his line count, and also that the number of lines per play has always been a hot topic of debate (e.g. Hart, 1932), I felt the calculation provided a reasonable enough basis for my selection. The dates given for each play are based on the earliest conjectured dates of composition (Dobson & Wells, 2001; Wells & Taylor, 1987).

Table 1: Arnold's soliloquy line count as a % of Shakespeare's plays

Play	% of play	Play	% of play
1. *Two Gentlemen of Verona* (1590)	13.30	20. *Much Ado About Nothing* (1598)	4.57
2. *Henry VI Part 3* (1591)	11.97	21. *A Winter's Tale* (1609)	4.55
3. *Cymbeline* (1610)	11.46	22. *Alls Well That Ends Well* (1604)	4.20
4. *A Midsummer Night's Dream* (1595)	10.95	23. *Troilus and Cressida* (1602)	4.13
5. *Macbeth* (1606)	9.89	24. *Henry V* (1598)	4.06
6. *Romeo and Juliet* (1595)	9.47	25. *Pericles* (1607)	4.06
7. *Twelfth Night* (1600)	8.58	26. *The Comedy of Errors* (1594)	3.47
8. *Timon of Athens* (1605)	8.36	27. *King John* (1596)	3.44
9. *The Merry Wives of Windsor* (1597)	7.47	28. *Henry VI Part 1* (1592)	3.33
10. *Hamlet* (1600)	7.23	29. *Titus Andronicus* (1592)	3.32
11. *Henry VI Part 2* (1591)	7.07	30. *The Tempest* (1611)	3.21
12. *Richard III* (1593)	6.59	31. *The Taming of the Shrew* (1591)	2.95
13. *Julius Caesar* (1599)	5.99	32. *Richard II* (1595)	2.82
14. *Love's Labour's Lost* (1594)	5.55	33. *Antony and Cleopatra* (1606)	2.57
15. *Henry IV Part 2* (1597)	5.44	34. *Henry VIII* (1613)	1.82
16. *King Lear* (1605)	5.29	35. *The Merchant of Venice* (1596)	1.54
17. *Othello* (1604)	4.83	36. *As You Like It* (1599)	1.35
18. *Henry IV Part 1* (1596)	4.67	37. *Coriolanus* (1608)	0.94
19. *Measure for Measure* (1603)	4.61		

As mentioned above, I decided to concentrate on four plays each from the genres of comedy, history and tragedy, two of each genre being from the early part of Shakespeare's career, and two from a later period. Looking at Table 2, it can be seen that most plays (in **bold**) selected themselves on the basis of percentage of soliloquy:

Table 2: Plays selected for analysis

	1590-1595	1596-1606
Tragedy	*Titus Andronicus (1592)*	*Hamlet (1600)*
	Romeo and Juliet (1595)	***Macbeth (1606)***
Comedy	***Two Gentleman of Verona (1590)***	***The Merry Wives of Windsor (1597)***
	A Midsummer Night's Dream (1595)	***Twelfth Night (1600)***
History	***Henry VI Part 2 (1591)***	***Henry IV Part 1 (1596)***
	Henry VI Part 3 (1591)	***Henry IV Part 2 (1597)***

I selected *Titus Andronicus* despite it having relatively little soliloquy, because it was the only other tragedy apart from *Romeo and Juliet* written in the early period and I felt it might reveal something about Shakespeare's earliest attempts at tragic soliloquy. Although *Cymbeline,* classed as a tragedy in the First Folio of 1623, comes third in the list, I rejected it as most critics consider it to be a romance or tragicomedy (e.g. Greenblatt *et al.*, 1997). Similarly, *Timon of Athens* has more tragic soliloquy than *Hamlet*, but the latter ("the one Shakespearian tragedy from which almost every speaker of English can quote at least one or two phrases" (Dobson & Wells, 2001:179)), particularly in respect of its soliloquies, obliged me to exclude the former, "which enjoys the dubious distinction of being perhaps the least popular play in the Shakespeare canon" (Dobson & Wells, 2001:475).

4. Methodology

Having selected plays for analysis, my next task was to choose a reliable electronic source for the texts. I decided to use WordHoard (Mueller *et al.*, 2006), a joint project of the Perseus Project at Tufts University, The Northwestern University Library, and Northwestern University Academic Technologies. The text is derived from *The Globe Shakespeare*, the one-volume version of the *Cambridge Shakespeare* (Clark *et al.*, 1891-3). Standardized spellings are used throughout, an important consideration when doing any form of corpus analysis. As I constructed my corpora of soliloquies and asides, I consistently compared the WordHoard text with *The Norton Shakespeare* (Greenblatt *et al.*, 1997) to ensure greater reliability. It was then a relatively simple task to copy WordHoard versions of the plays into Microsoft Word documents, and then proceed to separate out soliloquies and asides from interactional language based on a careful

reading of the 12 plays and comparison with the Norton text. I also removed all character names and stage directions using simple Find and Replace procedures.[2]

I thus created a soliloquy/aside document and an interactional language document for each play, and calculated the percentage of self-talk per play based on token[3] counts. For example, we can see from Table 3 (overleaf) that *Hamlet* contains 2,302 words of self-talk, of which 2,260 words are in soliloquies and 42 words in asides. I also named combined files, T SOL being all self-talk in tragedies, etc. Thus, the four tragedies contain 8,072 words of self-talk, a figure which represents 9% of all the words in these four plays. In addition, we can see that the 'Total T' row shows how many words of self-talk occur in tragedies written from 1590-95 (3,589) and those between 1596-1606 (4,483). The total amount of early self-talk across the four genres represents a total of 13,110 words, compared to 11,539 in the later group. The total amount of self-talk for all 12 plays is 24,649 words, and for all interactional talk, 236,149 words.

