I will proclaim myself what I am: Corpus stylistics and the language of Shakespeare's soliloquies

Abstract

This article reports on a corpus stylistic study of the language of soliloquies in Shakespeare's plays. Literary corpus stylistics can use corpus linguistic methods to test claims made by literary critics and identify hitherto unnoticed features. Existing literary studies of soliloquies tend to define and classify them, to trace the history of the form or to offer literary appreciation; yet they pay surprisingly little attention to the language which characterises soliloquies. By creating a soliloquy corpus and a dialogue corpus from 37 Shakespeare plays, and comparing the former against the latter using *WordSmith Tools*, I identify key language forms in soliloquies. Using an analytical framework broadly based on Halliday's ideational, interpersonal and textual metafunctions of language, I interpret my results and relate them, where possible, to literary critical interpretations. I also compare comedy, history and tragedy soliloquy corpora. My main findings show the following linguistic features to be characteristic of soliloquies in general: words relating to mental states and the body; pragmatic noise; linking adverbials and first-person pronouns. Characteristic forms in comedy, history and tragedy emphasise love, the monarch and the supernatural respectively. The empirical evidence presented here shows that Shakespeare regularly exploited certain language forms in soliloquies to represent expressions of doubt, resolve, introspection and strong emotion, among others. These forms not only add depth to characterisation, aid plot development and provide performance cues for actors, but may also conform to certain audience expectations.

Keywords

Soliloquy, soliloquies, Shakespeare, literary stylistics, keywords, corpus linguistics

Introduction

In this article, I identify characteristic language forms[1] in Shakespeare's soliloquies, commenting on their stylistic effects, and comparing these with claims made by literary critics. My findings are based on a corpus stylistic study of soliloquies and self-directed asides in 37 plays[2]. To provide some context, I present an extract from *The Merry Wives of Windsor* (III.v) in which Falstaff has just informed Brooke (Master Ford in disguise) that he has arranged for him to meet Mistress Ford (Master Ford's wife) so that he (Brooke) might 'enjoy' her. In prose soliloquy[3], Ford fumes at this threat to his reputation and resolves to find Falstaff and be very angry.

> Hum! ha! is this a vision? is this a **dream**? do **I sleep**?
> Master Ford, awake! awake, Master Ford! there's a hole made
> in your best coat, Master Ford. This it is to be married: this
> it is to have linen and buck-baskets! Well, **I will** proclaim **myself**
> what **I am**: **I will** now take the lecher; he is at my house; he
> cannot scape me; it is impossible he should; he cannot creep
> into a halfpenny purse, nor into a pepper-box; but, lest the devil
> that guides him should aid him, **I will** search impossible places.
> Though what **I am I** cannot avoid, **yet** to be what **I would not**,
> shall not make me tame: if **I** have horns to make me mad, let
> the proverb go with **me**; **I'll be** horn-mad.
>
> *The Merry Wives of Windsor* III.v

Besides being a contextually comic representation of a jealous, suspicious husband[4], Ford's soliloquy contains a number of language forms (in bold) which, as I will show, characterise soliloquies in general and produce particular effects. The repeated use of first-person pronouns (*I*, *me*, *my*, *myself*) suggests egocentrism; *dream* and *sleep* indicate a focus on states of unconsciousness; the use of *I will* / *I'll* shows the character resolving on courses of action or behaviours, thereby creating expectations for readers/audiences as to whether, when and how Ford will follow through with his resolutions and what consequences they

might have; self-definition (*I am*) indicates a degree of self-awareness, though this may be partial or inaccurate; finally, *yet* helps create the fiction of thought taking place as Ford weighs things up. My aim in this paper is to show, with empirical evidence, that these language forms, among others, characterise soliloquies across the Shakespeare canon.

1 Background

Given the renown of many passages of Shakespearean soliloquy, it is perhaps surprising that there are only four book-length studies (Arnold, 1911; Clemen, 1987; Hirsh, 2003; Skiffington, 1985), and these pay relatively little attention to language. Critics generally agree, however, that the language of soliloquies is somehow distinctive. For example, Clemen notes that 'Shakespeare, exploiting the conventions of the soliloquy, has nowhere moved farther away from "conventional language" than in his soliloquies' (1964: 24). While critics acknowledge distinctiveness, they fail to specify its precise linguistic nature: 'We see that [in soliloquy] normative notions of dialogue have been transformed into a profoundly subjective range of devices' (Gilbert, 1995: 222); [Soliloquy], 'since it needed to show the character resolving conflicting impulses, demanded a deliberate choice of syntax' (Hussey, 1992: 182). Similarly, Clemen draws attention to 'the pre-eminent question of the language of the soliloquies' (1987: 11), though does not answer his own question. Nonetheless, these authors do comment on a number of characteristic language forms.

The word 'thought' frequently occurs in soliloquies, particularly in *Hamlet*, emphasising the protagonist's 'indubitably introspective attitude', which serves to 'accentuate the meditative mood' (Arnold, 1911: 143). The theme of sleep also pervades many soliloquies, whether it be preparing for sleep, talking in one's sleep or apostrophizing sleep or sleepers (Arnold, 1911), as in Iachimo's soliloquy over the sleeping Imogen: *O sleep, thou ape of death, lie dull upon her* (*Cymbeline* II.ii). Arnold also observes that it is customary for a soliloquy to conclude with a signal that another character is approaching, highlighting *come(s)* as 'an easy way of acquainting the spectatnor with the new arrival' (1911: 54).

The apostrophe, 'an emotive address to an absent person, or to an inanimate object or abstraction, as if personified' (Adamson, et al. 2001: 281), is particularly common in 'impassioned and introspective soliloquies' (Arnold, 1911: 136) and often serves as a substitute in the absence of a hearer. Absent characters and loved ones are apostrophised, as

are the night, moon, sleep, heart, eyes, thoughts and the elements. Hirsh (2003) claims apostrophes imbue soliloquies with some of the dynamism of dialogue, thereby avoiding the risk of them being undramatic. Self-address by second-person pronoun is another common feature, **What dost thou**, *or what art thou, Angelo?* (*Measure for Measure* II.ii), as is address to part of oneself: *Hold, hold, **my heart***! (*Hamlet* I.v).

