



**Manchester
Metropolitan
University**

Murphy, Sean and Archer, Dawn and Demmen, Jane (2020) Mapping the links between gender, status and genre in Shakespeare's plays. *Language and Literature*, 29 (3). pp. 223-245. ISSN 0963-9470

Downloaded from: <https://e-space.mmu.ac.uk/625159/>

Publisher: SAGE Publications

DOI: <https://doi.org/10.1177/0963947020949438>

Please cite the published version

<https://e-space.mmu.ac.uk>

Mapping the links between gender, status and genre in Shakespeare's plays

Sean Murphy^{1,2,3}, Dawn Archer^{1,2} and Jane Demmen¹

¹Lancaster University, ²Manchester Metropolitan University and ³ University of Barcelona

Sean Murphy

Lancaster University, Manchester Metropolitan University and University of Barcelona

Department of Linguistics and English Language

County South, Lancaster University

Lancaster

LA1 4YL

s.e.murphy@lancaster.ac.uk

Department of Languages, Information & Communications

Manchester Metropolitan University

Manchester

M15 6LL

Sean.Murphy@mmu.ac.uk

Educació Lingüística i Literària i de Didàctica de les Ciències Experimentals i de la Matemàtica

University of Barcelona

Barcelona

08035

smurphy@ub.edu

Dawn Archer

Manchester Metropolitan University and Lancaster University

436, Geoffrey Manton Building, Manchester Campus

Rosamond St West

Manchester

M15 6EB

d.archer@mmu.ac.uk

0161 247 3887

Jane Demmen

Lancaster University

Department of Linguistics and English Language

County South, Lancaster University

Lancaster

LA1 4YL

j.e.demmen1@lancaster.ac.uk

Abstract

The AHRC-funded Encyclopedia of Shakespeare's Language (ESL) project has produced a resource allowing users to explore Shakespeare's plays in a variety of (semi-automatic) ways, via a web-based corpus query processor interface (CQPweb) hosted by Lancaster University. It enables users, for example, to interrogate a corpus of Shakespeare's plays using queries restricted by dramatic genre, gender and/or social status of characters, and to target and explore the language of the plays not only at the word level, but also at the grammatical and semantic levels (by querying part-of-speech or semantic categories). Using keyword techniques, we examine how female and male language varies in general, by social status

(high or low), and by genre (comedy, history and tragedy). Among our findings, we note differences in the use of pronouns and references to male authority (female overuse of ‘I’ and ‘husband’; male overuse of ‘we’ and ‘king’). We also observe that high-status males in comedies (as opposed to histories and tragedies) are characterised by polite requests (‘please you’) and sharp-minded ‘wit’. Despite many similarities between female and male usage of gendered forms of language (‘woman’), male characters alone use terms such as ‘womanish’ in a disparaging way.