Needless to say, there were a number of problem cases where a principled decision had to be taken for a passage's inclusion or exclusion as self-talk. I included the following as instances of self-talk:

- where a character continues to soliloquize, unaware of another character, as in *Titus Andronicus* III.1.22-26, when Lucius' entrance goes unnoticed by Titus. Similarly, in *Romeo and Juliet* II.2.23-30, Friar Laurence is not conscious of Romeo's presence. Or it may be the case that other characters are asleep, as in *A Midsummer Night's Dream* III.3.36-46;
- where a stage direction makes it clear that a character's words are an aside, as the character comes onstage but fails to see that other characters are already there, as when Proteus comes onstage in *The Two Gentlemen of Verona* I.3.45-50;
- where one character is aware of another's presence but their utterances are private, as in Romeo's speeches in the famous balcony scene in *Romeo and Juliet* II.1.67-74;
- where a character recites lines of verse, as Puck does in *A Midsummer Night's Dream* III.2.397-400;
- where asides occur mid-speech or even mid-line, as in *Henry VI Part 2* I.1.207 or *Henry IV Part 1* III.3.189-90 when Falstaff calls for the Hostess;
- where a character reads from a letter, as Malvolio does in *Twelfth Night* II.5. The contents of the letter are not counted as self-talk, but the character's commentary on the letter is;
- where a whole scene is a series of soliloquies, as in *Henry VI Part 3* II.5;

---

2 Another possibility would have been to manually tag the text for characters, soliloquies and asides and then use text extraction software to create soliloquy files, etc. Although this would have been time-consuming and unnecessary for the purposes of the present study, I would not rule out such a possibility for future studies as it may have analytical advantages over the present method.

3 Tokens are each word form which occurs in a text, irrespective of whether they are repeated or not (as opposed to types, which are multiple instances of the same token) (Stubbs, 2002).

- where madness is concerned, as in the cases of Lady Macbeth, *Macbeth* V.1; Ophelia, however, in *Hamlet* IV.5 explicitly directs her comments to the other characters, and cannot therefore be said to be engaging in self-talk;
- where soliloquy-type speeches are presumably overheard, as in *Macbeth* V.3.20-30 and V.5.16-27;
- where a character quotes another in soliloquy, as in *Twelfth Night* I.5.259-261;
- where there are speeches by formal choruses, as in *Henry IV Part 2* (Rumour and Epilogue), *Romeo and Juliet* (Prologue and Chorus II.0). *A Midsummer Night's Dream* (Puck's Epilogue). Hirsh (2003) justifies such audience-addressed speeches as soliloquies on the basis that they are spoken by a single actor who does not intend other characters to hear them.

Table 3: Token counts for self-talk in 12 Shakespeare plays

	1590-1595	Self-talk (sol./aside[4])	1596-1606	Self-talk (sol./aside)	Self-talk totals by genre (as % of plays)
Tragedy (T)	*Titus Andronicus (1592)*	827 (579/248)	*Hamlet (1600)*	2,302 (2,260/42)	**8,072 (9.00%) T SOL**
	Romeo and Juliet (1595)	2,762 (2,566/196)	*Macbeth (1606)*	2,181 (1,533/648)	
Total T		**3,589**		**4,483**	
Comedy (C)	*Two Gentleman of Verona (1590)*	2,982 (2,699/283)	*The Merry Wives of Windsor (1597)*	1,532 (1,354/178)	**7,845 (10.75%) C SOL**
	A Midsummer Night's Dream (1595)	1,735 (1,673/62)	*Twelfth Night (1600)*	1,596 (1,409/187)	
Total C		**4,717**		**3,128**	
History (H)	*Henry VI Part 2 (1591)*	2,026 (1,641/385)	*Henry IV Part 1 (1596)*	1,814 (1,814/0)	**8,732 (8.91%) H SOL**
	Henry VI Part 3 (1591)	2,778 (2,596/182)	*Henry IV Part 2 (1597)*	2,114 (2,045/69)	
Total H		**4,804**		**3,928**	
Total by period (as % of plays)	**13,110 (10.53%) EARLY SOL**		**11,539 (8.47%) LATER SOL**		**24,649 (9.45%) ALL SOL**
Total interactional language[5](as % of plays)					**236,149 (90.55%) ALL DIA**

---

4 All asides except those directed at another character.
5 Spoken and written.

Having checked that the word totals of the combined files corresponded with the figures in Table 3, I then saved the documents as plain text files and uploaded them to *Wmatrix* to create 'workareas'. I used the Early Modern English (EmodE) version, in which a variant detector regularizes words such as *i'* to *in*, or *hath* to *has* before part-of-speech tagging occurs. The CLAWS (Constituent Likelihood Automatic Word-tagging System)[6] tagger (96-97% accuracy) adds part-of-speech (POS) tags to the text before the USAS (UCREL[7] Semantic Analysis System)[8] tagger (92% accuracy) adds semantic field tags, based on 21 general domains and 232 semantic fields[9]. The next step in my analysis of linguistic features of soliloquies and asides was to compare smaller frequency lists against larger normative corpora (on the basis of the token counts in table 3). So, for example, I compared the smaller All SOL (all soliloquy) corpus with the larger All DIA (all dialogue) corpus to see how self- talk compared with interactional language use. The complete list of corpora I compared is shown in table 4.

