

Depictions of deception: a corpus-based analysis of five Shakespearean characters

Dawn Archer

Manchester Metropolitan University and Lancaster University, UK

Mathew Gillings

Vienna University of Economics and Business, Austria

Corresponding author: Dawn Archer, Manchester Metropolitan University, 436, Geoffrey Manton Building, Manchester Campus, Rosamond St West, Manchester, M15 6EB, UK
Email: d.archer@mmu.ac.uk

Abstract

Drawing on the Enhanced Shakespearean Corpus: First Folio Plus and using corpus-based methods, this paper explores, quantitatively and qualitatively, Shakespeare's depictions of five deceptive characters (Aaron, Tamora, Iago, Lady Macbeth, and Falstaff). Our analysis adopts three strands:

- Firstly, statistical keywords relating to each character are examined, to determine what this tells us about their natures more generally.
- Secondly, the wordlists produced for each of the five characters are drawn upon to determine the extent to which they make use of linguistic features that have been correlated with, or linked to, acts of deliberate deception in real-world contexts.
- Thirdly, we make use of the results identified during the two aforementioned strands by using them to identify particular (sequences of) turns that are worthy of more detailed analysis.

Here, we are primarily interested in (a) whether these keywords/deceptive indicators cluster or co-occur; and (b) whether these interactions are the same as those identified by other scholars exploring depictions of deception in Shakespeare from a literary perspective.

The findings indicate that deception-related features are indeed used collectively/in close proximity, by Shakespeare, at points where a character speaks to other characters disingenuously. They also suggest that Shakespeare's deceptive depictions do change stylistically, from character to character, in line with those characters' different characterisations and situations; that Shakespeare draws on atypical language features - such as self-oriented references - when it comes to some of his depictions of deception; and that Shakespeare uses these various stylistic features to achieve a range of dramatic effect(s).

Keywords

corpus linguistics, cues, deception, Shakespeare, stylistics

1. Introduction

This paper draws upon the Enhanced Shakespearean Corpus: First Folio Plus (ESC: Folio) (Culpeper et al., *in prep.*), a resource created as part of the AHRC-funded Encyclopaedia of Shakespeare's Language project (Grant Ref AH/N0024515/1). In combination with corpus-based methods, we explore, quantitatively and qualitatively, Shakespeare's depictions of three deceptive male characters and two deceptive female characters. Namely:

- Aaron and Tamora, from *Titus Andronicus*;
- Iago, from *Othello, the Moor of Venice*;
- Lady Macbeth, from *The Tragedy of Macbeth*;
- Falstaff, from *The First Part of King Henry IV*.¹

In line with Vrij's (2000: 6) definition of deception as 'a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue', our primary focus will be upon these characters' use of *deliberate* deception with the aim of intentionally misleading other characters. This means that the following types of deception are omitted from our analysis:

- accidental misunderstandings
- self-deception (i.e., characters believing something that is not true) and/or
- characters' use of disguises - with the exception of Tamora's portrayal as Revenge when confronting Titus (see 3.2).

By investigating 'depictions of deception', we are primarily interested in ascertaining how Shakespeare uses language in order to portray the five characters' duplicity. This might be at the point they use language duplicitously in an attempt to deceive other characters and/or at the point they admit their duplicity to the audience (but not to fellow characters). We also want to ascertain whether/the extent to which the five depictions show similarities with one another as well as with language features that are believed to point to deception today (for a description of such language features, see Section 4). Our quantitative analyses of the five characters thus adopt two strands, as outlined below:

- The first strand involves examining statistical keywords relating to each character to determine what this might tell us about their deceptive or manipulative natures on a general level (see Sections 3-3.6).

By statistical keywords, we mean words that occur in a text more frequently than we would otherwise expect they might by chance alone (see Demmen and Culpeper, 2015 for further information on the keyword approach). Our keywords have been identified by comparing the

wordlist of each character against the wordlist of the play s/he appears in (after omitting their turns from that play), using the three-step filtering process explained in Section 3. The keyword approach has already been used extensively to explore ‘aboutness’ (Phillips, 1989) and characterisation - including in a Shakespearean context (Culpeper, 2002, 2009; Archer and Bousfield, 2010). However, with the exception of, for example, Archer and Lansley (2015) and McQuaid et al. (2015), there has been little usage of the keyword approach in order to find ‘potential cues as to whether...someone is lying’ (Archer and Lansley, 2015: 245). This aspect of our work (see especially Sections 3-3.6) therefore seeks to extend this type of research, whilst also adding - what we would argue is - an important second strand to such investigations.

- The second strand involves drawing on the wordlists produced for each character again in order to determine the extent to which they make use of language features that have been found to occur when deliberate verbal deception takes place in real-world contexts (see Sections 4-4.2).

We should highlight, at this point, that there is no ‘Pinocchio’s nose’ when it comes to detecting deception. That is to say, there is no single verbal or behavioural feature that points towards deception each and every time. However, certain features (be they verbal or behavioural) have been found to be associated with (attempts at) deception - especially when they cluster (Archer and Lansley, 2015) within a short timeframe (Houston et al., 2012: 30) or stretch of text (Picornell, 2013). We should highlight, in turn, that much of the research into present-day deception detection starts from the premise that deceivers undergo three main processes:

- 1) they have an emotional reaction;
- 2) they experience an increase in cognitive load; and
- 3) they attempt to control their behaviour (Zuckerman et al., 1981).

Deception (detection) researchers have therefore looked for ways in which these three processes manifest themselves through language; DePaulo et al.’s (2003) meta-analysis reported on 158 cues, for example, and dozens more have been identified since then (see Section 4 for a discussion of specific linguistic manifestations of potential deception).

Even though much of the research within deception detection has been well-supported by years of evidence, it is also worth highlighting that the majority of studies have tended to focus on searching, at face value, for - what are often treated as “context-neutral” - *theory-led* features. This means that they do not fully take into account how such features have been operationalised by previous work, how speakers use them (including for what purpose), and/or how they differ

sociolinguistically (Gillings, 2020). Not surprisingly, recent work within the field has suggested that, when it comes to deception detection in particular, the predictive abilities of some features is likely to have been over-inflated. Luke (2019), for example, criticises researchers for not accounting – and hence adjusting some of their findings – for small participant numbers, publication bias and/or power differentials. In light of these criticisms, and in light of similar work within stylistics (e.g. Sorlin, 2016), we propose a new approach to investigating deception - using Shakespearean texts as a pilot - which not only adopts the two aforementioned quantitative strands but also engages in further qualitative analyses.

- The third strand is a form of qualitative analysis which, due to length constraints, will focus on two of the five characters (see Sections 5-5.2). Our interest, here, relates to whether and, if so how, Shakespeare's depictions of deception change stylistically, from character to character, and from context to context. In addition, we are interested in whether/the extent to which they draw upon the indicators of deception that have been identified in real-life deceivers, and whether/the extent to which the characters (like real-life deceivers) 'adapt to the changing landscape, altering and restructuring their communication blueprint as they encounter obstacles' (Picornell, 2013: 33).

Given our interest in characters' admissions of duplicity to the audience, but not to other characters (see above), we also highlight co-occurring features that would not happen in real life: that is, the playwright's use of soliloquies and asides (as a means of keeping audiences, but not other characters, *in the know*).

2. Background

This background section provides more details of the resource used in this study (2.1) and the keyword approach adopted (2.2).

2.1. *The Enhanced Shakespearean Corpus: First Folio Plus (ESC: Folio)*

The ESC: Folio is one of five corpora compiled as part of the Encyclopedia of Shakespeare's Language project², a five-year undertaking with the overall aim of releasing a series of corpus-based encyclopaedias exploring how Shakespeare's language would have been understood by his contemporaries. They include a frequency-based dictionary of words used within the works of Shakespeare and thematic explorations of his plays and characters. The ESC: Folio consists of what is widely considered to be Shakespeare's canon: the 36 plays published within the First Folio in 1623, plus the inclusion of *The Two Noble Kinsmen* and *Pericles*. The ESC: Folio is distinct

from other Shakespeare corpora in that it is not only marked-up with grammatical and semantic annotation, but individual characters have also been manually annotated with information about their gender and social rank, allowing much more sophisticated analyses to take place. In addition, the play-texts have undergone spelling regularisation in order to aid computer-assisted analysis, although there are options to view alternative spellings within the interface (see Archer and Findlay, forthcoming). The ESC: Folio, along with the ESC: Comparative Plays (Demmen, 2020) and ESC: EEBO-TCP Segment (Murphy, 2019) can all be accessed via CQPweb. Their addition to CQPweb means that the full range of corpus-based methods - collocation, keyword, and concordance analysis - can be carried out on them. More information about the design of the ESC: Folio can be found in Culpeper et al. (*in prep.*), and details of how to access all the resources can be found on the project website.