The recurring 'I' in the opening soliloquy of *Richard III* suggests 'the speaker can hardly cease looking at himself' (Clemen, 1987: 17). In *Twelfth Night*, Malvolio's five instances of *I will* sound 'almost like a statement of policy', says Clemen (1987: 57): ***I will be proud, I will** read politic authors, **I will** baffle Sir Toby, **I will** wash off gross acquaintance, **I will** be point-devise the very man.* (*Twelfth Night* II.v.). Hussey (1992) points out that one of the functions of a soliloquy is to reveal characters' future intentions, thereby preparing the reader/audience for subsequent developments. Regarding history, comedy and tragedy, critics have little to say on linguistic differentiation, except that '[t]he apostrophe is suggestive of the impassioned style of the soliloquies in the tragedies' (Clemen, 1987: 78).

While these literary critics rightly draw attention to common linguistic characteristics of soliloquies and highlight particular instances of certain language forms, there is no empirical evidence that such forms are characteristic of soliloquies in general. This study uses corpus linguistic and corpus stylistic methods to provide such evidence and relate it to the claims these critics make.

2 Corpus linguistics and corpus stylistics

Corpus linguistics aims to describe language use empirically. Generally, a smaller study (or 'target') corpus is compared against a larger reference corpus. The latter represents 'typical' usage against which the former can be measured. It is thus possible to identify distinctive linguistic features and move to interpret their significance. Literary corpus stylistics employs the methodology of corpus linguistics, but aims to relate linguistic description to literary appreciation (Mahlberg, 2013). One advantage of this approach is that it makes it possible to identify potentially significant textual features unnoticed by literary critics.

Techniques of analysis include searching for keywords, collocations[5] and lexical bundles[6]. Positive keywords have a significantly higher frequency in a study corpus than in a

reference corpus, while that of negative keywords is significantly lower. As all keywords are potentially indicators of style, Culpeper (2009) argues for categorisation based on Halliday's metafunctional view of language (e.g. Halliday, 1994), consisting of ideational, textual and interpersonal keywords. His analysis of *Romeo and Juliet* shows ideational keywords (*beauty*, *love*) to be typical of Romeo's speech; textual keywords (*if*, *yet*) to characterise Juliet's style, and more interpersonal keywords (*ah*, *O*, *madam*), to typify the Nurse's style. Such analyses can provide concrete evidence of the linguistic means through which characterisation is achieved.

Collocational analysis can often reveal evaluative associations. Stubbs (2002) finds, for example, that 'provide' frequently occurs with collocates describing things people need: 'help', 'money' and 'food'. Having the evaluative meaning 'good', 'provide' has a **positive semantic prosody** (Sinclair, 1996). By contrast, 'cause' is commonly found with 'problem(s)', 'damage', 'death(s)', and so has a **negative semantic prosody**. Fischer-Starcke's (2009) study of *Pride and Prejudice* revealed that in the semantic field of family relationships ('sister', 'family', 'daughter'), 'connection' differs from other words in more frequently having a negative semantic prosody (collocates included 'inferior', 'unequal', 'low') reflecting characters' concern with social status and prestige.

Lexical bundles are 'sequences of word forms that go together in natural discourse' (Biber *et al.* 1999: 990). Three-word bundles may be regarded as 'extended collocational association' (1999: 992). Common examples are 'I don't know' and 'do you want'. In Mahlberg's (2013) study of Dickens' novels, she argues that bundles such as *his hand upon his shoulder* or *his hand to his forehead* act as textual building blocks with local textual functions and can serve to compare linguistic techniques of characterisation with assertions found in literary appreciation.

For this study, I created corpora based on typologies of soliloquy and dialogue developed by Williams (1983) and Hirsh (2003). The main differentiating criteria were textual indications of direction of address. I classed instances of self-address (including self-directed asides) as 'soliloquy', and other-address (including other-directed asides) as 'dialogue'. I then compared the soliloquy corpus (55,870 tokens; 7,338 types[7]) with the dialogue corpus (762,447 tokens; 23,380 types). I also compared soliloquies across genres, based on the First Folio classification of comedies (15 plays, 20,413 tokens), histories (10 plays, 13,803 tokens) and tragedies (12 plays, 21,654 tokens), ignoring other categorisations such as tragicomedies, romances or problem plays. As a test for statistical significance, I

chose the log-likelihood test (Dunning, 1993) over the chi-square test as it is favoured as a better indicator of keyness (Scott, 2008). I experimented with different levels of significance and chose a significance or *p* value of 0.01 (a 1% risk of encountering fluke results).

3 Results

My analysis with WordSmith Tools produced 57 positive keywords (capitalised by the program and shown in Table 1), which I grouped using the aforementioned metafunctional categorisation. Limitations of space require me to restrict discussion to the forms in bold.