Table 4: Corpora used for key analysis

Smaller corpus		Larger corpus
All SOL		All DIA
Comedy SOL		History SOL
Tragedy SOL	compared with	History SOL
Tragedy SOL		Comedy SOL
Later SOL		Early SOL

I carried out these comparisons at the word level, the POS level, and the semantic level. *Wmatrix* employs the log-likelihood (or G2) statistic to calculate significant key items.[10] The advantage of the G2 statistic is that it does not assume a 'normal distribution' of words in a text, nor does it over-estimate the significance of rare events, as the commonly-used Pearson chi-squared test does (Leech *et al.*, 2001). My aim in this study was to look for statistically highly significant items with log-likelihood (LL) values of above 15.13 ($p < 0.0001$ 1 d.f.).

There were two further stages in my analysis. One was to look for n-grams (consecutive word sequences) among a number of selected findings from the previous stage by using a Multilingual Corpus Toolkit (Piao *et al.*, 2002). This allowed me to uncover further layers of meaning than those provided by individual words. The other was to carry out concordances on selected findings using *WordSmith Tools* (Scott, 1999) to determine frequent collocations. Both of these procedures are referred to in the next section.

---

[6] See http://www.comp.lancs.ac.uk/ucrel/claws/ for further information.

[7] University Centre for Computer Corpus Research on Language.

[8] See http://www.comp.lancs.ac.uk/ucrel/usas/ for further information.

[9] See http://www.comp.lancs.ac.uk/ucrel/usas/semtags.txt for the complete list.

[10] See http://ucrel.lancs.ac.uk/wmatrix.html for further details

5. Findings and discussion

Before analysing the *Wmatrix* output, the first thing I noticed was that my own token count data generally gave higher percentages than Arnold's line counts (see Figure 2), particularly in the case of *The Two Gentlemen of Verona*, and to a lesser extent, *Henry IV Part 1* and *Macbeth*. This was probably due to my inclusion of asides as instances of self-talk, *The Two Gentlemen of Verona* having no fewer then 57 lines / 283 words of asides. The data also provides some clues to one of my research questions, namely how soliloquies evolved over Shakespeare's career.

With the selected plays arranged in chronological order of composition, the graph in Figure 1 shows that both Arnold's and my own 'Lineal' lines display a noticeable tendency towards a general reduction in self-talk as a percentage of the play, *Titus Andronicus* and *Macbeth* being exceptions to this trend.
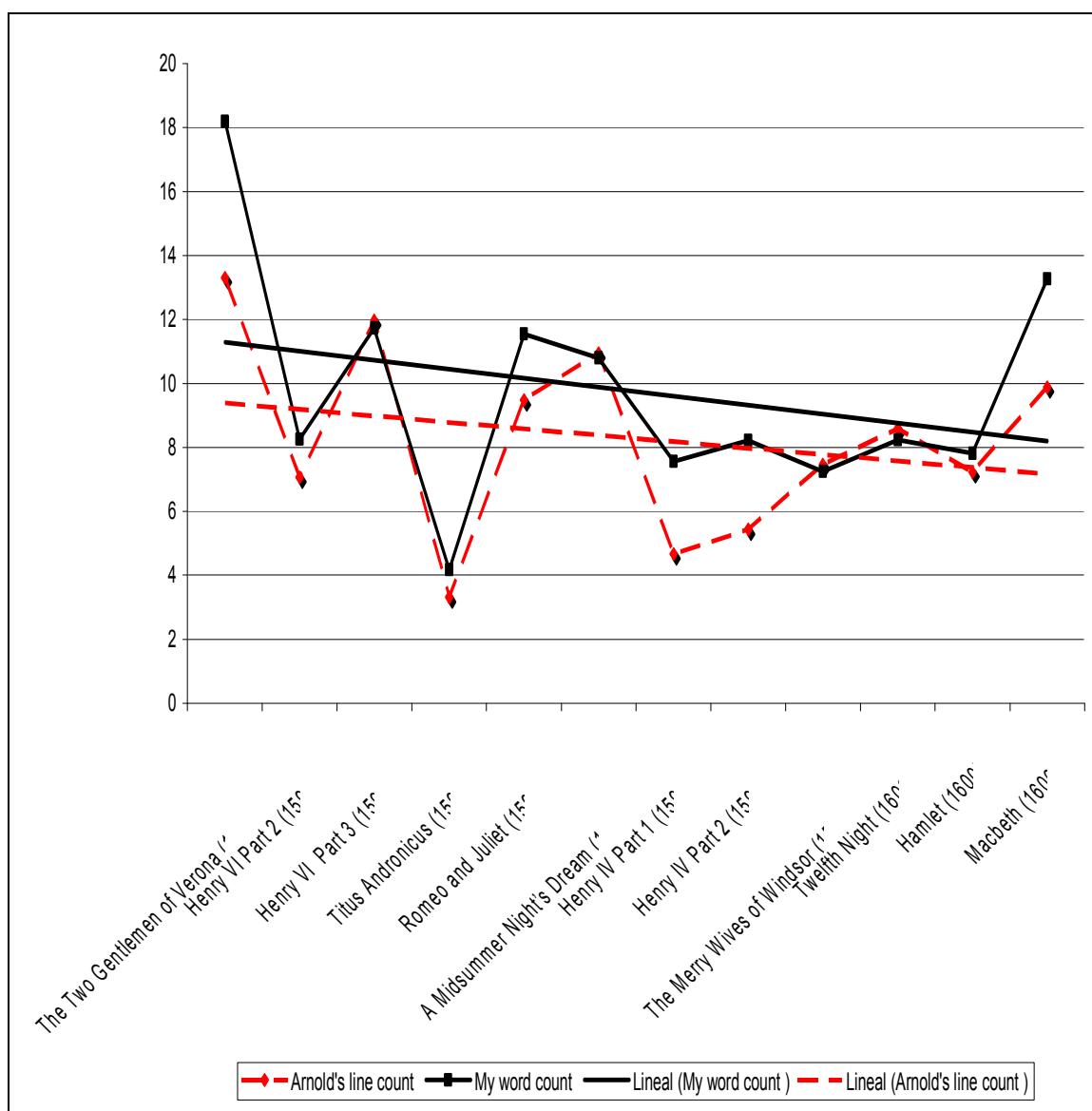