2.2. The keyword approach

Keyword analysis highlights which words appear statistically more frequently (or infrequently) in one corpus over another. It is a tried-and-tested method within corpus stylistics, having been applied to fictional work such as romantic fiction (Tribble, 2000), Jane Austen's novels (Fischer-Starcke, 2009) and Shakespeare's plays (Culpeper, 2002, 2009; Archer and Bousfield, 2010). The 'important point of interest' for such researchers is not simply 'undertak[ing] statistical comparisons between incidences of the same word' across two texts or corpora 'in order to establish significant differences' (Culpeper and Demmen, 2015: 92). Rather, these researchers are interested in what keywords can tell them about *style* more generally, as well as characters' emotional and cognitive experiences, as depicted by an author or playwright, more specifically (cf. Archer and Lansley, 2015: 245; see also Section 3.6). Applied to our particular study, one would therefore expect that *if* there is a stylistic difference between the language of characters that are speaking duplicitously and those that speak 'plainly' (see, e.g. Graham, 1991; Maley and Tudeau-Clayton, 2010; Grennan, 1987; Cook, 1998; and also Section 4), this is likely to appear as one or more keywords.

Of course, any study making use of keywords also requires a close reading of any relevant concordance lines. In our case, for example, we want to determine what, if anything, the keywords of five Shakespearean characters, when viewed in context, can tell us about the respective characters' act(s) of duplicity (see Sections 3-3.6). Given this initial step constitutes an investigation of how deception works stylistically, it is not guaranteed that the keyword method will also instantly point us towards any of the theory-led cues to deception. However, we do highlight where this occurs (at various points in Sections 4-5).

3. Keyword results for the five characters

The keyword analyses reported in Sections 3-3.5 have been carried out using the new CQPweb interface (see 2.1), and (as outlined in the Introduction) involved comparing the wordlists of the character under analysis against the wordlist from the rest of the play with that character's turns omitted. Table 1, below, captures the statistical keywords for each of the five characters under investigation: Aaron and Tamora from *Titus Andronicus*; Iago from *Othello, the Moor of Venice*; Lady Macbeth from *The Tragedy of Macbeth*; and Falstaff from *The First Part of King Henry IV*. The keywords have been identified using a three-step filtering process: namely, two statistical measures - Log-Likelihood (LL) and Log Ratio (LR) – in combination with a third step of only including keyword results that occur a minimum of five times. LL measures *statistical significance*, that is, how much evidence we have for a difference between two wordlists. The results presented here are filtered using a LL cut-off of 6.63, meaning each keyword has a minimum confidence level of 99%. The second measure, LR, is an *effect size* statistic used to sort the keyword list such that the quantitatively largest differences are at the top of that list (<http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/>). Because LR is a binary logarithm of the ratio of relative frequencies, each increase by 1 indicates a *doubling* of how many times *more* frequent the word is, with respect to a particular character's wordlist, when compared against the *full play* wordlist minus that character's turns. In order to build the analysis on only the most prominent differences among those shown to have a sufficient evidence base (using the LL filter), this paper reports on keywords with an LR ranging between 1 and seven (making them twice to one-hundred-and-thirty-two times more frequent in the relevant character's wordlist, relatively speaking). We have restricted the analysis to keywords with a minimum frequency of five, on the understanding that it is difficult to say, with any kind of confidence, how a word is used in a character's speech when it occurs less frequently than this (it is difficult to make claims about style or 'aboutness', for example). Whilst we acknowledge that low-frequency words may be semantically related and therefore contribute to a style, this relatively liberal cut-off point allows us to identify keywords that – because of combining a significance and effect size with a minimum frequency filter – are both *quantitatively* and *qualitatively* robust (see also Archer and Findlay, forthcoming).

As Table 1 reveals, this procedure attracts large variation in the number of keywords (i.e., between four and twenty-three per character):

Table 1: Characters' key words (followed by 'frequency/LL/LR' values)

Character	Key words (ordered by LR value)
Aaron	gold (8/31.28/6.6), villainy (5/14.44/4.92), black (6/12.91/3.6), keep (7/11.25/2.82), set (5/7.21/2.6), Empress (14/10.34/1.65)
Tamora	ear (5/8.2/2.7), fear (5/7.47/2.51), Andronicus (15/17.21/2.09), sweet (11/9.88/1.79), revenge (9/7.96/1.77), at (13/8.41/1.46)
Iago	mark (8/18.09/5.07), sleep (7/10.58/3.88), wine (6/8.6/3.65), angry (5/6.68/3.39), awhile (5/6.68/3.39), thyself (5/6.68/3.39), drink (7/7.85/2.88), follow (8/7.54/2.48), purse (8/7.54/2.48), money (10/8.98/2.39), lieutenant (17/8.46/1.57), Roderigo (16/7.3/1.48), which (25/8.73/1.25)
Lady Macbeth	without (6/8.76/2.52), bed (5/7.01/2.45), would (15/8.69/1.38), your (27/11.25/1.14)
Falstaff	Hal (35/88.52/5.87), banish (8/23.53/5.74), fifty (6/17.64/5.33), wag (5/14.7/5.07), reward (5/14.7/5.07), Poins (9/20.49/4.91), counterfeit (10/19.64/4.07), plague (14/26.9/3.97), company (11/19.37/3.62), tavern (5/7.37/3.07), pick (7/9.94/2.97), rogue (13/17.67/2.86), Bardolph (6/7.75/2.74), cup (6/7.75/2.74), coward (12/15.99/2.64), lad (8/7.54/2.16), sack (12/11.31/2.16), hang (11/7.6/1.74), old (14/8.84/1.64), honour (12/7.46/1.63), if (57/33.86/1.58), give (27/15.49/1.54), out (27/7.53/1.27)

What should be immediately apparent to readers is that only two of the above keywords (i.e., **villainy** and **counterfeit**) have obvious connections with deception and manipulation at first glance - and for Aaron and Falstaff only. Only one of the two, moreover, has a denotative meaning similar to deception. That words denotatively associated with deception appear as keywords at all is probably due to the fictional nature of the data: as when, for example, a character confides in their audience during a soliloquy or aside (see, e.g., Sections 3.1 and 3.5). In line with their real-life counterparts, however, the five characters also use keywords that seem to have little to do with deception and manipulation at the outset but sometimes point to deception, on their part, when considered in their context-of-use. In what follows, we focus on those keywords that link with deception and manipulation, paying particular attention to points where the five characters use language duplicitously in an attempt to deceive other characters and/or admit their duplicity to the audience (as outlined in the Introduction). This means that some of the keywords in Table 1 will not be discussed, even though they might provide important 'aboutness' information and/or important characterisation information beyond the five characters' duplicitous natures.

3.1 Aaron

Two of Aaron's keywords - **black** and **villainy** – hint at the dark-skinned Moor's propensity for iniquity. He does not merely choose wickedness over goodness, for example, but appears to delight in it. After tricking Titus into chopping off his hand he tells the audience: 'how this *villainy* / Doth fat me with the very thoughts of it, / Let fools do good, and fair men call for grace, / Aaron will have his soul *black* like his face' (3.1). As well as providing an example of a character

admitting his deception to the audience, this particular example reveals how such keywords can point us to explicit characterisation cues (Culpeper, 2001: 164). Not all of the character statements that Aaron makes about himself or others are deception related though. Aaron uses **black**, for example, in reference to his baby son, after the infant is described (by others) as a devil. He also (implicitly) challenges Chiron and Demetrius to recognise their ‘brother’ as their equal, drawing on the keyword **black** (4.2). The differences in characterisation(s), provided by this one keyword, support Quennell and Johnson’s (2013: 1) assessment that Aaron is not a completely evil character, in spite of being described as ‘the evil Moor of Christian tradition...distinguished by...cruelty’ and ‘the direct descendent of the figure of Vice in the medieval morality plays’. Many of Aaron’s actions in the play are deceptive, nonetheless, as revealed by the keywords **villainy** and **gold**. Aaron explains how he is planting a ‘bag of *gold*’ (2.3) by the pit where Bassianus’s body is dumped in order to implicate Martius and Quintus in his murder. He also informs the audience that hiding the gold may appear illogical initially, such that ‘He that had wit would think that [he] had none, / To bury so much *Gold* under a tree’ (2.3). However, he goes on to explain ‘that this *gold* must coin a stratagem, / Which cunningly effected, will beget / A very excellent piece of *villainy*: And so repose sweet *Gold* for their unrest’ (2.3). **Gold**, in this case, provides us with an implicit characterisation cue, thereby allowing us ‘to infer [...] character information’ about Aaron from his ‘linguistic behaviour’ (Culpeper, 2001: 164). Namely, that Aaron is prepared to go to extreme lengths to be rid of enemies: albeit doing so *cunningly* (i.e., duplicitously). **Gold** also provides the audience with a rationale for such behaviour, for Aaron confides his longing to ‘be bright and shine in pearl and gold’ (2.1), thereby hinting at his ambitions for wealth and power. Aaron uses **villainy**, in turn, as an explicit characterisation cue when he describes Tamora as having a ‘sacred wit / To *villainy* and vengeance’ (2.1) that is a match for his. The structure serves to reveal something about himself as well as his opinion of Tamora (his lover), of course: namely, that – from Aaron’s perspective – ‘villainy’ and ‘vengeance’ are skills to be admired/honed/practiced.