| N	Key word	Freq.	Keyness	N	Key word	Freq.	Keyness	N	Key word	Freq.	Keyness
1	**YET**	184	50,32	20	OUT	134	15,48	39	FIND	54	9,28
2	**DREAM**	33	48,86	21	ANOTHER	44	15,33	40	**MIND**	39	9,12
3	**O**	242	47,61	22	DOG	24	14,79	41	**COMES**	61	8,84
4	**I'LL**	201	47,15	23	**AM**	194	14,54	42	**BODY**	28	8,65
5	**MYSELF**	83	40,58	24	**NATURE**	43	14,23	43	**THUS**	74	8,43
6	**I**	1600	38,56	25	**EYES**	70	14,15	44	THAT	810	8,28
7	THIS	574	33,72	26	**SUN**	31	14,11	45	WAY	57	8,21
8	**SLEEP**	43	29,95	27	DEVIL	31	14,11	46	STILL	53	8,15
9	**SHE**	220	28,82	28	RUN	28	13,85	47	ITSELF	29	7,95
10	**BRAIN**	21	27,64	29	NIGHT	70	13,28	48	**BLOOD**	64	7,72
11	A	1107	26,24	30	THE	1984	13,13	49	BETTER	58	7,66
12	**HER**	339	24,14	31	**THOUGHTS**	32	13,05	50	UNDER	31	7,65
13	**CONSCIENCE**	24	21,75	32	CANNOT	77	11,16	51	LIE	38	7,52
14	**FIE**	28	21,62	33	WHEN	173	10,87	52	**THOUGHT**	39	7,31
15	MY	968	21,01	34	**EARTH**	36	10,48	53	NOW	228	7,23
16	**AND**	1818	19,49	35	THEIR	173	10,37	54	LIKE	157	7,04
17	MAKES	46	18,48	36	**LIGHT**	33	10,07	55	FALSE	31	6,96
18	HELL	27	17,43	37	**WORLD**	63	9,55	56	SAYS	27	6,87
19	**LOVE**	181	16,58	38	GOLD	28	9,52	57	**WIT**	29	6,71

Table 1. Positive keywords in soliloquies.

Ideational keywords: DREAM, SLEEP, CONSCIENCE, THOUGHTS, THOUGHT, MIND, WIT; BRAIN, EYES, BODY, BLOOD; NATURE, SUN, EARTH, LIGHT, WORLD; LOVE; COMES.

Interpersonal keywords: O, FIE.

Textual keywords: YET, AND, THUS; I, MY, MYSELF, SHE, HER.

Ideational keywords
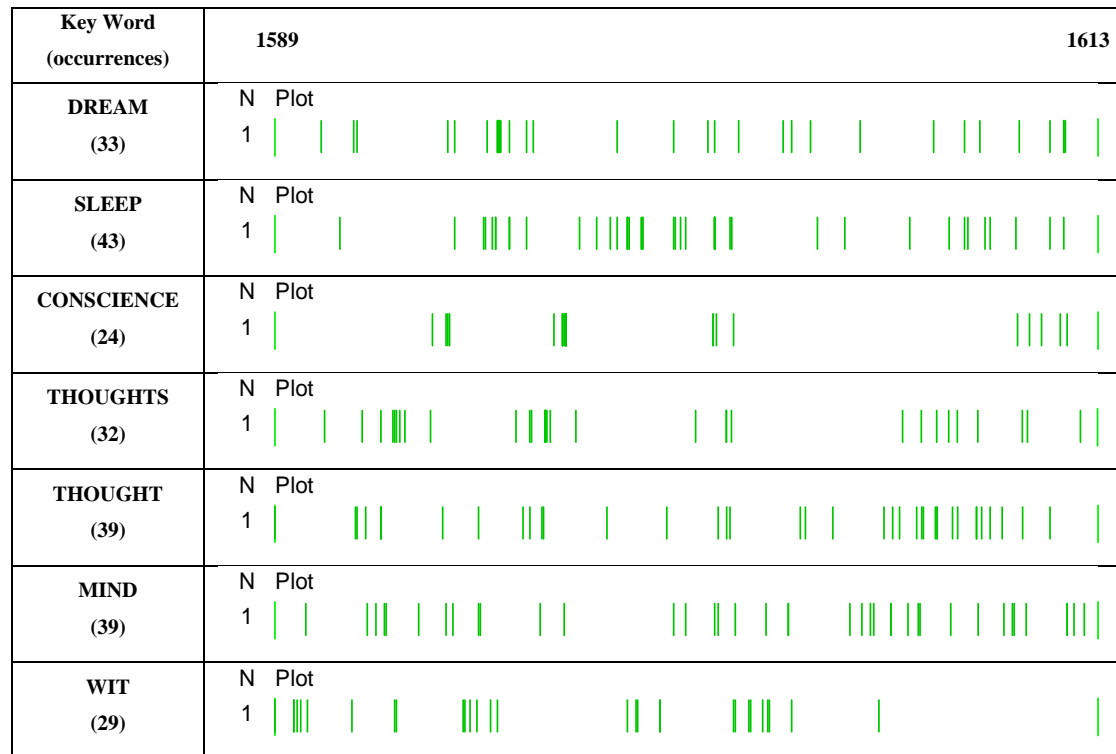
Figure 1. Dispersion plots for words relating to mental states and processes across time.

DREAM, SLEEP, CONSCIENCE, THOUGHTS, THOUGHT, MIND, and WIT form a semantic group of words related to mental states and processes. To discover whether these words occur evenly throughout Shakespeare's soliloquies, I created dispersion plots, illustrations which show the degree to which words are uniformly distributed across the chronologically-arranged corpus (Figure 1). DREAM, SLEEP, THOUGHT and MIND are relatively evenly distributed across the plays, whereas CONSCIENCE, WIT, and to a lesser extent THOUGHTS have more uneven distributions. The 10 comedic instances of CONSCIENCE belong to one speech by Launcelot Gobbo in *The Merchant of Venice*; in total the word only appears in soliloquy in four other plays: *Richard III*, *King John*, *Hamlet*

and *Cymbeline*. WIT does not appear in soliloquy after *King Lear* (1605), as it is more frequent in early to mid-career comedies (19) than tragedies (6) and histories (4). Perhaps the intelligence and sharp-mindedness implied by WIT tends to be more a feature of comedy than other genres. The relatively even distribution of DREAM, SLEEP, THOUGHT(S) and MIND across the canon suggests that these words are stable features, and that soliloquists are portrayed as introspective and preoccupied with cerebral activity.