Figure 1: Self-talk as a % of selected plays

How could this be explained? Did Shakespeare gradually become less interested in the convention of the soliloquy, or was there a conscious or even a subconscious change in style which required fewer words? Skiffington (1985:117) offers the following explanation:

> Primitiveness is undeniably resident in numerous speeches from the histories presumed to have been written within the first five or six years […].The greatest incidence of sophistication occurs in tragedy soliloquy of the later fifteen years or so, as in […] *Hamlet* […] and *Macbeth*.

He goes on to assert that both primitiveness and sophistication are present in all soliloquies but the early plays comprise "extravagances in language, such as elaborate conceits, catachresis, the veritable piling-on of images, and occasional floods of Senecan rhetoric." He uses the following example from *Henry VI Part 3*, in which Henry ponders the simplicity of a shepherd's existence, to make his point:

> O! God! methinks it were a happy life,
> To be no better than a homely swain;
> To sit upon a hill, as I do now,
> To carve out dials quaintly, point by point,
> Thereby to see the minutes how they run,
> How many makes the hour full complete;
> How many hours brings about the day;
> How many days will finish up the year;
> How many years a mortal man may live.
> When this is known, then to divide the times:
> So many hours must I tend my flock;
> So many hours must I take my rest;
> So many hours must I contemplate;
> So many hours must I sport myself;
> So many days my ewes have been with young;
> So many weeks ere the poor fools will ean;
> So many years ere I shall shear the fleece:
> So minutes, hours, days, months, and years,
> Passed over to the end they were created,
> Would bring white hairs unto a quiet grave.
> Ah, what a life were this! how sweet! how lovely!
> (*Henry VI Part 3* II.5.21-41)

The repetition of *How many* and *So many hours must I* seems laboured and artificial.[11] Hussey (1982) criticizes the soliloquy as "too 'poetic' and too rhetorical", which suggests unwarranted prolixity. By contrast, Macbeth's regretful overheard soliloquy is replete with ordinary, common imagery.

---

[11] The Norton Shakespeare notes that this soliloquy of 54 lines from the Folio text is reduced to 13 lines in the Octavo text. Perhaps even audiences in the 1590s found it tiresome!

Seyton! — I am sick at heart,
When I behold — Seyton, I say! — This push
Will cheer me ever, or disseat me now.
I have lived long enough: my way of life
Is fall'n into the sear, the yellow leaf;
And that which should accompany old age,
As honour, love, obedience, troops of friends,
I must not look to have; but, in their stead,
Curses, not loud but deep, mouth-honour, breath,
Which the poor heart would fain deny, and dare not.
Seyton!
(*Macbeth* V.3.20-30)

The contrast between the succinctness of "a 'real' man fallen from former grace and noblesse, as contrasted to poetic posturing suitable for fallen kings in romantic chronicle plays" (Skiffington, 1985: 129) may be representative of a reduction in quantity of soliloquy, but an increase in quality, possibly influenced by the popularity of essays by Montaigne and Cornwallis (Shapiro, 2005)[12]. It remained to be seen if the *Wmatrix* data would confirm or refute such a thesis (see section 5.3).

5.1 Soliloquies versus interactional language

5.1.1 Key words

Comparison of soliloquies/asides with interactional language enabled me to gauge the 'aboutness' of the former. I was particularly interested in common words with a relatively high frequency in both corpora, so I discounted low frequency items like proper names (e.g. Demetrius), and words with a highly localized significance in particular plays (e.g. *shoe* in Launce's soliloquy in *The Two Gentlemen of Verona* II.3.1-28). Table 5 shows the seven most positive key words (p < 0.0001) in soliloquies/asides:

Table 5: Positive key words in soliloquies/asides (p < 0.0001 1 d.f.)

Top 7 key words	Observed frequency in All SOL	Observed frequency in All DIA	Log-likelihood score
lord	16	850	+75.46
Jove	11	16	+51.88
O	120	621	+33.01
yet	83	404	+27.04
I	780	6203	+22.63
a	496	3775	+21.69
my	465	3554	+19.62

---

[12] I am grateful to Professor Keith Johnson for drawing my attention to this.