3.2 *Tamora*

Keywords like **sweet** are not as obviously connected with deception and manipulation as words like **villainy** (see Section 3.1, above, and also Section 4). Tamora’s repetitive use of it, in the first scene especially, hints at her feminine wiles, nonetheless. She appeals to her ‘*sweet*’ Emperor to ‘pardon what is past’, and encourages him to make up with Titus (1.1). When Saturninus remains slow to forgive certain members of the Andronicus family publicly, Tamora then advises her ‘*sweet* emperor’ they ‘must be friends’, reaffirming to her ‘*sweet* heart’ she ‘will not be denied’

(1.1). Spectators recognise that Tamora is being duplicitous here, however, as she has previously revealed to them her secret desire to ‘find a day to massacre’ the ‘cruel father...and his traitorous sons’ (1.1). In Act 4, Scene 4, Tamora actually brags about her sweet-talking abilities to her cuckolded husband (Saturninus). She tells him she ‘will enchant the old Andronicus [i.e., Titus], With words more *sweet*, and yet more dangerous Than baits to fish, or honey-stalks to sheep’. The keyword **ear** is significant in this regard too, with Tamora claiming she ‘can smooth and fill his aged *ear*, With golden promises, that were his heart Almost Impregnable, his old *ears* deaf, Yet ... both *ear* and heart’ would ‘obey her tongue’ (4.4). The keyword **fear** appears in Tamora’s attempted ruse of Titus (in combination with **ear**). As Revenge, she tells him menacingly, ‘There’s not a hollow Cave or lurking place...Where bloody Murder or detested Rape, can couch for *fear*, but I will find them out, And in their *ears* tell them my dreadful name’ (5.2). Tamora’s other uses of **fear** hint at this particular character’s (almost reckless) fearlessness, with two instances of ‘*fear not*’ (1.1., 2.3), one instance of ‘Why should you *fear*?’ (4.4), and an imperative to Saturninus to ‘bury all [his] *fear* in [her] devices’, given her ability to ‘temper [Titus] with all the Art’ she has (4.4).

3.3 Iago

Like Tamora (above), Iago’s keywords do not appear to relate to deception at first glance, even though many of them do in his case. Let’s consider six, beginning with the keywords, **money** and **purse**. All ten instances of the former, and eight instances of the latter, occur in a speech Iago has with **Roderigo** (a third keyword for the character). The speech was focused upon ‘put[ting] *money* in [his] *purse*’ (1.3), and was designed to achieve two things from Iago’s perspective: to convince Roderigo to follow Othello’s army to Cyprus so that (1) he might bring about Othello’s demise and (2) have Roderigo on hand as a scapegoat. During their interaction, Iago confesses to Roderigo both his hatred for ‘the Moor’ and his desire that Roderigo ‘Cuckold him’. He keeps his manipulation of Roderigo covert until he kills him at the end of the play after Roderigo fails to slay Cassio. **Lieutenant** is a fourth keyword for Iago. The military position, awarded to Cassio, is the reason Iago wants revenge on both him and Othello. When speaking to Cassio directly, Iago refers to him flatteringly as ‘good *Lieutenant*’ (x2) (2.3). By this point, however, spectators have already been privy to Iago’s imagery of a ‘little web’ as a trap, with himself as the spider and Cassio as the ‘Fly’ he will ‘ensnare’ (2.1). The keyword, **sleep** (and its associated terms), is used repetitively (i.e., x7) and manipulatively by Iago at the point he sucks Othello into his lie about Desdemona and Cassio (3.3). He tells Othello, he ‘could not *sleep*’, before hinting at an illicit affair between Desdemona and Cassio. He explains that ‘There are a kind of men, / So loose of

Soul, that in their *sleeps* will mutter / Their Affairs: one of this kind is Cassio: / In *sleep* I heard him say, sweet Desdemona, / Let us be wary, let us hide our Loves, / And then (Sir) would he gripe, and ... kiss ...' (3.3). It is at this point, Othello and Iago begin to plot together to kill Cassio and Desdemona. Iago goes on to feign surprise in front of Emilia and Desdemona, querying repeatedly why Othello 'is/be *angry*' (3.4) - **angry** being another keyword. As Suchet (1989: 195) notes, however, he knows only too well by this point that, if 'Othello is showing his anger *publicly*', it means 'he's [not only] taken the bait' but 'swallowed [it] whole' (original italics). Iago's conniving only unravels when Roderigo's failure to kill Cassio unleashes a chain of events that results in Iago not only killing the former but also his own wife. At the play's end, he defiantly claims he will 'never speak [a] word' (5.2) that might explain what must have motivated such cruel actions, on his part, thereby echoing a self-disclosure he had made to Roderigo (and spectators) about never allowing 'outward action' to reveal his true intention, adding somewhat cryptically, 'I am not what I am' (1.1). Early modern contemporaries would have recognised the inversion of God's statement, 'I am that I am' (Exodus 3.4), and understood that Iago was self-identifying with (and, perhaps, even as) a devil here.

3.4 Lady Macbeth

Lady Macbeth has the fewest keywords of the characters, with only four in total (see Table 1). Although (like Iago) those keywords seem somewhat innocuous, **would** and **without**, in particular, point us towards her manipulative disposition. For example, she uses both when ruminating on whether Macbeth will take (what she perceives to be) the necessary steps to make himself king (after reading about the Weird Sisters' prophesy). She thinks he '*would*'st be great' and is 'not *without* ambition', and knows that he '*would*st highly / That *would*st [he] holily', but also feels he '*would*st not play false' in order to 'win' (1.5). Her doubting Macbeth leads her to resolve she will do all she can to convince her husband to do whatever is required to seize the crown. This includes asking the spirits to 'unsex' her and 'fill' her 'Of direst cruelty' (1.5) so that, having put her natural femininity aside, she will be able to do the bloody deeds necessary to seize the crown (if Macbeth cannot). Lady Macbeth eventually comes to realise their murderous path has been for nothing, however. Drawing once again on the keyword, **without**, she laments that 'Naught's had; all's spent / When our desire is got *without* content' (3.2). Her later sleepwalking scene, near the end of the play (5.1), then ironically inverts the earlier murder scene where she urges Macbeth to wash his hands and get to bed (2.2) – **bed** also being a keyword for Lady Macbeth. It is now she who cannot stay in and hence is '*without* the *bed*' (cf. Macbeth's earlier statement he would 'sleep no more').

3.5 Falstaff

Falstaff has the most keywords of our five characters (23 in total), with several alluding to the (extensive) self-presentation (Marelj, 2019) he engages in throughout the play. One such example of his self-presentation – ‘I am no *counterfeit*’ (5.3) – constitutes the antithesis of what Falstaff’s turns actually suggest about him, given he lies about fighting off attackers, imitates the king, pretends to be robbed, and falsely claims he has killed Hotspur. Note that Falstaff uses the keyword, **counterfeit**, nine times in total in Act 5, Scene 3, as he pretends to be dead in order to avoid losing his life, whilst also musing over whether he is really counterfeiting a dead man, if such counterfeiting means he will live. This and similar instances thus allude to Falstaff’s typical (but not always successful) modus operandi of using his wit in an attempt to turn around any discussion or argument to suit himself (see, e.g. Hayes, 2004: 91-100). Falstaff’s self-centredness can be further evidenced via his uses of the keyword, **honour**. In his soliloquy on the topic, for instance, he highlights its meaninglessness if measured by the ‘grinning *honour*’ of a corpus (5.2). Falstaff tends to treat other characters badly, preferring to ridicule them using the keywords **hang**, **plague** and **cowards**. He is particularly nasty to his perceived competitor for Prince Hal’s attention, **Poins** (another keyword for Falstaff). Four of his keywords - **lad**, **Hal**, **rogue** and **company** – suggest he has a genuine affection for the Prince, however. He confesses to being ‘bewitched / with the *Rogue’s company*’ (2.2), for example. In addition, ‘I/me’ are strong collocates of ‘Hal’ and are used to preface a deferential statement (‘Before I knew thee / *Hal*, I knew nothing’ (2.1)) and polite request (‘O *Hal*, prithee give *me* leave to breathe awhile’ (5.2)). The impression therefore given by Falstaff, when his keywords are viewed collectively, is that of a self-centred (vice-ridden) character who nonetheless attempts to stay in the good graces of the future king. However, one keyword, **sack**, suggests he often struggles to achieve the latter in practice. Indeed, he irritates Prince Hal when he attempts ‘to *sack a city*’ (5.2) with a bottle rather than a sword. His vow to ‘purge, and leave *Sack*’ (5.3) thus constitutes one of his less convincing lies.

3.6 An assessment of the usefulness of keyword analysis in detecting depictions of deception

Archer and Lansley (2015: 245) have argued that the patterns, uncovered by keyness methods, can point not only ‘to the “aboutness” of a given text’ (cf. Phillips, 1989) but also ‘individuals’ emotional and cognitive experiences (Pennebaker, 2011)’ and, more importantly, perhaps, given the focus of this paper, ‘potential cues as to whether...someone is lying’. Our keyword analyses seem to support such a claim – but only if users are careful to take ‘into consideration the context

in which [a particular key]word is used' (McQuaid et al., 2015: 6), including to whom. For it is only by so doing that we can differentiate those instances of a particular keyword that are pointing to deception from those that are not: cf., for example, Aaron's keyword, **black**, which highlights intended duplicity on his part but also his love for his son (see Section 3.1).