Soliloquists are also physically preoccupied. The keyness of BRAIN, EYES, BODY and BLOOD attest to this. One might expect BLOOD to be more characteristic of tragedy and history soliloquies because of themes of death and lineage, and EYES to be more typical of comedy, which concerns itself with love, beauty, attraction, etc. The data in Table 2 supports these assumptions. BRAIN is more frequent in tragedy and history than comedy. Collocates emphasise size and weight (*lack-brain*, *too little*, *too much*, *gross*). Hamlet's *book and volume of my brain* (I.v), with its use of hendiadys (two nouns joined by *and* to express a subordinate relationship: 'voluminous book') is a good example. The most frequent collocates of EYES are MINE (14) and HER (13). MINE EYES seems to collocate more with ideas of terror (*pluck out, bloody business, monstrous apparition*), or rejection and destruction (*fools, shun, blasted*), whereas HER EYES shows a preference for expressions of admiration (*jewel-like, feast, blessed*). These tendencies are not unexpected: the phrase MINE EYES usually indicates what the soliloquist perceives visually and hence feels (generally negative); HER EYES frequently refers to what the speaker believes to be the inherent qualities of a loved one's eyes (positive). BODY accounts for 61% of occurrences in history soliloquies, collocating most frequently with MY, often negatively: *Where sits deformity to mock my **body*** (*Henry VI Part 3* III.ii), perhaps emphasising vulnerability. BLOOD occurs frequently in tragedy (47%), and history (33%). It has many senses in Shakespeare, but two predominate:

    i.   blood (n.) passion, feeling, strong emotion [especially sexual]
    ii.   blood (n.) blood-letting, bleeding

<div align="right">Crystal and Crystal (2002)</div>

Distributions of sense i. (tragedy: 3; history: 0; comedy: 8) and sense ii. (tragedy: 13; history: 10; comedy: 1) confirm expectations as to the use of BLOOD in different genres: *But I am in So far in **blood**, that sin will pluck on sin.* (*Richard III* IV.ii).

	Comedy	History	Tragedy	Total
BRAIN	3 (14.3%)	8 (38.1%)	10 (47.6%)	21 (100%)
EYES	30 (42.9%)	14 (20%)	26 (37.1%)	70 (100%)
BODY	5 (17.9%)	17 (60.7%)	6 (21.4%)	28 (100%)
BLOOD	13 (20.3%)	21 (32.8%)	30 (46.9%)	64 (100%)
Totals	51 (27.9%)	60 (32.8%)	72 (39.3%)	183 (100%)

Table 2. Occurrences and distribution of BRAIN, EYES, BODY and BLOOD by genre, showing absolute frequency and relative weighting.

The occurrence and distribution NATURE, EARTH, LIGHT, WORLD and SUN across genres is shown in Table 3. NATURE, EARTH and LIGHT occur more frequently in tragedy. In the case of NATURE, this may be because tragedy often depends on exploitation by evil characters such as Lady Macbeth: *Yet do I fear thy **nature**; / It is too full o'the milk of human kindness* (Macbeth I.v). Nature, in the sense of 'natural law', can equally be exploited by self-seekers such as Edmund: *Thou, **Nature**, art my goddess; to thy law / My services are bound.* (King Lear I.ii). Arresting imagery surrounding EARTH can be found in 13 instances: *Where bloody Tybalt, yet but green in **earth**, Lies festering in his shroud* (*Romeo and Juliet* IV.iii). LIGHT is particularly common in soliloquies in *Othello* (8 instances), with its themes of light and dark, seeing and not seeing. The cumulative effect of the use of NATURE, EARTH and LIGHT in tragic soliloquy seems to emphasise a natural order under threat. With WORLD, most common in historical soliloquy, speakers reflect on struggles for power: *Vain pomp and glory of this **world**, I hate ye.* (*Henry VIII* III.ii). A common theme is the uncertainty and treachery of events: *O **world**, thy slippery turns!* (*Coriolanus* IV.iv). SUN collocates with a similar number of positive and negative concepts in tragedy and history, but with more positive words (*celestial, glorious, beauty*) than negative ones in comedy, as one might expect.

	Comedy	History	Tragedy	Total
NATURE	7 (16.3%)	7 (16.3%)	29 (67.4%)	43 (100%)
EARTH	7 (19.5%)	8 (22.2%)	21 (58.3%)	36 (100%)
LIGHT	11 (33.3%)	5 (15.2%)	17 (51.5%)	33 (100%)
WORLD	11 (17.5%)	29 (46%)	23 (36.5%)	63 (100%)
SUN	11 (35.5%)	8 (25.8%)	12 (38.7%)	31 (100%)
Totals	47 (22.8%)	57 (27.7%)	102 (49.5%)	206 (100%)

Table 3. Occurrences and distribution of NATURE, EARTH, LIGHT, WORLD and SUN in soliloquies by genre, showing absolute frequency and relative weighting.

The keyness of COMES confirms Arnold's (1911) observation that it is a common device to signal a character's approach (see Section 2). In this respect, it is important to note that in Shakespeare's time, play manuscripts were cut into individual 'parts' and speeches pasted onto scrolls, with cues such as 'comes' added for the actors who were about to enter onstage (Thomson, 1997). Lexical patterning (see Figure 2) shows collocates of COMES organised according to frequency. The top word in each column is most frequently found in that position.

N	L3	L2	L1	Centre	R1
1	AND	HERE	WHO	COMES	HERE
2		BUT	HE		
3			SHE		
4			HERE		

Figure 2. Lexical patterning of COMES in soliloquies.

Clusters such as WHO (IS) COMES HERE? (11) and (HERE/WHERE) S/HE COMES (11) serve as conventionalised entrance announcements to signal the end of a soliloquy or to effect thematic change, often by focusing attention on the character entering.

O and FIE can be termed 'pragmatic noise', expressions of strong emotion, with pragmatic and discoursal functions (Culpeper and Kytö, 2010). O seems to function both as an interjection, independently expressing emotion (*O! it is foul.*) and as a particle, focusing attention on what follows (*O heavy burden!*). Table 4 shows collocates of *O* and punctuation.