It is perhaps unsurprising to find vocatives (*lord*, *Jove*), an interjection (*O*), and words relating to self (*I*, *my*). *O* most frequently collocates with an adjective and noun:

> O excellent motion (3), O bosom black (3), O wretched state (2), O weary night (2), O reverend tribunes (2), O piteous spectacle (2), O limed soul (2), O heavenly Julia (2), O gentle sleep (2), O exceeding puppet (2), O bloody times (2).

This emphatic nature of these vocatives, interjections and pronouns not only acts as a powerful directorial indicator to actors, but also invites playgoers/readers to "live with the emotions of the characters" (Taavitsainen, 1998:195) and empathize with their joy or predicament.

The keyness of the item *yet* (as a sentence rather than temporal adverb) and its most frequent collocate *and* (25 of the 83 occurrences) reveals how self-talkers are often troubled by doubt and anxiety, a point which has previously been noted in relation to Juliet's character in *Romeo and Juliet* (Culpeper, 2001). The following examples reinforce this view:

> Laertes: (aside to Claudius) My lord, I'll hit him [Hamlet] now.
> Claudius: (aside to Laertes) I do not think't.
> Laertes: (aside) **And yet** it is almost against my conscience.
> (*Hamlet* V.2.238-240)

> Malvolio: M, O, A, I; this simulation is not as the former: **and yet**, to crush this a little, it would bow to me, for every one of these letters are in my name.
> (*Twelfth Night* II.5.122-124)

> Falstaff: I have forsworn his company hourly any time this two and twenty years, **and yet** I am bewitched with the rogue's company.
> (*Henry IV Part 1* II.2.15-17)

Investigation of the ten most negative key words revealed that six could be termed as 'interactional pronouns' (*you*, *your*, *we*, *thou*, *our* and *us*) - perhaps not a very surprising finding, but one which emphasizes the interpersonal function of such words in dialogue.

5.1.2 Key word classes

Table 6 shows positive key parts of speech. By examining the concordance for *comes*, the most common *-s* form of a lexical verb, it is possible to see the reason for its keyness, as 12 of the 20 occurrences collocate with *here* and serve as stage directions for the characters (Clemen, 1987; Hirsh, 2003):

> Falstaff: I have led my ragamuffins where they are peppered; there's not three of my hundred and fifty left alive; and they are for the town's end, to beg during life .
> [Enter Prince Harry]
> But who **comes here**?
> (*Henry IV Part 1* 5.3.35-38)

Table 6: Positive key parts of speech (p < 0.0001 1 d.f.)

Key word classes (*most common words within the class*)	Observed frequency in All SOL	Observed frequency in All DIA	Log-likelihood score
-s form of lexical verb (*comes, makes, gives, lies, speaks*)	358	2478	30.27
general adjective (*good, true, sweet, fair, great*)	1572	13004	28.66
1st person sing. subjective personal pronoun (*I*)	767	6105	22.05
plural common noun (*eyes, men, tears, words, thoughts*)	932	7683	17.79

General adjectives are key items in soliloquies/asides, with both positive (*good, true, sweet, fair*) and negative (*poor, old, dead, cold*) semantic prosodies. This suggests an evaluative function of soliloquies, allowing the audience to see how soliloquizers view people and events. As regards the third most key item, an N-gram analysis revealed that *I will (not) / I shall / will I / shall I* account for 160 of the 767 occurrences of *I* (20.9%), suggesting that a typical characteristic of self-talk is to reveal the speaker's intentions (Clemen, 1987). Plural common nouns are also significantly more frequent in soliloquies/asides, perhaps indicating a tendency for speakers to generalize as in the following examples:

> Helena: Love looks not with the **eyes**, but with the mind.
> (*A Midsummer Night's Dream* I.1.234)

> Aaron: Let **fools** do good, and fair **men** call for grace
> (*Titus Andronicus* III.1.203)

> King Henry: And let our **hearts** and **eyes**, like civil war,
> Be blind with **tears**, and break o'ercharged with grief.
> (*Henry VI Part 3* II.5.77-78)

This could be related to a moralizing function of many soliloquies, which may be partly due to their origins in medieval morality plays (Arnold, 1911).

5.1.3 Key concepts

Semantic analysis (Table 7) reveals that by far the most key semantic category in soliloquies/ asides is anatomy and physiology.

Table 7: Positive key concepts (p < 0.0001 1 d.f.)

Key semantic fields (*most common words within the field*)	All SOL	All DIA	LL score
anatomy and physiology (*eyes, heart, blood, sleep, hand*)	549	3579	+63.83
mental object: conceptual object (*thoughts, dream, thought, matter, subject*)	69	298	+29.88
colour and colour patterns (*golden, pale, black, light, white*)	116	643	+25.79
relationship: intimate/sexual (*love, kiss, cuckold, lovers, in love*)	111	636	+22.02
evaluation: true/false (*lie, lies, false, forsworn, deceive*)	70	360	+19.61

Nowhere is this more evident than in *Macbeth*, as the partial concordance in Figure 2 shows:

I may pour my spirits in thine	**ear** ;	And chastise with the valour o
chastise with the valour of my	**tongue**	All that impedes thee from the g
d fill me from the crown to the	**toe**	topfull Of direst cruelty ! make
direst cruelty ! make thick my	**blood**;	Stop up the access and passage
ct and it ! Come to my woman 's	**breasts**,	And take my milk for gall , yo
gement here ; that we but teach	**Bloody**	instructions , which , being tau
our poisoned chalice To our own	**lips**.	He 's here in double trust ; F
l blow the horrid deed in every	**eye**,	That tears shall drown the win
horrid deed in every eye , That	**tears**	shall drown the wind . I have no
efore me , The handle toward my	**hand**?	Come, let me clutch thee . I
ceeding from the heat-oppressed	**brain**?	I see thee yet , in form as pa

Figure 2: Anatomy and physiology semantic concordance in *Macbeth*

It has been suggested that in terms of cognitive metaphor theory, the body-based figurative language that pervades *Macbeth* and gives it such thematic intensity, is indicative of a CONTAINER schema (Freeman, 1998), of which the body is the most basic container, and the one which Macbeth violates by stabbing Duncan to death, a view which this data supports. The other key concepts (mental concepts, colour, intimate relationships, truth and falsity) would seem to back the contention that Shakespearian soliloquies are very much a "psychophysical blend of the abstract and the concrete" (Clemen, 1987).

5.2 Soliloquies compared by genre

The three-way comparison of genres involved comparing the smaller corpus with the larger one in each case. As I was interested in both corpora, I looked for both positive and negative key words, POS and semantic fields. The findings given in table 8

(overleaf) suggest that comedic self-talk is, as one would expect, very much about interpersonal and intimate relationships.

The keyness of *master* reminds us that these are Elizabethan comedies about masters and servants, yet it is the latter who wonder about the former, rather than vice-versa. Soliloquizing characters in both comedies and histories are more likely to use *I* than tragic heroes, but in histories, they also compare things, as in Falstaff's proverbial *The **better** part of valour is discretion* (*Henry IV Part 1* V.4. 117-118), and contemplate *war* (*soldiers, battles*). Tragic self-talk, on the other hand, seems more concerned with existential questions of *death* and religion (*heaven, soul, hell*) as the protagonists wrestle with their consciences and ethical dilemmas.

Table 8: Positive and negative key words in soliloquies compared across genres
(p < 0.0001 1 d.f.; * p < 0.001 1 d.f.)

Genres	Keyness	Word	POS	Semtag
Comedy History	+	*he, she, love*	3rd person sing. subjective personal pronoun (*he, she*)	pronouns; relationship: intimate/sexual
	-	*and, the*	plural common noun	warfare, defence and the army; weapons
Tragedy History	+	*Romeo, She*	*is**	religion and the supernatural*
	-	*I*	1st person sing. subjective personal pronoun (*I*)	evaluation: good/bad*
Comedy Tragedy	+	*I, her, master*	1st person & 3rd person sing. subjective personal pronoun (*I, s/he*)	pronouns (*I*)
	-	*death*, upon*, heaven**	plural common noun	life and living things (*death*)

5.3 Early soliloquies versus later soliloquies

My final task was to compare the later soliloquy corpus (11,458 tokens) with the early soliloquy corpus (13,110 tokens), the results of which are given in Table 9.

Table 9: Key words in the later soliloquy corpus compared to the early soliloquy corpus

	Keyness	Freq.	POS	Semtag
Later (1596-1606)	+	of	*of* (as preposition)	evaluation:- good/bad (*well*)
Early (1590-1595)	-	love	singular proper noun (*Jove, God*)	relationship: intimate/sexual (*love*)

The most interesting finding here was the overuse of *of* as a preposition in the later soliloquy corpus. Examples include *bank and shoal of time*, *a deal of scorn*, *whole school of tongues*, etc. Why did Shakespeare, consciously or subconsciously, begin to use *of*

significantly more frequently in instances of self-talk? To answer this question, I followed the example of Michel Stubbs (2005) and looked at recurrent lexico-grammatical patterns with *of*. *WordSmith Tools* (Scott, 1999) showed that *of* collocates most frequently with *the*, two places to the left in both corpora (early: *the* 61 *of* 207; later: *the* 103 *of* 325).

If we compare a representative selection of 24 such strings from each corpus (Table 10), we can appreciate the relative simplicity in the early soliloquies, many of the examples being titles or geographical names (5, 8, 10, 12, 17, 21), or fairly conventionalized comparisons (2, 3, 9, 14, 20). Relatively few are 'poetic' in that they bring together uncommonly matched words to create striking and memorable images, exceptions, perhaps, being 6, 7, 23, 24. Some have a poetic feel, but 1, 4, 11, 13, 15, 16, 18, 19, 22 may not be entirely original expressions. By contrast, many of the phrases in the later soliloquies are characterized by their unusually powerful figurative imagery (26, 27, 31, 32, 35, 37, 39, 41, 43, 45) and their alliterative and assonantal patterns (25, 28, 29, 30, 36, 44). Such an interpretation fits well with Skiffington's (1985) notion of "early primitiveness" and "later sophistication", but we can also reasonably conjecture that as Shakespeare's career developed, he may have become more aware of the fact that the 'the (noun) of (noun phrase)' structure afforded not only greater poetic possibilities, but also the potential for more varied expression and the suggestion of greater depth of feeling and heightened emotional states.