Given the emphasis, above, on emotional and cognitive experience(s), it is worth reiterating that we are exploring fictional characters (rather than real people) in this paper. This allows us a greater level of insight into what they were feeling and thinking. Indeed, characters like Aaron and Iago sometimes tell us their thoughts, feelings, plans, and motivations for those plans through their soliloquies or asides to spectators. Characters' keywords are found within these soliloquies and asides, and sometimes point to important characterisation cues (Culpeper, 2001) as well as intentions to deceive or admissions of previous duplicity. By way of illustration, Aaron's mention of **gold** is used to signal his cunning plan to deceive others (*Titus Andronicus* 2.1) as well as the motive(s) that lay behind his actions. Whilst this confirms soliloquies and asides as a good place to focus, when attempting to understand characters' emotional and cognitive experiences, they will not constitute examples of deception *per se*, given that characters tend to confide their 'true feelings and motives' (Clarke, 2013: 209) via such dramatic devices. Section 5 reports, nonetheless, on their capacity to point us to speeches of interest when looking for deception because they contain character reports of previous deceptive moves or future, intended moves on their part.

As previously established, Falstaff was the only character to have a keyword – **counterfeit** – that had any denotative associations with deception prior to analysis. One observation we have made, in respect to this particular keyword (worth reiterating here) is that it was used repetitively in the same scene. Falstaff, like several other characters, also drew on two or more keywords in quick succession: sometimes within the same turn, and sometimes as part of a sequence of turns or a particular scene. This appears to suggest that, were we to develop a predictive method for detecting deception (to be used on factual data), but based in part on keywords, it would be worthwhile to cultivate a means of identifying stretches of language where a number of keywords co-occur (see Sections 5 and 6). This would only constitute a first step, however. We identify keywords as a potential first step *only* here, and in the Introduction, as keywords highlight salient terms. Some might be to do with 'aboutness' (Phillips, 1989), when viewed in their context-of-use. Others might highlight 'individuals' emotional and cognitive experiences', mindstyles (McIntyre and Archer, 2010; McIntyre, 2011) or potential deception cues (Archer and Lansley, 2015: 245). Given that keywords only offer a *way in* to a potential study, be it to begin exploring deception (Archer and Lansley, 2015), mindstyle (McIntyre and Archer, 2010), cognitive

metaphor (Archer et al., 2009), etc., we believe there is much to be gained by combining this method with forensic linguistic knowledge about deception and its detection.

Previous researchers who have done this include Archer and Lansley (2015) and McQuaid et al. (2015), both of whom analysed public appeals. McQuaid et al. (2015) opted to pay attention to certain semantic categories that, having achieved an LL score of 6.63 or higher, were believed to be relevant to this particular context (e.g. level of knowledge, descriptions of personality characteristics and physical attributes, self and other references, and the use of discourse markers). Archer and Lansley (2015), in contrast, used the keyword methodology to supplement results based on coders annotating public appeals for 'Points of Interest' or PIns. That is, 'relevant occurrences of twenty-seven criteria'³ (validated through previous research) occurring across two or more communication channels, be it the Face, Body, Voice, the Language channels or the Autonomic Nervous System (Archer and Lansley, 2015: 231) within a seven-second window. Our methodology is slightly different again. Having engaged in a keyword analysis as a first step, we now draw upon the same wordlists used for the keyword analyses to determine whether/the extent to which:

1. Our five characters make use of language features associated with deliberate deception in real-world contexts (see Sections 4.1 and 4.2).
2. These features co-occur with characters' keywords (see Section 5).

Accordingly, Section 4 (below) provides an overview of the sorts of language features that the extant literature has associated with deception (and its detection) to date.

4. Language features associated with deception

As discussed in the Introduction, there are a number of linguistic features that have been identified by previous research as pointing towards potential deception. These include but are not be limited to:

- changes in pronoun usage (cf. DePaulo et al., 2003; Newman et al., 2003; Bond and Lee, 2005);
- increases in the use of negation, including spontaneous negation (Schafer, 2007);
- negative emotion terms or disparaging terms (Vrij, 2000; Newman et al., 2003);
- emotiveness more generally (Burgoon and Quin, 2006);
- refusals to discuss certain topics, the use of unnecessary connectors (*then*, *next*, etc.) to pass over information (DePaulo et al., 2003), and other types of distancing behaviour (Hancock

et al., 2013; McQuaid et al., 2015) such as the use of hedging devices (Adams and Jarvis, 2006);

- repetition (Wright Whelan et al., 2013);
- decreases in temporal and spatial indicators (McQuaid et al., 2015);
- the provision of extraneous information (Wright Whelan et al., 2013), in an attempt to keep an interaction on ‘safe’ topics for the would-be deceiver;
- changes in the frequency of and, in particular, an increasing use of disfluencies (Arciuli et al., 2010), exclusivisers/particularisers such as *just* (McQuaid et al., 2015) and/or qualifiers such as *absolutely, to be honest*, etc. (Enos et al., 2007);
- use of stance or modality markers (e.g. *know, think*) (McQuaid et al., 2015; Zuckerman et al., 1981) in order to create a particular narrative for the would-be deceiver;
- an increase in motion words (walk, go, move, etc.) as a means of dealing with cognitive complexity (Newman et al., 2003).

In the next section, we outline the extent to which a selection of these features occurred in the interactions of the five Shakespearean characters. Our starting point will be their pronoun usage, given that pronouns were keywords for two of the five characters. As Table 1 reveals, although they tended to overuse different pronouns (i.e., Iago – *thyself*; and Lady Macbeth –*your*), both were examples of other-oriented references. Like McQuaid et al. (2015), we will thus be dividing the different pronoun forms, within a character’s speech, into (singular and plural) self-oriented references and (singular and plural) other-oriented references. The five categories adopted (see below) are based on Gillings (2020). As there is a low(er) degree of evidence that some pronouns (i.e. *my/mine/myself/ours/ourselves/possessives*) link reliably to deception (given their lack of frequency in the Gillings dataset and little discussion of them by other deception detection researchers), they are not included in this analysis.

Table 2: Explanation of self-references and other-references

Cue	Part-of-speech tag	Examples
Self-references (singular)	PPIS1 and PPIO1	I and me
Self-references (plurals)	PPIS2 and PPIO2	we and us
Other-references (singular)	PPHO1 and PPHS1	him/her and he/she
Other-references (plural)	PPHS2 and PPHO2	they and them
Other-references (other)	PPY	You
	PPH1	It

The sorts of divisions noted in Table 2 mean that we can assess whether our results reflect the widely-held belief that (modern) deceptive speakers tend to use fewer (singular) self-references,

and more other-references, compared to their more truthful counterparts (see, e.g. Newman et al., 2003). To do this meaningfully, however, we are required to have other (more truthful) Shakespearean characters against whom we can compare our ‘deceptive’ characters. Accordingly, we have identified five additional characters known for their sincerity, as signalled by their associations with ‘blunt’ or ‘plain’ speaking (see, e.g. Graham, 1991; Maley and Tudeau-Clayton, 2010; Grennan, 1987; Cook, 1998) - Othello, Emilia, Humphrey, Brutus and Katherine.⁴ We will outline the extent to which they use the same language features as our five ‘deceptive’ characters in the sections that follow. For ease of comparison within this analysis, we will refer to these characters collectively as either ‘deceptive’ or ‘plain’.

4.1 Characters’ use of self-oriented and other-oriented references

Table 3 captures our comparison of the five ‘deceptive’ characters’ use of self-referencing pronouns with our five ‘plain’ characters’ usage. Normalised figures calculated according to instances per 4,000 words (i.e., the average number of words spoken by all the characters combined) are included in the Table (in italics), alongside raw figures, to make it easier for readers to compare both individual totals for each character and collective totals. Note, for example, that – collectively speaking – our deceptive characters use 171.18 self-referencing pronouns per 4,000 words, in comparison to the ‘plain’ speakers’ 193.94:

Table 3: Use of self-referencing pronouns by ‘deceptive’ characters (compared with ‘blunt’ characters)

Language feature	Aaron	Tamora	Iago	Lady Macbeth	Falstaff	Othello	Emilia	Humphrey	Brutus	Katherine
Pron. – self ref sing. (<i>I, me</i>)	88	72	320	43	333	289	73	87	205	145
	125.7	145.8	152.5	90.2	245.7	185.4	161.1	148.5	147.2	223.9
Pron. – self ref plural (<i>we, us</i>)	12	14	25	12	18	13	14	8	62	4
	17.1	28.4	11.9	25.2	13.3	8.3	30.9	13.7	44.5	6.2
Individual Total	100	86	345	55	351	302	87	95	267	149
	142.9	174.2	164.4	115.4	259	193.7	192	162.2	191.7	230.1
Collective (mean) Total (normalised)	171.18						193.94			

Whilst this means that, as predicted in relation to deceptive speakers, our ‘deceptive’ characters use slightly fewer self-referencing pronouns than our ‘plain’ characters, it is also worth noting that the range of the ‘deceptive’ characters’ self-referencing pronouns (from 115.4 to 259) is wider than the ‘plain’ characters’ (from 162.2 to 230.1) range. This is largely due to one character: Falstaff.⁵ Our keyword analysis of this particular character can help to shed light on why he may be an atypical deceptive character: it confirms Falstaff to be obsessed with himself (see Section 3.5). By way of illustration, Falstaff uses the ‘I’/‘me’ pronouns 333 times in total, with the result that their LL value is high: 78.73. Although these self-referencing pronouns are not included in Table 1, because of having an LR value of 0.80, the figures help to underscore, nonetheless, that Falstaff’s use of them is high, especially when compared with the other ‘deceptive’ characters – two of whom had other-referencing pronouns as keywords (see Table 1 and Section 4).