Keywords

Early Modern English, Shakespeare, plays, CQPweb, gender, genre, status, rank

1. Introduction

Corpus linguistic approaches to analyse Shakespeare’s plays are well established. Indeed, they range from fine-grained studies focusing on particular characters (e.g., Archer and Bousfield 2010; Culpeper 2009) to analyses of language features across the whole body of plays (e.g. Beatrix Busse 2006 on vocatives; Ulrich Busse 2002 on second-person pronouns). As well as adding some much-needed empirically based findings to a large and long-established body of qualitative literary critical work, these quantitative studies have provided useful insights into the way Shakespeare used language to construct different types of individuals, settings and plots. One area where corpus linguistic research remains under-represented, however, is that of the (quantitative) study of language and gender in Shakespeare’s plays.¹ By *gender* we refer not only to the binary distinction of character representations as biologically female or male but, more importantly, to the socially-constructed linguistic characteristics associated with femininity or masculinity (e.g. ‘womanish’). As Culpeper (2001: 12) states, such characteristics are used by people to make ‘sense of others’, thus providing us with insightful commentary on women’s societal roles (whether in a particular place or at a particular time). The period in which Shakespeare was writing was underpinned by social hierarchy and patriarchy (Findlay 1999: 127-163; Nevalainen and Brunberg 2003: 32–38). Women had much less socio-economic power and fewer rights than men, for example in marriage choices and the ownership of property, and they were excluded from the spheres of politics and the law. There was also great inequality between people born into the upper social ranks and those at the lower end, and little or no social mobility. Consequently, gender is cross-cut by social status, and this can be observed in the characters constructed in drama of the period. In her dictionary of women in Shakespeare’s plays, Findlay (2010) details the ways in which characters are gendered by rank title, kinship and social roles. ‘Lady’ denotes a female of a particular social rank. Age can be a factor too. ‘Woman’, ‘lady’ and ‘girl’ are all descriptors identifying someone as female. ‘Woman’ and ‘lady’ imply a female of adult age, whereas ‘girl’ implies a relatively young female, with connotations that might include youth, beauty, inexperience, vulnerability, and so on (see Archer and Gillings, this volume). Although some of the above are evidenced through lists of *dramatis personae* in critical editions, and/or in the narrative of plays, many of the characters in (publicly-accessible) Shakespeare corpora or, indeed, comparative corpus data more generally are not formally categorised according to rank, kinship or social role. This helps to explain why there has not been a comprehensive,

comparative corpus linguistic study of language and gender in Shakespeare until now. There is thus the need for resources like those reported in this paper, provided by the AHRC-funded Encyclopedia of Shakespeare's Language (ESL) project.

The project resources (explained in detail in Section 2.1) allow users to explore Shakespeare's plays in a variety of (semi-automatic) ways, via a web-based corpus query processor interface (CQPweb) hosted by Lancaster University (detailed in Section 3.1). The project's corpora are particularly useful when it comes to empirically based, gender-focused studies, given the inclusion of information on gender - as well as social rank - for each character. Indeed, such additions enable quantitative-based investigation to determine the kind(s) of language female and male Shakespearean characters overuse (statistically speaking).

Following Culpeper (2001) and Findlay (2010), we assume that the language used by female and male characters - in Early Modern English² plays - can help us to understand what constituted usual and appropriate social behaviour for an Early Modern audience (and what did not). We are interested in instances when the playwright's language usage coincided with societal expectations about women and men for his contemporaries, and when it served to flout audiences' assumptions for dramatic effect (because it challenged norms). We thus pay attention to the ways in which Shakespeare reinforced his characters' identities through descriptions/evaluations of gender and social rank, and the connotations that arose in consequence. We draw upon the corpus linguistic concept of keywords to do so (following Culpeper 2009). Keywords occur in a word list with relatively high or low statistical frequency and are generated by comparing one word list with another (see Section 3.2). We examine the keywords characterising (all) female and male speech across the 38 plays. We then broaden our scope by examining how female and male language varies according to characters' social status (high or low), and in different genres (comedy, history and tragedy). Building on existing research utilising semantic category analysis (e.g. Archer et al. 2009), we focus, in particular, on gendered forms of language ('girl', 'woman', 'lady', etc.) and analyse collocational patterns to reveal differences in use by female and male characters. We discuss our results in detail in Sections 4.1 to 4.4, following both our explanation of the ESL project/corpora used in this study (see Section 2), and our method of extracting results from these ESL resources (see Section 3).