Table 10: Collocations of '*the* (noun) *of* (noun phrase)' in the early and later soliloquy corpora

Early (1590-1595)		Later (1596-1606)	
1.	**the** agent **of** her heart	25.	**the** adoption **of** abominable terms
2.	**the** beauty **of** the sun	26.	**the** badge **of** pusillanimity
3.	**The** brightness **of** her cheek	27.	**the** blanket **of** the dark
4.	**the** cradle **of** the fairy queen	28.	**the** cankers **of** a calm world
5.	**the** Duke **of** Suffolk	29.	**the** canopies **of** costly state
6.	**the** fury **of** this mad-bred flaw	30.	**the** dread **of** something after death
7.	**the** honey **of** thy breath	31.	**the** flame **of** bold rebellion
8.	**the** house **of** Lancaster	32.	**the** grief **of** a wound
9.	**the** manner **of** it	33.	**the** honey **of** his music
10.	**The** name **of** Henry the Fifth	34.	**The** insolence **of** office
11.	**The** name **of** valour	35.	**the** milk **of** human kindness
12.	**the** realms **of** England, France	36.	**The** observed **of** all observers
13.	**the** remembrance **of** my former love	37.	**The** pangs **of** despised love
14.	**the** secrets **of** the state	38.	**the** perfumes **of** Arabia
15.	**the** shepherd **of** thy lambs	39.	**the** ports **of** slumber open wide
16.	**the** soul **of** love	40.	**the** purging **of** his soul
17.	**the** state **of** Normandy	41.	**the** rearward **of** the fashion
18.	**the** terror **of** the place	42.	**the** salt **of** most unrighteous tears
19.	**the** thoughts **of** desperate men	43.	**the** seeds **of** Banquo kings
20.	**the** time **of** night	44.	**the** smile **of** safety
21.	**the** title **of** John Mortimer	45.	**the** table **of** my memory
22.	**the** Venus **of** the sky	46.	**the** uses **of** this world
23.	**the** wings **of** night	47.	**the** valour **of** my tongue
24.	**the** yoke **of** inauspicious stars	48.	**the** warming **of** the blood

6. Conclusions

In this paper, I have aimed to make a methodological contribution to the study of soliloquies by making accurate word counts of self-talk and its percentages in plays, and I have carefully tracked its development over Shakespeare's career. It is clear that there are important linguistic differences between self-talk and interactional language (such as the overuse of interjections like *O*, and the expression of doubt with *and yet*). We have also seen what soliloquisers do (they give implicit stage directions, they reveal future intentions, and they generalize). In terms of topics, they appear to talk at great length about anatomy and physiology, thoughts, colours, love and deception. Comedic soliloquies are very much about love and relationships; tragic heroes contemplate religion and the supernatural; and historical figures seem to contemplate themselves (*I*). In later soliloquies, we find an increasing number of '*the* (noun) *of* (noun phrase)' structures, displaying a growing maturity in poetic expression as Shakespeare's career as a dramatist progressed.

Many avenues for further research suggest themselves, the most obvious being to include more plays in the analysis, preferably all 38 in the canon, bearing in mind that a number of plays were probably co-authored. In terms of genre, it would also make sense to add the category of Romances to study plays such as *Pericles*, *A Winter's Tale*, etc. Gender-based comparisons (Sobhan Raj Hota & Moshe Koppel, 2006) could potentially yield significant results regarding the nature of female and male soliloquies. It would also be interesting to compare verse soliloquy with prose soliloquy, and individual character's soliloquies with their dialogic speech, perhaps even across plays, as in the case of Falstaff in *Henry IV 1*, *Henry IV 2*, *Henry V* and *The Merry Wives of Windsor*.

The limitations of automatic analysis of texts should always be recognized; after all, computers will still only do what humans tell them to do. Nonetheless, I believe that I have shown that a corpus linguistic approach to studying Shakespearean self-talk can highlight features of soliloquies and lead to new perspectives, thus enhancing the value of literary corpus stylistics.

References

Alexander, C.M.S. (ed.). (2004). *Shakespeare and Language*. Cambridge: Cambridge University Press.

Archer, D., Culpeper, J. & Rayson, P. (2005). *Love - a familiar or a devil? An exploration of key domains in Shakespeare's comedies and tragedies*, AHRC ICT Methods Network Expert Seminar on Linguistics. Lancaster University.

Arnold, M.L. (1911). *The Soliloquies of Shakespeare: A study in technic*. New York: The Columbia University Press.

Barnbrook, G. (1996). *Language and Computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Brown, R. & Gilman, A. (1989). Politeness theory and Shakespeare's four major tragedies. *Language in Society* 18, 159-212.

Burrows, J. (2002). The Englishing of Juvenal: Computational stylistics and translated texts. *Style* 36(4), 677-679.

Clemen, W. (1964). *Shakespeare's Soliloquies: The presidential address of the Modern Humanities Research Association, 1964*. London: Cambridge University Press.

Clemen, W. (1987). *Shakespeare's Soliloquies*. London: Methuen.

Crystal, D. & Crystal, B. (2004). *Shakespeare's Word*s. London: Penguin.

Crystal, D. & Crystal, B. (2005). *The Shakespeare Miscellany*. London: Penguin.

Culpeper, J. (2001). *Computers, Language and Characterisation: An analysis of six characters in Romeo and Juliet.* Paper presented at the ASLA symposium "Conversation in life and in literature", Uppsala, Sweden.