If we compare their (and Falstaff’s) use of other-referencing pronouns, with the ‘plain’ characters’, we find (perhaps unsurprisingly, given the aforementioned keywords results) that their collective total is higher than the ‘plain’ characters, with the former using 238.06 other-referencing pronouns per 4,000 words, and the latter using only 158.96 per 4,000 words (see Table 4).

Table 4: Use of other-referencing pronouns by ‘deceptive’ characters (compared with ‘blunt’ characters)

Language feature	Aaron	Tamora	Iago	Lady Macbeth	Falstaff	Othello	Emilia	Humphrey	Brutus	Katherine
Pron. – other ref sing (<i>him/her, he/she</i>)	37	28	270	23	64	145	53	18	95	41
	52.8	56.7	128.7	48.2	47.2	93.0	116.9	30.7	68.2	63.3
Pron. – other ref plural (<i>they, them</i>)	16	23	39	20	40	22	18	12	29	25
	22.8	46.6	18.6	42.0	29.5	14.1	39.7	20.5	20.8	38.6
Pron. – 2 nd pers. other (<i>you, thou, thee, thy</i>)	99	70	253	65	213	137	23	43	90	8
	141.4	141.8	120.6	136.3	157.1	87.9	50.7	73.4	64.6	12.4
Individual Total	152	121	562	108	317	304	94	73	214	74
	217	245.1	267.9	226.5	233.8	195	207.3	124.6	153.6	114.3
Collective Total (normalised)	238.06					158.96				

This means that, like deceptive speakers more generally, our ‘deceptive’ characters appear to be putting the focus on other characters, away from themselves, by using more other-references on average. Note that two of Iago’s results are particularly high in this regard, even though Table 1 shows only one other-referencing pronoun as a keyword: ‘thyself’. However, other-referencing pronouns (‘he’/‘him’, ‘his’ and ‘you’) did have high LL values in this character’s case (i.e., 24.45, 12.01 and 8.88 respectively). This suggests that Iago, even more than our other characters, is particularly adept at taking the focus away from himself. In fact, we found ‘I’ to be a negative keyword for Iago, with an LL value of -12.94 (and LR value of -0.33), meaning that he underuses it, statistically speaking, when compared with the other characters from *Othello*.

4.2 Characters’ use of additional language features associated with deception

Thanks to the addition of annotation at both the part-of-speech level and semantic field level, as part of the AHRC-funded *Encyclopedia of Shakespeare’s Language* project, we are able to explore some of the additional features associated with deception that we outlined in Section 4. This has been carried out using the new interface within CQPweb (see Section 2.1). As Table 5 highlights, we are focusing on six (linguistic) features: (grammatical) negation, modal verbs, the stance markers *think* and *know*, the exclusiviser *just*, and hedging devices (i.e., approximators, diminishers and minimizers). Naturally, for such an analysis, the addition of part-of-speech and semantic tagging within the ESC: Folio was invaluable for three main reasons. Firstly, it has allowed us to search for all forms quickly and easily, using just one tag. Secondly, it has ensured that our searches are standardised (i.e. does not introduce another layer of subjective (researcher) interpretation as to what might be deemed a *hedging device* and what might not). Thirdly, it allows us to contribute to the growing body of research within corpus-based deception detection which also draws upon the same tagging systems, CLAWS and USAS (Archer and Lansley, 2015; McQuaid et al., 2015, Gillings, 2020). This latter point is important: there is now substantial and consistent evidence that *just* tagged as an exclusiviser appears more frequently in deceptive communication (Archer and Lansley, 2015; McQuaid et al., 2015), for example. Yet, further analysis of those concordance lines suggests it is mostly being used as a hedge (Gillings, 2020). The USAS tagging, here, allows researchers to find those instances efficiently for further interrogation.

Table 5: Characters' use of language features associated with deception

	A	T	I	LM	F	O	E	H	B	K
negation (not)	30	19	107	28	86	77	30	24	82	29
	42.8	38.5	51.0	58.7	63.4	49.4	66.2	41.0	58.9	44.8
modal verbs	83	48	231	43	135	140	46	56	166	59
	118.5	97.2	112.9	90.2	99.6	89.8	101.5	95.6	119.2	91.1
stance marker (know)	6	7	21	2	10	15	5	3	13	8
	8.6	14.2	10	4.2	7.4	9.6	11	5.1	9.3	12.4
stance marker (think)	4	2	19	3	6	15	9	5	10	3
	5.7	4.1	9.1	6.3	4.4	9.6	19.9	8.6	7.2	4.6
exclusivisers (e.g., just)	0	2	0	5	5	5	0	0	4	2
	-	4.1	-	10.5	3.7	3.2	-	-	2.9	3.1
Hedging	11	3	35	5	10	20	3	2	15	7
	15.7	6.1	16.7	10.5	7.4	12.8	6.6	3.4	10.8	10.8
individual totals (+ normalised)	134	81	413	86	252	272	93	90	290	108
	191.3	164.2	199.7	180.4	185.9	260.8	205.2	153.7	208.3	166.8
collective (mean) total (+ normalised)	193.2 (184.3)					170.6 (198.96)				

Once again, our table includes the results for both the 'deceptive' characters and 'plain' characters to allow a comparison between the two. Note that, contra our earlier findings (see Tables 3 and 4), our 'deceptive' characters – collectively speaking at least – are no longer behaving in similar ways to (modern) deceptive speakers by using more of these features. Indeed, the normalised figures show that the 'plain' characters use the six language features, when combined, slightly more than the 'deceptive' characters (i.e., 198.96 times compared with 184.3 times [per 4,000 words]). If we break these figures down, however, we find that the five 'deceptive' characters do use three of the six features more than their five 'plain' counterparts (collectively speaking). These are:

- Modals (103.68 times compared with 99.44 times [per 4,000 words]).
- The exclusiver *just* (3.66 times compared with 1.84 times [per 4,000 words]).

- Hedging devices (11.28 times compared with 8.88 times [per 4,000 words]).⁶

There are also interesting differences if we look at individual character's use of the six language features. Iago uses the most hedging devices of both the 'deceptive' characters and 'plain' characters (i.e., 16.7 [per 4,000 words]), for example. In contrast, Lady Macbeth uses *just* more than any other character, who in turn uses it more than double the amount as the next highest user: Tamora (i.e., 10.5 times compared with 4.1 times [per 4,000 words]). Two 'deceptive' characters, in addition to Tamora, are also a second-highest user of one or more of the six language features. Falstaff uses *not* 63.4 times (per 4,000 words), for example, and Aaron uses modal verbs and hedging 118.5 times and 15.7 times (per 4,000 words) respectively. Tamora is the highest user of the stance marker *know* (i.e., 14.2 times [per 4,000 words]), even though 'plain' characters use this stance marker the most overall (i.e., 9.48 times compared with 'deceptive' characters' 8.8 times [per 4,000 words]).⁷

The above findings support our earlier keyword analyses, in the sense that there appears to be interesting differences between Shakespeare's depictions of our five characters' attempts at deception, stylistically speaking. To gain a sense of how nuanced these differences are, Sections 5-5.2 will go on to explore certain stretches of two of the 'deceptive' characters' turns, but qualitatively. Our specific focus will be upon speeches where keywords and pronouns feature (and preferably co-occur). This is based upon:

- our findings, from Sections 3-3.5, suggesting that keywords may be a route into studying how Shakespeare stylistically portrays his deceptive characters (assuming, of course, that any relevant keywords are checked, first, in order to determine their use in context), and
- our findings from Section 4.1 suggesting that pronouns are being used by our 'deceptive' characters in similar ways to how (modern) deceptive speakers use pronouns.
- a keenness to determine whether additional deception-related language features co-occur with each other and/or with pronouns and keywords (but see also endnote 6).

5. Evidence of keyword/language feature clustering

Due to length constraints, our qualitative analyses focus on four of Aaron's turns, taken from Act 3, Scene 1 of *Titus Andronicus*, and seventeen of Falstaff's turns, taken from Act 2, Scene 4 of *The First Part of Henry IV*. Our choice of these particular characters is based on their use of keywords that obviously connect to deception and manipulation at first glance (i.e., **villainy** and

counterfeit). The results thus far suggest that the character, Aaron, acts as deception researchers anticipate deceptive individuals would. He uses fewer self-references than other-references, for example (e.g. 217 compared with 142.9 instances per 4,000 words: see Section 4.1). One particular other-referencing pronoun - 'you' – also has a high LL value (of 10.51). Modal verbs and hedging devices feature quite frequently too, such that Aaron uses them more than all but Brutus (with respect to modal verbs) and Iago (with respect to hedging), with 118.5 and 15.7 instances per 4,000 words respectively (see Section 4.2). Our results thus far suggest that Falstaff is not typical deceptively speaking, however. For example, although he uses 'not' more than all but *Othello*'s Emilia (i.e., 63.4 times per 4,000 words: see Section 4.2), he uses more self-references than other-references (e.g. 259 compared with 233.8 instances per 4,000 words: see Section 4.1).