O + vocative	*O!*	*O,*	*O* + other	Total
131	88	19	4	242
54.1%	36.3%	7.9%	1.7%	100%

Table 4. Collocates and punctuation occurring with O.

The most frequent accompaniment to *O* is a vocative expression (131 instances/54.1%), and the most frequent type of vocative (45 instances) is the rhetorical figure of the apostrophe, which can take one of two forms:

i.   + [adjective(s)] + [personified inanimate object or quality]
ii. + [personal pronoun] + [adjective(s)] + [vocative noun / personified inanimate object or quality]

Common occurrences of type (i) are: **O powerful love**; **O wicked world**; type (ii) instances include: **O thou blessed moon**; **O you leaden messengers**. These are somewhat unusual vocatives in that the addressee is an inanimate quality or object. It is also worth noting the *thou* form used for singular *moon*, and the *you* form for plural *messengers*. The apostrophe was a classical tradition widely used in the Renaissance and liberally exploited by Shakespeare. The reasons for this may be several: first and foremost, Shakespeare may have found the apostrophe to be a useful device for representing the diverse attentions of thought processes; secondly, he may have been keen to demonstrate his training in the classical tradition; thirdly, sophisticated members of the audience may have expected such deference to the classics.

There are 7 instances of what Culpeper and Kytö refer to as 'self-reflexive vocatives', that is to say vocative expressions in which addresser and addressee are the same person. Examples include *O me!* and *O coward conscience!*. Having the addresser and addressee as the same person can be accounted for in a split-self paradigm (see Emmott, 2002). In *O coward conscience!*, for example, we have a 'self as judge' and a '(part of) self as adjudged'. Address to other characters by name (*O Caesar!*, *O sweet Maria*) is common (21 instances), as is address to supernatural forces (*O God!, O heavens!*) (20 instances). Culpeper and Kytö term these 'exclamatory vocatives'.

The form *O!* as a strong interjection appears 88 times and is frequently followed (16 times) by a hypothetical form (*O! that I had my wish*, *O! would thou were*), a means by which a character reveals how her wish-world differs from the real world of the play[8]. Another common pattern (11 instances) is *O!* + imperative (*O! be some other name*), signalling perhaps a desire on the part of the character to control events.

With regard to FIE, Culpeper and Kytö (2010) point out its original use as the sound made on perceiving a disagreeable smell[9]. In Early Modern English, FIE was often used to pour scorn on something or someone (Taavitsainen, 1995). Jack Cade does both: *Fie on ambition! fie on myself.*(*Henry VI Part 2* IV.x). It is sometimes repeated as an expression of exasperation: *I must be round with him, now he comes from hunting. Fie, fie, fie, fie!* (*Timon of Athens* II.ii).

Textual keywords

To prove you a cipher. To be whipped;	and yet	a better love than my master.
God pardon him! I do, with all my heart;	And yet	no man like he doth grieve my
virtues excellent, None but for some,	and yet	all different. O! mickle is the
many time this two-and-twenty years,	and yet	I am bewitched with the rogue's
me before to Romeo? I fear it is:	and yet,	methinks, it should not, For he
I have no great devotion to the deed;	And yet	he has given me satisfying
thou would holily; would not play false,	And yet	would wrongly win; thou'dst
folio. And three times as much more,	and yet	nothing at all. A message well
lock and key of villainous secrets;	And yet	she'll kneel and pray; I have
let me clutch thee: I have thee not,	and yet	I see thee still. Art thou not,

Figure 3. A sample of the contrastive cluster AND YET.

Linking adverbials such as YET and THUS help the reader/audience make sense of relations between different parts of the text (Biber, *et al.*1999). YET more than any other word, characterises Shakespearean soliloquies. The *OED* distinguishes three basic senses of *yet*[10]. By far the most common (125 instances / 68%), according to the *OED* definition, is to 'introduc[e] an additional fact or circumstance which is adverse to, or the contrary of what would naturally be expected from, that just mentioned' (*OED* 'yet' III.9). At one extreme, we can find Richard III contradicting himself, while at the same time indicating a highly perturbed mental state: *I am a villain.* **Yet *I lie; I am not.*** (*Richard III* V.iii). More generally, however, *yet* signals doubt and hesitancy, particularly when used with *and*, its most frequent collocate. This can be seen clearly in the sample shown in Figure 3. AND YET accounts for 31% of all contrastive uses of YET. All but one are clause initial, the preceding punctuation indicative of a pause in speech, thereby creating the fiction that thought is taking place (Herman, 1995). AND YET can also serve as a fulcrum for antithetical parallelism: *would not play false,* ***And yet*** *would wrongly win* (*Macbeth*, I.v). Such examples demonstrate how Shakespeare used repetition, synonyms and semantically related words either side of AND YET to highlight inherent contradictions about characters and events, and perhaps to demonstrate poetic skill.

Approximately 90% of the 74 instances of THUS in the corpus mean 'In this way', as in *It is the bloody business which informs /* ***Thus*** *to mine eyes.* (*Macbeth* II.i), and around 10% mean 'accordingly, consequently'[11], as in: *Thy drugs are quick.* ***Thus*** *with a kiss I die.* (*Romeo and Juliet* V.iii). The *Macbeth* example is listed as a sub-sense, *thus* meaning *in the manner now being indicated or exemplified.* This is important for two reasons: first, because of the verisimilitude it lends to the situation, namely that Macbeth actually sees the dagger with its *gouts of blood*; and secondly, because of the information it provides as to the character's thought processes. We can gain much insight from the logical deductions speakers make. When Hamlet observes

Thus conscience does make cowards of us all,
And thus the native hue of resolution
Is sicklied o'er with the pale cast of thought

Hamlet III.i

he is not only making a general observation about the human condition, but obliquely revealing the reason for his own inaction. It may be that the first sense of *thus* predominates because for soliloquy to work, the playwright must make the character 'think through' a situation, thereby providing psychological depth and motive.