2. Background

2.1. The *Encyclopedia of Shakespeare's Language* project

The ESL project (2016-2019)³ aims to provide new, empirically-based insights into Shakespeare's plays, notably by examining Shakespeare's language in the context of language used in plays by other contemporaneous playwrights and in wider Early Modern English of the same period. Three corpora comprise the Enhanced Shakespearean Corpus (ESC). The first is the ESC: First Folio Plus ('ESC: Folio'; Culpeper et al. in preparation). It includes 38 play-texts⁴ by William Shakespeare: the 36 plays in the First Folio of 1623, sourced from Internet Shakespeare Editions,⁵ plus the quartos of *Pericles* and *The Two Noble Kinsmen* (both of which have a longstanding presence in the Shakespeare canon).⁶ The second corpus is the ESC: Comparative Plays ('ESC: Comp'; Demmen, 2020). It comprises 46 play-texts by 24 playwrights approximately contemporaneous with Shakespeare, with approximately similar proportions of comedy, history and tragedy. The third corpus is the

ESC: EEBO-TCP Segment ('ESC: EEBO'; Murphy, 2019). It comprises the digitised texts from the Early English Books Online - Text Creation Partnership⁷ from the period 1560-1640, apart from those contained in the ESC: Folio and the ESC: Comp. These corpora are publicly available for scholars, students, schools, theatre groups, actors and anyone with an interest in Shakespeare's plays, other drama and the English language of the period.

2.2. Annotation and normalisation

The ESC texts are marked up and annotated with XML tags (see Bray et al. 2006; Hardie 2014a). Each utterance is marked with an opening speaker ID tag and a close tag. Act and scene boundaries, stage directions, front matter, end matter and paratext (e.g. prologues and epilogues) are also marked with XML tags.

Spelling normalisation (regularisation) has been applied to the ESC texts. Early Modern English spelling variation has been normalised using VARiant Detector spelling normalisation software (VARD 2; Baron and Rayson 2008). In the ESC: Folio this was carried out manually on a word-by-word basis. Due to time constraints, in the ESC: Comp and ESC: EEBO it was carried out automatically, the software having been trained using manual regularisations made in the ESC: Folio. Spelling normalisation is designed to improve the usability of the play-texts with corpus linguistic software tools, many of which identify results by the orthographic matching of word-forms (see Demmen, 2020). The ESC texts have also been annotated with grammatical part-of-speech tags using a customised version of the Constituent Likelihood Automatic Word-tagging System (CLAWS; see Leech et al 1994). CLAWS tags are alphanumeric codes in square brackets which correspond to over 200 part-of-speech classifications (CLAWS tagset version 6 was used; see <http://ucrel.lancs.ac.uk/claws6tags.html>). For example, [JJ] denotes an adjective, [NN] a noun and [VV] a verb. In the ESC Folio, every word was manually checked for accuracy at the highest level of the tag (e.g. a word tagged NN1 (singular common noun, e.g. girl) and another NN2 (plural common noun, e.g. girls) were both checked to ensure that the initial 'N' (noun) tagging was correct). Again, due to time constraints, tagging of the ESC: Comp and ESC: EEBO was carried out automatically using a version trained on the ESC: Folio data.

Semantic annotation of the ESC texts was conducted using the UCREL⁸ Semantic Analysis System (USAS; Rayson et al., 2004) in the Wmatrix suite of corpus linguistic software tools (Rayson, 2008). USAS assigns a semantic category label in the form of an alphanumeric tag to each word, using a taxonomy of 232 categories of meaning grouped into 21 main semantic fields⁹. This is discussed further in Section 3.3.

The play-texts in the ESC: Folio and the ESC: Comp were annotated with XML tags for 'social' categories (i.e., whether speaking on their own or as part of a group, their gender, and their social rank) using the taxonomy shown below in Table 1.

Table 1. Character categories used in the ESL project

Category	Possible values
Speaker(s)	Singular (s) or multiple (m)

Gender of speaker	Male (m), female (f), assumed male (am), assumed female (af), problematic (p)
Status/social rank of speaker	Monarch (0), Nobility (1), Gentry (2), Professional (3), Other Middling Groups (4), Ordinary Commoners (5), Lowest Groups (6), Supernatural Beings (7), Problematic (p)

The possible values for gender include an assumed gender identity to account for characters who disguise themselves as a member of a different gender to that of their main character. The categories relating to a character's status/social rank mainly draw upon the scheme developed by Archer and Culpeper (2003), and reflect the nature of status in pre-industrialised Early Modern English society and the way in which Shakespeare's contemporaries wrote about it. That scheme has been slightly adapted, however, in order to capture particular Shakespearean features, in particular, a *Supernatural Beings* category to account for ghosts, gods, fairies, etc. (see Murphy, 2017).