5.1 Analysis of Aaron's turns

Act 3, Scene 1 has already been identified to be an important scene for Aaron, in keyword terms, due to his excitement at tricking Titus. A relevant turn, in this regard, is: 'how this *villainy* / Doth fat me with the very thoughts of it, / Let fools do good, and fair men call for grace, / Aaron will have his soul *black* like his face', containing the keywords **villainy** and **black** (see Section 3.1). Act 3, Scene 1 starts with Titus pleading for mercy (in vain) for the sons Aaron has framed for murder. Titus then witnesses another son, Lucius, being sentenced to banishment when his attempt to free Martius and Quintus fails. As the mutilated Lavinia cries, Titus ruminates on what to do in order to help her, even contemplating cutting off his own hands and biting off his own tongue (so as to better comprehend her plight). It is at this point, Aaron arrives and (as it turns out falsely) claims Saturninus is willing to let Martius and Quintus go free were one of the Andronicus men to cut off their hand and send it to the Emperor. Although Lucius and Marcus both offer their own hands, Titus is quick to ensure Aaron takes his in exchange for the lives of his sons. As Aaron leaves, he gleefully utters the line noted above (containing the keywords, **villainy** and **black**). A messenger then arrives, shortly thereafter, with Titus's hand in a sack, along with the severed heads of his sons. As noted in Section 3.5, 'asides' like Aaron's not only provide us with a useful insight into these character's 'true feelings' (Clarke, 2013: 209) and psychological make-up but also their plans and motivations for such plans (including plans to deceive others). Accordingly, Table 6 (below) captures Aaron's turns relating to the Emperor's (apparent) willingness to make a bargain with the Andronici, and the keyword, pronouns and other deceptive features found within each turn. This includes two asides.

Table 6: Characteristics of Aaron's turns

No.	Turn	Keywords	Other-referencing pronouns	Self-referencing pronouns	Additional deceptive features
1.	Titus Andronicus, my Lord the Emperor, / Sends thee this word, that if thou love thy sons, / Let Marcus, Lucius, or thyself old Titus, / Or any one of you, chop off your hand / And send it to the King: he for the same / Will send thee hither both thy sons alive / And that shall be the ransom for their fault.		thee (x2), thou, thy (x2), thyself, you, your, he		will, shall
2.	Nay come agree, whose hand shall go along / For fear they die before their pardon come.		they,		shall
3.	If that be called deceit, I will be honest, / And never whilst I live deceive men so: / But I'll deceive you in another sort, / And that you'll say ere half an hour pass.		you (x2)	I (x3)	will, 'll (x2)
4.	I go Andronicus, and for thy hand/ Look by and by to have thy sons with thee: / Their heads I mean: Oh How this villainy / Doth fat me with the very thoughts of it. Let fools do good, and fair men call for grace, / Aaron will have his soul black like his face.	villainy, black	thy (x2), thee, his (x2)	I (x2), me	will

Although Table 6 contains four turns only, there is clear evidence of two keywords, ten different forms of other-referencing pronouns (all but five of which occur more than once) and three modal forms (all of which are used more than once). The pronoun, *their*, also occurs but is not highlighted here as it was not included in our analysis. Aaron's use of self-referencing pronouns is highlighted in Table 6, of course. Notice, however, that these particular references are far less frequent. With the exception of one ('my', which was not included in our analysis in Section 4.2 and is thus not highlighted above) they occur, moreover, as part of Aaron's asides, alongside words that make clear this is a trick on Aaron's part, namely, *deceit*, *deceive* (x2) and **villainy**. This provides us with two useful findings:

1. The identification of (a cluster of) deception-related features within dramatic devices like asides (in line with Aaron informing the audience of his duplicitous motives).
2. Clear evidence that features - which previous research has associated with deception - were used collectively/in close proximity, by Shakespeare, at points where Aaron is known to speak to other characters disingenuously.

Our first finding (above) is more than likely representative of fictional work only, given that real-life deceivers tend to want to remain covert (as a means of hiding their deceptive intent). The second finding, however, tallies with Archer and Lansley's (2015: 232-233, 236) discovery that behaviour 'suggestive of inconsistency', and hence deception, in real-life contexts, tends to cluster 'within a seven-second window' (see also Houston et al., 2012).⁸ As we will be discussing the notion of deceptive indicators clustering in our Conclusion section, we will not discuss this in more detail here, beyond suggesting that it is the clustering effect that arguably gives this particular depiction of deception some sense of realism. As Picornell (2013: 33) notes, nonetheless, a real-life, would-be deceiver tends to have to 'adapt to the changing landscape, altering and restructuring their communication blueprint as they encounter obstacles'. Aaron does not encounter obstacles until much later in the play. This explains, in part, the scene we have thus chosen for Falstaff, which does see him have to overcome such obstacles (see Section 5.2), thereby allowing us to determine how Shakespeare's deceptive depictions change stylistically, from character to character, and from context to context (see Introduction).

5.2 Analysis of Falstaff's turns

Although we did not discuss Act 2, Scene 4, in Section 2.5, when analysing Falstaff's keywords initially, Falstaff's turns are replete with keywords, in this scene, with all but four (**honour, wag, reward, tavern**) featuring, and all but four of the nineteen that do feature (**counterfeit, Bardolf, hang, pick**) occurring repeatedly. **Banish** occurs in this scene only, moreover, and is repeated eight times by the character.

When we are first introduced to Falstaff in the scene at the Boar's Head, he is venting at the other characters. After (repeatedly) being asked why this is so, he gives a false report of a robbery. Table 7 below captures the first 17 turns that make up that explanation, prior to being told, by the Prince, that his 'Lies are like the Father that begets them, / gross as a Mountain, open, palpable' (2.4).

Table 7: Characteristics of Falstaff's turns

No.	Turn	Keywords	Other-referencing pronouns	Self-referencing pronouns	Additional deceptive features
1.	Here be four of us, have / taken a thousand pound this Morning,		us		
2.	Where is it? / taken from us, it is: a hundred / upon poor four of us.		us (x2)		
3.	I am a Rogue, if I were not at half Sword with / a dozen of them two hours together. / I have escaped by	rogue,	them (x2), they (x2)	I (x6)	not (x2), would

	/ miracle. / I am eight times thrust through the Doublet, / four through the Hose, my Buckler cut through and / through, my Sword hacked like a Handsaw, ecce signum. / I never dealt better since I was a man: all would not do. / A plague of all Cowards: let / them speak; if they speak / more or less than truth, they / are villains, and the sons / of darkness.	if (x2), plague, coward			
4.	You Rogue, they were bound, every man of / them, or I am a Jew else, an / Hebrew Jew.	rogue	you, they, them	I	
5.	And unbound the rest, and then come in the other.				
6.	I know not what ye call all		ye		know, not
7.	All? / I know not what ye call all: but if I / fought not with fifty of the, I am a bunch of Radish: / if there were not two or three and fifty upon poor old / Jack, then am I no two-legged Creature.	old , fifty	ye	I (x4)	know, not (x3)
8.	Nay, that's past praying for, I have peppered / two of them: Two I am sure I have paid, two Rogues / in Buckram Suits. / I tell thee what, Hal, if I tell thee a / Lie, spit in my face, call me a Horse: thou knowst my old / word: here I lay, and thus I bore my point; four Rogues in Bukram let drive at me.	rogue (x2), Hal , if , old	them, thee (x2)	I (x7), me (x2)	knowst
9.	Four, Hal, I told thee four.	Hal	thee	I	
10.	These four came all affront, and mainly thrust / at me; I made no more ado, but took all their seven / points in my Target, thus.		their	I, me	
11.	In Buckram.				
12.	Seven, by these Hits, or I am a Villain else.			I	
13.	Dost thou hear me, Hal?	Hal	thou	me	
14.	Do so, for it is worth the listening to: these / nine in Bukram, that I told thee of.		thee	I	
15.	Their Points being broken.		their		
16.	Began to give me ground: but I followed me / close, came in foot and hand; and with a thought, seven of / the eleven I paid.	Give		I (x2), me (x2)	
17.	But as the Devil would have it, three misbegotten / Knaves, in Kendal Green, came at my Back,	Hal	thou, thy	Me	would, couldst, not

	and / let drive at me; for it was so dark, Hal, thou couldst not see thy Hand.				
--	--	--	--	--	--

It was not just the Prince's intervention, above, that indicated Falstaff's audience of characters did not believe his lies. Indeed, there were comments throughout - from various other characters too – at the points where Falstaff made glaring errors with respect to how many assailants he had actually fought off (which ranged from two to eleven). Not surprisingly, these seventeen turns are characterised by a number of features associated with deception, in addition to eight keywords, which, in five cases, are used repeatedly: **Hal** (x4), **fifty** (x2), **plague**, **rogue** (x4), **coward**, **old** (x2), **if** (x5), **give**. The deception-related features include:

- Twenty-one instances of other-referencing pronouns, specifically, 'they' (x3), 'them' (x4), 'ye' (x2), 'you', 'thou' (3), 'thee' (x4), 'thy' and 'it' (x3), as well as 'their' (a term not included in our analysis in Section 4.2)
- Modals - specifically, 'would' (x2) and 'couldst'
- Three instances of the stance marker 'know/st'
- Seven instances of 'not'.