Dickson, A. (2005*). The Rough Guide to Shakespeare*. London: Rough Guides.

Dobson, M. & Wells, S. (eds.) (2001). *The Oxford Companion to Shakespeare*. Oxford: Oxford University Press.

Dunton-Downer, L. & Riding, A. (2004). *Essential Shakespeare Handbook*. London: Dorling Kindersley.

Engler, B., Hohl, R., Gebhard, C., Rosenthaler, L. & Fava, M. (2003). *Hyperhamlet*. Basel: Department of English, University of Basel.

Freeman, D. (1998). 'Catch[ing] the nearest way': Macbeth and cognitive metaphor. In Culpeper, J., Short, M. & Verdonk (eds.), *Exploring the Language of Drama: From text to context*. London: Routledge, 96-111.

Gilbert, A. (1995). Shakespearean self-talk, the Gricean maxims and the unconscious. *English Studies* 76(3), 221-237.

Gilbert, A. (1997). *Shakespeare's Dramatic Speech*. New York: The Edwin Mellen Press.

Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Philadelphia Press.

Greenblatt, S.J., Cohen, W., Howard, J.E., Maus, K.E. & Gurr, A. (eds.) (1997). T*he Norton Shakespeare: Based on the Oxford Edition.* New York and London: W.W. Norton.

Hart, A. (1932). The number of lines in Shakespeare's Plays. *Review of English Studies* VIII(29), 19-28.

Hirsh, J. (2003). *Shakespeare and the History of Soliloquies*. London: Associated University Presses.

Hussey, S.S. (1982). *The Literary Language of Shakespeare*. London: Longman

Kermode, F. (2000). *Shakespeare's Language*. London: Penguin.

Kerrigan, M. (ed.) (2002). *To Be or Not to Be: Shakespeare's soliloquies*. London: Penguin.

Leech, G. (1969). *A Linguistic Guide to English Poetry*. London: Longman.

Leech, G., Rayson, P. & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. Harlow: Longman.

Leech, G. & Short, M. (1981). *Style in Fiction: A linguistic introduction to English fictional prose.* London: Longman.

Louw, B. (1997). The role of corpora in critical literary appreciation. In Wichmann, A., Fligelstone, S., McEnery, T. & Knowles, G. (eds.), *Teaching and Language Corpora.* London: Longman, 240-251.

McEnery, T. & Wilson, A. (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Mueller, M., Parod, W., Cousens, J., Burns, P. & Norstad, J. (2006). *WordHoard*. Evanston, Ill.: Northwestern University.

Newell, A. (1991). *The Soliloquies in Hamlet: The structural design*. London: Associated University Presses.

Percival, B., Brozel, M., Richards, D., & Fraiman, E. (2006). *Shakespeare Retold*. UK: Acorn.

Piao, S., Wilson, A. & McEnery, A. (2002). A multilingual corpus toolkit. Paper presented at AAACL-2002, Indianapolis, Indiana, USA.

Porter, J. (1979). *The Drama of Speech Acts: Shakespeare's Lancastrian Tetralogy*. London: University of California Press.

Rayson, P**.** (2003). Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. Ph.D. thesis, Lancaster University. Also available at: http://www.comp.lancs.ac.uk/computing/users/paul/phd/phd2003.pdf

Rayson, P. (2007) Wmatrix: A web-based corpus processing environment, Computing Department, Lancaster University. http://www.comp.lancs.ac.uk/ucrel/wmatrix/

Rudanko, J. (1993). *Pragmatic Approaches to Shakespeare: Essays on "Othello", "Coriolanus" and "Timon of Athens"*. London: University Press of America.

Scott, M. (1999). *WordSmith Tools (Version 3.0)*. Oxford: Oxford University Press.

Semino, E. & Short, M. (2004). *Corpus Stylistics: Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.

Shapiro, J. (2005). *A Year in the Life of William Shakespeare*. London: Faber and Faber.

Skiffington, L.A. (1985). *The History of English Soliloquy: Aeschylus to Shakespeare*. London: University Press of America.

Sobhan Raj Hota, S.A. & Moshe Koppel, I.Z. (2006). *Performing Gender: Automatic Stylistic Analysis of Shakespeare's Characters*. Paper presented at the Digital Humanities Conference, Paris.

Stubbs, M. (2002). *Words and Phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Stubbs, M. (2005). Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature* 14(1), 5-24.

Taavitsainen, I. (1998). Emphatic language and romantic prose: Changing functions of interjections in a sociocultural perspective. *European Journal of English Studies* 2(2), 195-214.

Wallace, D. (2003). *In Search of Shakespeare*. UK: BBC.

Wells, S. & Taylor, G. (1987). *William Shakespeare: A textual companion*. Oxford and New York: Clarendon Press: Oxford University Press.

Wright, W. A., Clark, W. G. & Glover, J. (eds.). (1891-3). *The Works of William Shakespeare* (2nd ed.). London: Macmillan.

Wynne, M. (2005). Stylistics: corpus approaches. In *Encyclopedia of Linguistics*. Oxford: Elsevier.

Zyngier, S. (1999). "Smudges on the canvas?" A corpus stylistics approach to Macbeth. Paper presented at the PALA XIX Conference, Poetics, Linguistics and History: Discourses of War and Conflict, Potchefstroom University, South Africa.