Deciding what counts as a textual word is not always clear-cut. This is particularly true in the case of socially deictic pronouns, words which make reference to the participants in soliloquy events. The second person pronouns 'thou' or 'you', for example, are interpersonal, though of course the addressee is rarely present. In dialogue, 'I' might be considered interpersonal, but in the specific case of soliloquy is not so much *inter*personal as *intra*personal. Third person pronouns such as 'she' and 'her' are often anaphoric and thus clearly textual. For the sake of simplicity, I decided to group all personal pronouns as textual.

Perhaps the least surprising finding of this research is that I is a textual key word. 1,600 instances of I (1,801 including I'LL) attest to the notion that soliloquy tends to convey a sense of egocentricism. I examined the corpus for immediate collocates of I, and three-word clusters in which I occurs. The results of this examination can be seen in Table 5, Columns A and B.

Column A: immediate collocates of *I*						Column B: three-word clusters with *I*		
N	Word	Total	L1	Centre	R1	N	Cluster	Freq.
1	I	1,602	1	1,600	1	1	I WILL BE / I'LL BE	37
2	'LL/WILL/SHALL	419	55	0	364	2	I KNOW NOT	13
3	AM	184	33	0	151	3	I WILL NOT	13
4	HAVE	152	26	0	126	4	BUT I WILL	11

Table 5. Collocates of *I* one place to the left (L1) or right (R1) and three-word clusters with *I* in soliloquy.

Column A shows that the most frequent collocates of I are, by a long way, the forms WILL, 'LL and SHALL[12]. This finding supports the idea that one of the purposes of soliloquy is to prime readers/audiences for subsequent plot developments as well as providing an insight into characters' plans. The WILL and SHALL auxiliaries were 'gaining ground in the first person in the Early Modern period' (Nevalainen, 2006: 95). However, auxiliary

'will' derives from a main verb meaning 'wish' or 'want', and 'shall' indicating obligation or necessity (Blake, 2002). This led to some overlap as 'English was trying to express volition, obligation and futurity through only two auxiliary verbs; some confusion of functions was inevitable' (Blake, 2002: 125). Column B shows that the following clusters: I WILL BE / I'LL BE and I WILL NOT are particularly prevalent: *Therefore,* **I will be** *sudden and dispatch* (*King John* IV.i); **I'll have** *her but* **I will not** *keep her long* (*Richard III* I.ii).

The preponderance of so many of these forms would seem to indicate that one of the ways in which soliloquy works is to make the reader/audience aware of the speaker's resolve, be it volition, obligation or future intention, thus characterising her, and creating expectations as to plot development. This determination may be expressed by inverting pronoun and future auxiliary after adverbs such as *now* for emphasis: *Now* **will I** *[go] to the Goths* (*Titus Andronicus* III.i). Note also the ellipsis of a verb of motion after auxiliary 'will'.

Unsurprisingly, AM is a frequent collocate of I (184 instances). Common three-word clusters include I AM A and I AM NOT: **I am a** *villain* (*Much Ado About Nothing* II.iii); **I am not** *naturally honest* (*Winter's Tale* IV.iv). Such clusters show characters revealing their state of self-knowledge, allowing readers/audience to assess its veracity.

Negative keywords

Negative keywords (see Table 6) are significantly infrequent in the corpus and can also reveal the stylistic nature of soliloquies by showing us what is not present.. Negative ideational keywords like SIR, GOOD, LADY, BROTHER, and PEACE occur in clusters such as PRAY YOU SIR, MY GOOD LORD, THE LADY, MY BROTHER and PEACE BE WITH YOU, suggesting interpersonal address or relationships. Other keyword clusters emphasise the act of speaking to or directing another: I TELL THEE, SPEAK WITH YOU, GO WITH ME and GIVE ME YOUR. These directives emphasise the transactional nature of a talk exchange between two or more participants, thus accounting for the unusual infrequency of these items in soliloquies.

Negative key word	Frequency in soliloquy	Frequency in dialogue	Keyness
GIVE	69	1290	-7,1
PEACE	21	524	-8,9
BROTHER	20	519	-9,7
LADY	23	607	-11,9
SPEAK	40	1136	-26,3
GO	62	1695	-36,1
TELL	27	1051	-41,1
GOOD	107	2787	-52,9
LORD	49	2669	-150,0
SIR	22	2524	-222,4
YOUR	93	6,663	-461,86
YOU	179	13,676	-992,82

Table 6. Negative keywords in soliloquy.

Generic differences

To test for significant variation among comedy, history and tragedy soliloquies, I divided my soliloquy corpus into sub-corpora, based on the First Folio categorisation. Comparison of each genre with those of the other two gave the results shown in Table 7.

COMEDY		HISTORY		TRAGEDY	
Key word	Keyness	Key word	Keyness	Key word	Keyness
LOVE	59.09	HENRY	54.94	T	51.08
I	44.91	KING	23.55	GODS	26.93
SHE	35.62	MANY	17.31	O	26.61
HER	33.69	AND	13.60	THOU	26.15
THY	-11.20	LOVE	-19.44	LOVE	-17.88
THOU	-15.71	HER	-23.25	I	-42.79

Table 7. Main positive and negative keywords in comedic, historic and tragic soliloquy.