There are thirty-five instances of self-referencing pronouns in the seventeen turns, however – i.e., 'I' (x29), 'me' (x3) and 'us' (x3) – as well as seven instances of 'my' (a term not included in our analysis in Section 4.2). Such a finding suggests that understanding deception requires deception-focussed researchers to pay as much attention to how particular pronouns are used, in context, as they do their frequencies. The statistical collocates of 'I' are helpful here as, in the above seventeen turns - and/or in the play overall - Falstaff consistently uses 'I' with 'll', 'am', 'give', 'prithee', 'tell', 'fear', 'know', 'would', 'have', 'him' and 'thee'.⁹ 'I am' structures, in turn, are paralleled with 'if' constructions in the example above, thereby helping to give these turns their exaggerated, hyperbolic feel. Falstaff's 'playing loose with the truth and exaggerating things' in this way is not just a part of his characterisation, of course, but also a device via which Shakespeare can expose him 'as a comic fraud' for the purposes 'of humour' (Berger, 2011: 19-20). Whilst the humour gives this depiction of deception a different feel from Aaron's more menacing depiction in *Titus Andronicus* (cf. Section 5.1), we believe there are touches of realism nonetheless. By way of illustration, Falstaff seeks to escape embarrassment on learning that it was the prince and Poins who robbed him at Gad's Hill. He does this by claiming he 'knew' it was Hal all along, given he knows him 'as well as' his father, and then by feigning deference: 'Was it for me to kill the Heir apparent? / Should I turn upon the true Prince? / Why, thou knowst / I am as valiant as Hercules, but beware instinct, the Lion / will not touch the true Prince: Instinct is a

great matter' (2.4). Deception-related linguistic features drawn upon by Shakespeare, here, include negation ('not'), modals ('should' and 'will') and stance markers ('know'), as before. According to Marelj (2019: 43), 'know' does not merely point to 'Falstaff's stupendous, improvised lies', though. Along with the theatrical pun on 'true', she argues it gives 'rise to implicatures' about Hal being a 'counterfeit prince' but 'genuine (true) thief' (Marelj, 2019: 43). Interestingly, this paralleling of veracity with comprehension is a feature of one of Falstaff's earlier exchanges with Prince Hal too. He tells him, somewhat tongue-in-cheek: 'O, thou hast damnable iteration, and art indeed / able to corrupt a Saint. / Thou hast done much harm unto / me Hal ... Before I knew thee / Hal, I knew nothing: and now I am (if a man should speak / truly) little better than one of the wicked. / ... I am a / Villain' (1.2). As it involved Falstaff projecting his own faults onto the Prince, one might also see it as another occasion where the humour in *The First Part of King Henry IV* takes on a darker slant, both dramatically and psychologically speaking (cf. Cardullo, 2006).

Our qualitative analysis of Falstaff's turns not only appears to reaffirm the second finding highlighted in Section 5.1 (based on our analysis of Aaron) – that deception-related features are, indeed, used collectively/in close proximity, by Shakespeare, at points where a character speaks to other characters disingenuously – but also allows us to advance three additional findings. Namely, that:

1. Shakespeare's deceptive depictions *do* change stylistically, from character to character, in line with those characters' different characterisations, situations, etc.
2. Shakespeare also draws on atypical language features - such as self-oriented references - when it comes to some of his depictions of deception (see Sections 4.1 and 5).
3. He uses these various stylistic features to achieve a range of dramatic effect(s).

The first finding (noted above) is somewhat predictable, but shows, nonetheless, that different characters draw upon the same deceptive indicators, but to different degrees, and also sometimes draw upon different indicators when engaged in or reporting on (previous and/or intended) acts of duplicity: much like their real-life counterparts. Shakespeare's depiction of Falstaff's duplicitous acts also see him (attempting to) 'adapt to the changing landscape...as [he] encounter[s] obstacles' (cf. Picornell, 2013: 33). Falstaff is largely responsible, in turn, for our second finding, having used many more self-oriented references than the other 'deceptive' characters. In Section 4.1, we highlighted how Falstaff's keywords can help to shed light on his *a*-typicality compared to the other 'deceptive' characters, as they tend to confirm Falstaff as being obsessed with himself (see Section 3.5). This feeds into our third finding, which highlights another important observation, which may seem obvious – but is often glaringly absent from the deception

detection literature. Namely, that features commonly associated with deception are also used by characters – as well as real-life speakers – to do more than just deceive.

6. Concluding comments and future studies

As well as demonstrating the usefulness of a new Shakespearean resource, the ESC: Folio, this study has provided analyses that are likely to be of interest to several audiences. We have contributed to the small body of work, within stylistics, that has drawn upon fiction to explore the language of deception and manipulation (see, e.g. Sorlin, 2016). We have also extended such work by purposefully exploring - in this case, Shakespearean - drama using a number of linguistic features linked with 'real world' deception across a variety of contexts (see especially Sections 4 and 5). Our aim, here, was to contribute to the work within English Literature that has sought to understand the playwright's grasp of and ability to depict, appositely, human traits such as falseness, cruelty, gullibility, folly and manipulation (see, especially Sections 5.1 and 5.2). However, we would caution against a straightforward mapping of these features of deception from the 'real world' onto fictional (and in this case, dramatic) contexts - without careful consideration of whether it is appropriate to do so. We highlight this caveat as many of the features of deception discussed above have been identified through the use of experimental work in psychology (over the past thirty years), with the result that there is no obvious connection to Early Modern English drama. What we hope our analysis shows, however, is that - by searching for these in combination with corpus linguistic techniques such as keyword analysis and a close reading of relevant concordance lines - we are able to go one step further towards assessing the *realism* of Shakespeare's depictions of deception.

Due to our focus on the language level, this study is also likely to be of interest to audiences interested in deception more generally. It is important to emphasise, here, that this work differs significantly from traditional deception detection work in psychology in many ways. Context is paramount within this study, for example. Indeed, rather than only making claims and predictions based on the different frequencies of a collective set of features, we argue that deception is best investigated by careful examination of features *alongside* their wider context within a scene (or, even, a turn). We also identify how dramatic devices that would not occur in real life – in particular, soliloquies and asides – can be used by playwrights (Shakespeare included) to keep their audiences *in the know* (when other characters remain in the dark).

Our future work aims to further develop the analysis we carried out in Section 5 in particular. That analysis suggests that treating features of deception as a *set* is perhaps not fruitful (see, e.g.,

endnote 6). In response, we are advocating an approach based on identifying a selection of co-occurring features - including keywords. As Archer and Lansley (2015) note, identifying language features, and especially keywords, in real time – as a means of identifying potential attempts at deception – ranges from difficult to impossible currently. Given that this pilot provides us with a strong indication that deceptive features do cluster during speeches where characters are *known* to be being deceptive (because they confess as such, during asides), we believe that an important next step is to investigate the co-occurrence of features as robustly as possible. This paper has drawn upon the notion that behaviours ‘suggestive of inconsistency’, akin to possible deception, cluster ‘within a seven-second window’ (Archer and Lansley, 2015:232–3, 236). As noted in endnote 8, Archer and Lansley (2015) pay attention to a seven-second time span following a stimulus, such as a question, based on an understanding of being able to ‘reasonably conclude...the behaviour is directly associated with [that] stimulus’ (Houston et al., 2012:30). In future work, we plan to test this further, on both fiction like Shakespeare’s plays, which are designed to be performed, as well as the turns (verbal and written) of real-life interlocutors. We will aim, for example, to (i) measure the density of a set of features across a dataset as a whole and then at pivotal points, which may or may not have been previously identified as deceptive, and (ii) assess the best (predictive) span size when capturing a ‘clustering’ effect. The ultimate aim is to determine whether a predictive model based on the clustering of features is possible and, if so, how close to real time this could be operationalised. This work is therefore likely to be of interest to a number of practitioners (from the forensic and security fields especially) in addition to academics.

Acknowledgments

We are grateful to the anonymous reviewers for their comments on this paper. Any remaining infelicities are ours.

Declaration of conflicting interests

We, the author(s), declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The research presented in this article was supported by the UK's Arts and Humanities Research Council (AHRC), grant reference AH/N002415/1. Before the end of 2020, the Shakespeare Corpus and the Comparative Corpus will be made publicly available, first via the CQPweb interface and then through download.

ORCID IDs

Dawn Archer: <https://orcid.org/0000-0002-4547-6518>

Mathew Gillings: <https://orcid.org/0000-0002-6114-0117>

References

Archer D and Bousfield D (2010) “‘See better, Lear?’ See Lear better! A corpus-based pragmatical investigation of Shakespeare’s King Lear. In: McIntyre D and Busse B (eds) *Language and Style*. London: Palgrave, pp. 183-203.

Archer D and Findlay A (in press). Keywords that characterise Shakespeare’s (anti)heroes and villains. In: Jonsson E and Larsson T (eds) *Voices of English: Tapping into records past and present*. Amsterdam/Philadelphia: John Benjamins.