Comedy

1.	I [hunt] after love	9.	love is still most precious in itself
2.	Love bids me forswear	10.	Love, lend me patience
3.	Love is a familiar; Love is a devil	11.	Alas, how love can trifle with itself!
4.	The boy Love is perjured	12.	O love! be moderate
5.	Love looks not with the eyes	13.	omnipotent love!
6.	If I can check my erring love	14.	O powerful love!
7.	Love, set on thy horns	15.	Love, lend me wings
8.	how wayward is this foolish love		

Table 8. Metaphorical uses of love in comic soliloquies

LOVE strongly defines comedic soliloquy. It is often used metaphorically, as can be seen in Table 8. In particular, certain metaphors can be noted: LOVE IS AN ANIMAL (1, 7), and LOVE IS AN ANIMATE ENTITY (for example, 3, 4, 5), who can be wayward (6, 8), powerful (13, 14), lacking good sense (8, 11), able to make and grant requests (2, 10, 12, 15), and more besides. Above all, LOVE is frequently personified in comic soliloquy, particularly implicitly as the goddess, Venus or the god, Cupid (15)[13].

The positive keyness of LOVE, I, SHE and HER contrasts with the negative keyness of THY and THOU, not only confirming intuitive perceptions of the archetypal comic soliloquist as an introspective lover, but reminding us that second-person pronouns are more likely to be used in interactional contexts, at least in comedy (but, see the section on Tragedy below).

History

The keyness of HENRY and KING (especially THE KING) in history is highly predictable, not only because 7 of the 10 plays concern a king called Henry, but also because the comedies and tragedies contain no such-named character. Other key names not shown in the table are YORK, EDWARD and RICHARD, thus emphasising the identity, office and relevance of certain male characters. The negative keyness of LOVE and HER conforms to expectation as Shakespeare's histories concern power struggles among men, while women 'stand for permanence and fidelity against shifting political sands' but are essentially 'impotent' (Dusinberre, 1975: 294; 297).

A word may be key because of repeated use by one character, thus giving the misleading impression that it is characteristic of the genre as a whole.

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;

Henry VI, Part 3 II.v

Whether or not such obvious parallelism is indicative of a certain 'early primitiveness' (Skiffington, 1985) in soliloquy design, MANY serves as a useful reminder that corpus findings should always be treated with caution and subjected to qualitative analysis.

Tragedy

Perhaps the most curious finding of all in my comparison of soliloquy between genres is the keyness of T, the contracted form of IT, with the 'I' elided. This occurs 34 times in tragedy, particularly in the forms DO'T (6) and ON'T (5): *If swift thought break it not, a swifter mean Shall outstrike thought; but thought will do***'t***, I feel* (*Antony and Cleopatra* IV.vi); *Best draw my sword; and if mine enemy But fear the sword like me, he'll scarcely look on***'t** (*Cymbeline* III.vi). Clearly, use of the contraction 'T has prosodic advantages, by eliminating an unwanted extra syllable. A prosaic explanation may be that it is simply a 'space-saver' (Crystal, 2008: 53), a means by which a compositor ensured that a line of text did not exceed the column measure, though that would not account for its keyness in tragedy. One explanation may be that the playwright intended it to represent the speech or thought style[14] of a character whose emotions are affected by tragic circumstances such as destitution or infidelity: *O gods! Who is***'t** *can say, "I am the worst?"* (*King Lear* IV.i); *He has done my office: I know not if***'t** *be true* (*Othello* I.iii). I have already shown that O is a key word in soliloquies. This is especially true of tragedy, and confirms Clemen's (1987) assertion that apostrophes particularly suggest the passionate style of the tragic soliloquies. The most common forms of vocative address are to named characters, e.g. O Brutus! (13) and GOD or GODS (9). With regard to the keyness of THOU[15] (and THY and THEE) in tragic soliloquy, it is important to note that often, the referent can be another character or an inanimate object. In *Othello*, Shakespeare cleverly plays with both:

If I quench **thee**, **thou** flaming minister [candle],

I can again **thy** former light restore

Should I repent me: but once put out **thy** light [Desdemona's life],

Thou cunning'st pattern of excelling nature,

I know not where is that Promethean heat

That can **thy** light relume.

Othello V.ii

Note the contrast here with comic soliloquy. While comic soliloquists do not generally use second-person pronouns, tragic soliloquists do. My contention is that tragedy often deals with forces seemingly beyond the characters' control (though it is their human failings which lead to the tragedy). By using forms which address supernatural, absent or inanimate referents, Shakespeare suggests that characters seek to communicate with the wider universe in an attempt to explain or justify their feelings and actions.

4 Conclusions

In this article, I have used corpus stylistic techniques to identify characteristic language forms in Shakespeare's soliloquies. I have not only been able to confirm claims made by literary critics as to the distinctiveness of soliloquies from dialogue and the prevalence of certain forms in such passages, but importantly, to quantify what these forms are and to offer a stylistic analysis of their significance. The keyness in soliloquies of forms such as 'thought', 'sleep', 'comes', 'I' and apostrophes confirms the claims of critics like Arnold (1911), Clemen (1987) and Hirsh (2003). However, this study goes much further, by offering a stylistic analysis of these forms in context. It also identifies other key forms which critics have failed to notice. In this respect, I have shown that 'thought and 'sleep' belong to a set of key words in soliloquies related to mental states and processes. These include 'dream', 'conscience' and 'mind'. Other key semantic sets are related to the body ('brain', eyes', 'body', 'blood') and nature ('sun', 'earth', 'light', 'world'), as well as 'love', which is, not surprisingly, key in comedy. Further interesting findings include the keyness of pragmatic noise features ('O', 'fie'), linking adverbials ('yet', 'and', 'thus') and certain pronouns ('my',

'myself', 'she', 'her'). This study also breaks new ground by systematically comparing soliloquies across the genres of comedy, history and tragedy and identifying those language forms characteristic of each. The most surprising finding in this regard is the keyness in tragedy of 't`' ('it' with the elision of the 'i') in conjunction with a verb such as 'done', perhaps indicating heightened emotion under duress.

Apart from making a valuable contribution to Shakespeare studies, by applying corpus stylistic techniques to such canonical texts, this research underscores the value of literary corpus stylistics in confirming claims made by literary critics, and in bringing hitherto unrecognised key linguistic features to light, thereby challenging both critics and stylisticians to interpret their significance so as to enhance our appreciation of literary texts.