Archer D, Culpeper J and Rayson P (2009) Love – ‘A familiar or a devil?’ An exploration of key domains in Shakespeare’s comedies and tragedies. In: Archer D (ed) *What’s in a word-list?: Investigating word frequency and keyword extraction*. Ashgate: Routledge, pp. 137-157.

Archer D and Lansley C (2015) Public appeals, news interviews and crocodile tears: an argument for multi-channel analysis. *Corpora* 10(2): 231–258.

Arciuli J, Mallard D and Villar G. (2010) “‘Um, I can tell you’re lying’: Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics* 31(3): 397-411.

Berger AA (2011) *The Art of Comedy Writing*. New Brunswick and London: Transaction Publishers.

Bond GD and Lee AY (2005) Language of lies in prison: linguistic classification of prisoners’ truthful and deceptive natural language. *Applied Cognitive Psychology* 19(3): 313–329.

Burgoon J and Qin T (2006) The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology* 25(1): 76–96.

Cardullo B (2006) *A Critical Edition of Two Modern Plays on the Dramatic Character of Sir John Falstaff*. Lewiston, NY: Edwin Mellen Press.

Clarke K (2013) Reading *As You Like It*. In: Goodman L and Owens WR (eds) *Shakespeare, Aphra Benn and the Canon*. London: Routledge, pp. 193-250.

Cook A (1998) The ordering effect of dramatized history: Shakespeare and Henry VIII. *The Centennial* 42(1): 5-28.

Culpeper J (2001) *Language and characterisation: people in plays and other texts*. Harlow: Longman.

Culpeper J (2002) Computers, language and characterisation: an analysis of six characters in Romeo and Juliet. In: Melander-Marttala CO and Kyto M (eds) *Conversation in Life and in Literature: Papers from the ASLA Symposium*. Uppsala: Association Suedoise de Linguistique Appliquée, pp. 11-30.

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, Hardie A, Demmen J, Timperley M, and Hughes J (in prep) Supporting studying Shakespeare: Structural markup and grammatical annotation enhancing a corpus of the First Folio. *Digital Scholarship in the Humanities*.

Demmen J (2020) Issues and challenges in compiling a corpus of Early Modern English plays for comparison with those of William Shakespeare. *ICAME Journal* 44(1): 37-68.

Demmen J and Culpeper J (2015) Keywords. In: Biber D and Reppen R (eds) *The Cambridge Handbook of Corpus Linguistics*. Cambridge: Cambridge University Press, pp. 90-105.

DePaulo BM, Lindsay JJ, Malone BE, Muhlenbruck L, Charlton K and Cooper H (2003) Cues to deception. *Psychological Bulletin* 129(1): 74–118.

Enos F, Shriberg E, Graciarena M, Hirschberg J and Stolcke A (2007) Detecting deception using critical segments. In: *Interspeech Proceedings*, ACL, pp.2281-2284.

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.

Gillings M (2020) A corpus-based investigation into verbal cues to deception and their sociolinguistic distribution. PhD Thesis, Lancaster University, UK.

Graham KJE (1991) "Without the Form of Justice": Plainness and the Performance of Love in "King Lear". *Shakespeare Quarterly* 42(4): 438-461.

Grennan E (1987) The Women's Voices in "Othello": Speech, Song, Silence. *Shakespeare Quarterly* 38(3): 275-292.

Hancock JT, Woodworth MT and Porter S (2013) Hungry like the wolf: a word-pattern analysis of the language of psychopaths. *Legal and Criminological Psychology* 18(1): 102–114.

Hayes DW (2004) *Rhetorical Subversion in Early English Drama*. New York: Peter Lang.

Houston P, Floyd M and Carnicero S. (2012) *Spy the Lie*. New York: St Martin's Press.

Luke TJ (2019) Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science*, 14(4): 646-671.

Maley W and Tudeau-Clayton M (2010) *This England, That Shakespeare: New Angles on Englishness and the Bard*. Ashgate: Farnham.

Marelj J (2019) *Shakespearean Character: Language in Performance*. London and New York: The Arden Shakespeare, Bloomsbury Publishing.

McIntyre D (2010) Dialogue and characterization in Quentin Tarantino's Reservoir Dogs: A corpus stylistic analysis. In McIntyre D and Busse B (eds) *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 162–182.

McIntyre D and Archer D (2010) A corpus-based approach to mind style. *Journal of Literary Semantics* 39(2): 167-182.

McQuaid SM, Woodworth MT, Hutton EL, Porter S and ten Brinke L (2015) Automated insights: verbal cues to deception in real-life highstakes lies. *Psychology, Crime and Law*, 21(7): 617-631.

Murphy S (2019) Shakespeare and his contemporaries: Designing a genre classification scheme for Early English Books Online 1560-1640. *ICAME Journal* 43(1): 59-82.

Newman ML, Pennebaker JW, Berry DS and Richards JM (2003) Lying words: predicting deception from linguistic styles. *Personality and Social Psychology Bulletin* 29(5): 665–675.

Pennebaker JW (2011) *The Secret Life of Pronouns: What Our Words Say about Us*. New York: Bloomsbury Press.

Phillips M (1989) *Lexical Structure of Text*. Birmingham: University of Birmingham.

Picornell I (2013) Cues to deception in a textual narrative context: lying in written witness statements. PhD Thesis, Aston University, UK.

Quennell P and Johnson H (2013) *Who's Who in Shakespeare*. London and New York: Routledge.

Schafer J (2007) Grammatical Differences between Truthful and Deceptive Written Narratives. PhD Thesis. Fielding Graduate University, USA.

Sorlin S (2016) *Language and Manipulating in House of Cards: A Pragma-Stylistic Perspective*. London: Palgrave MacMillan.

Suchet D (1989) Iago in Othello. In: Jackson R and Smallwood R (eds) *Players of Shakespeare 2: Further Essays in Shakespearean Performance by Players with the Royal Shakespeare Company*. Cambridge: Cambridge University Press, pp.178-199.

Tribble C (2000) Genres, keywords, teaching: Towards a pedagogic account of the language of project proposals. In: Burnard L and McEnery T (eds) *Rethinking language pedagogy from a corpus perspective*. Frankfurt am Main: Peter Lang, pp. 75-90.

Vrij A (2000) *Detecting Lies and Deceit: The psychology of lying and the implications for professional practice*. Chichester: John Wiley & Sons.

Wright Whelan CW, Wagstaff GF and Wheatcroft JM (2013) Highstakes lies: verbal and nonverbal cues to deception in public appeals for help with missing or murdered relatives. *Psychiatry, Psychology and Law* 21(4): 523–537.

Zuckerman M, DePaulo BM, and Rosenthal R (1981) Verbal and nonverbal communication of deception. In: Berkowitz L (ed) *Advances in Experimental Social Psychology, Vol. 14*. New York: Academic Press, pp. 1-59.

Bionotes

Dawn Archer is Professor of Pragmatics and Corpus Linguistics. Her publications include ‘Love – ‘a familiar of a devil’? An Exploration of Key Domains in Shakespeare’s Comedies and Tragedies’ (with Culpeper and Rayson), in the collection, *What’s in a word-list? Investigating word frequency and keyword extraction* (2009). Within the *Encyclopedia of Shakespeare’s Language* project, she is lead editor for volumes focussing on “keywords” in plays, characters and genres.

Mathew Gillings is an Assistant Professor. His research interests are in all aspects of corpus linguistics, most recently applying the method to the study of deception, Shakespeare’s language, and politeness. He is the administrator and communications officer for the *Encyclopedia of Shakespeare’s Language* project.

¹ Those wishing to read character summaries can find supplementary material here: <http://wp.lancs.ac.uk/shakespearelang/publications/>

² More information about the Encyclopedia of Shakespeare’s Language project, including the various tools, datasets, and resultant publications, can be found on the project website: <http://wp.lancs.ac.uk/shakespearelang/>

³ ‘Relevant’, here, meant ‘appear[ing] to point to inconsistencies with respect to...the story’ the speaker was ‘trying to convey, their apparent baseline and context’ (Archer and Lansley 2015: 231).

⁴ Othello and Emilia appear (alongside Iago) in *Othello*. Humphrey appears as the Duke of Gloucester in *2 Henry VI*. Brutus appears in *Julius Caesar* and Katherine of Aragon appears in *Henry VIII*.

⁵ Note, however, that Humphrey’s normalised figure of 162.2 instances per 4,000 words is lower than three of the deceptive characters, namely, Tamora, Iago and Falstaff.

⁶ One thing we might deduce from such findings is that, if/when the above are used to predict deception, it might be more profitable to look for them separately (instead of as a set of collective features that can be assessed based on higher frequencies alone).

⁷ Collectively, ‘deceptive’ characters used negation 50.88 times and *think* 5.92 per 4,000 words. In comparison, ‘plain’ characters used them 52.06 times and 9.98 times respectively.

⁸ The argument for paying ‘especially close attention to [a] person’s cross-channel behaviour immediately’ after some sort of stimulus, such as a question, ‘(i.e., up to seven seconds)’ (Archer and Lansley, 2015: 232) is based on an understanding of being able to ‘reasonably conclude’ such ‘behaviour is directly associated with the stimulus’ (Houston et al., 2012:30).

⁹ The eleven statistical collocates of ‘I’ have LL scores of at least 15.2 and LR scores ranging between 2.226 and 7.436.