References

Adamson S, Hunter L, Magnusson L, Thompson A and Wales K (eds) (2001) *Reading Shakespeare's Dramatic Language: A Guide*. London: Thomson Learning.

Anthony L (2014) AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Arnold ML (1911) *The Soliloquies of Shakespeare: A Study in Technic*. New York: The Columbia University Press.

Biber D, Johansson S, Leech G, Conrad S and Finegan E (1999) *Longman Grammar of Spoken and Written English*. London: Longman.

Blake NF (2002) *A Grammar of Shakespeare's Language*. Basingstoke and London: Macmillan.

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

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

Crystal D and Crystal B (2002) *Shakespeare's Words: A Glossary and Language Companion*. London: Penguin.

Crystal D (2008) *Think on my Words: Exploring Shakespeare's Language*. Cambridge: Cambridge University Press.

Culpeper J (2009) Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14(1): 29-59.

Culpeper J and Kytö M (2010) *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: Cambridge University Press.

Dunning T (1993) Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1): 61-74.

Dusinberre J (1975) *Shakespeare and the Nature of Women*. London and Basingstoke: Macmillan.

Emmott C (2002) 'Split selves' in fiction and in medical 'life stories': Cognitive linguistic theory and narrative practice. In: Semino E and Culpeper J (eds). *Cognitive Stylistics*. Philadelphia and Amsterdam: John Benjamins, pp. 153-182.

Fischer-Starcke B (2009) Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics* 14(4): 492-523.

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

Halliday MAK (1994) *An Introduction to Functional Grammar* (2nd edn). London: Edward Arnold.

Herman V (1995) *Dramatic Discourse: Dialogue as Interaction in Plays*. London: Routledge.

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

Hussey SS (1992) *The Literary Language of Shakespeare*. London: Longman.

Mahlberg M (2013) *Corpus Stylistics and Dickens's fiction*. New York: Routledge.

Nevalainen T (2006). *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.

OED = Oxford English Dictionary Online (2nd edn). Oxford: Oxford University Press. http://www.oed.com.ezproxy.lancs.ac.uk/ (accessed via Lancaster University library).

Scott M (2008) *WordSmith Tools* (Version 5.0). Liverpool: Lexical Analysis Software.

Semino E (1997) *Language and World Creation in Poems and Other Texts*. Harlow: Longman.

Semino E and Short M (2004) *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.

Sinclair J (1996) The search for units of meaning. *Textus* 9: 75-106.

Shakespeare W and Craig WJ (1916) *The Complete Works of William Shakespeare*. Oxford: Clarendon Press.

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

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 (1995) Interjections in Early Modern English: from imitation of spoken to conventions of written language. In: Jucker AH (ed.). *Historical Pragmatics:*

Pragmatic Developments in the History of English, (Pragmatics and Beyond New Series 3), 439-465. Amsterdam & Philadelphia: John Benjamins.

Thomson P (1997) 'Rogues and rhetoricians: Acting styles in early English drama'. In: Cox JD and Kastan DS (eds) *A New History of Early English Drama*. New York: Columbia University Press, pp. 321-335.

Tissari H (2006) 'Is love a tender thing?' Metaphors of the word 'love' in Shakespeare's plays. *Studi Linguistici e Filogloici Online,* 4(1): 131-174.

Williams R (1983) On dramatic dialogue and monologue (particularly in Shakespeare). In: *Writing in Society*, 31-64. London: Verso.

Notes

[1] By 'language forms', I refer not only to words (which I define as a letter or a string of letters with a blank space or a punctuation mark before and after it/them), but also to elements of language such as contracted word forms (for instance, 'I'm'), collocations (strong associations between words, for example, 'true love'), rhetorical expressions (such as, 'O!'), metaphors (for example, 'Love is blind'), and so on.

[2] The 37 plays include all 36 plays in the 1623 First Folio, plus *Pericles*. The source text I used is the Shakespeare Corpus developed by Mike Scott, based on the OUP edition of 1916. I chose this version as it was ready for corpus processing and spelling variations (which can represent a serious problem for corpus analysis based on frequency counts) had been regularised.

[3] Prose soliloquies are typical of secondary or lower-status characters (though Falstaff is a notable exception). Soliloquies in verse are generally given to major or higher-ranking characters.

[4] Though comic, Ford's jealousy and suspicion of his wife prefigure the more tragic and serious cases of Othello, and Leontes (*The Winter's Tale*).

[5] Collocations have been defined as words with a statistical tendency to co-occur (Hunston, 2002).

[6] Also commonly referred to as n-grams, clusters, chains, idioms, clichés, conversational routines, recurrent word sequences and multi-word units. Researchers use terms differently. For example, in the AntConc corpus toolkit (Anthony, 2014), clusters are based on a search condition, whereas n-grams are clusters of a specified length. See Moon (1998) and Culpeper and Kytö (2010) for further discussion of these terms.

[7] 'Tokens' refers to the total number of individual words; 'types' refers to the total number of distinct words (taking account of words which are repeated).

[8] See Semino (1997) for a discussion of possible world theory.

[9] "fie, int." OED Online. Oxford University Press, March 2014. (accessed 16.04.2014).

[10] "yet, adv. (and adj.) and conj." OED Online. Oxford University Press, March 2014. (accessed 16.04.2014).

[11] "thus, adv." *OED Online*. Oxford University Press (accessed 16.04.2014).

[12] In table 2A, I combined the results for WILL, 'LL and SHALL to highlight the salience of these similar forms as a feature of self-talk.

[13] See Tissari (2006) for a discussion of metaphors involving the personification of love in Cupid.

[14] See Semino and Short (2004) for futher discussion of speech and thought representation.

[15] The overuse of THOU means that the archaic form of 'are', ART, is key as well.