3. Method

3.1. CQPweb

The ESC corpora are accessed through a web-based Corpus Query Processor interface (CQPweb) designed by Andrew Hardie and hosted by Lancaster University¹⁰ (which is publicly accessible via the Resources tab on the ESL project website). The corpus data can be interrogated using 'standard' or 'restricted' queries, using Simple Query Language¹¹ (SQL) or CQP syntax.¹² A standard query interrogates all the data in the ESC: Folio; restricted queries facilitate searches on particular Sections and/or variables, including:

- pre- and post-1600, allowing the user to compare results from early versus late plays;
- choice of dramatic sub-genres (comedy, history and tragedy, with or without a fourth category of 'problem' plays);
- selection of one or more individual plays-texts;
- play-texts in the ESC: Folio (see Section 2.1) or in a separate Quarto dataset which includes versions of Shakespeare's plays that were published as separate editions prior to the First Folio);
- gender of speaking characters (male, female, unclear, or assumed gender identity; see Section 2.2);
- social status categories (see Section 2.2).

For our investigation, we restricted the corpus to the ESC: Folio, and examined the female-only and male-only speaker data. We looked at the data across all dramatic sub-genres, and examined findings in each. We also performed further searches restricted by high and low social status groupings.

3.2. Keyword analysis and (sub)corpora creation

Keyword analysis is a well-established technique in corpus linguistics for identifying words that indicate style (see, e.g. Baker 2004; Culpeper and Demmen 2015). As noted in our

introduction, we generated keyword lists by comparing word frequencies in a target wordlist against a reference wordlist, and applied statistical measures. This paper uses both Log-Likelihood (LL) and Log Ratio (LR) as statistical measures¹³. LL was set at 6.63 or above (99% confidence level). LR was then used to sort the keyword list so that the quantitatively largest differences were at the top of the list (with each increase by 1 indicating a *doubling* of how many times *more* frequent the word was in the target wordlist when compared with the reference wordlist). This allowed us to gauge not only the amount of evidence we had for the existence of an effect (LL), but also the size of that effect (LR). The significance cut-off point was 0.01%, using Šidák correction, with a minimum frequency of 3 in each frequency list.

Keyword studies of Shakespeare’s plays include Culpeper’s (2009) and Archer and Bousfield’s (2010) comparisons of the language styles of characters in single plays. Murphy (2015) has profiled soliloquy dialogue relative to interactional dialogue, and Scott and Tribble (2006) have analysed language features based on all of Shakespeare’s plays. The ESC: First Folio Plus in CQPweb not only facilitates such comparisons, but also allows researchers to create corpora to meet their own research needs. Indeed, this is the approach taken in this paper. Our first step involved taking the ESC: First Folio Plus corpus and creating appropriate sub-corpora. The creation of subcorpora in CQPweb is a simple procedure that involves selecting the ‘Create/edit subcorpora’ option from the left-hand menu, defining a ‘new subcorpus via corpus metadata’, naming the new subcorpus, and selecting the ‘restrictions for your subcorpus’. We named one subcorpus ‘female’, and selected the following text-type restriction: ‘Source – First Folio’; and the following restriction on ‘Speaker ID’: ‘Sex – f’. This enabled us to ‘Create [the] subcorpus from selected categories’. On the next screen, it is important to click on ‘Compile’ in the ‘Frequency list’ column. For the purposes of this study, we created subcorpora based on the variables of gender, genre and social status (SS).¹⁴ These subcorpora are shown in Table 2 (social status categories ‘7’ (Supernatural) and ‘p’ (problematic) were not included as they are not gender-specific).

Table 2. Basic subcorpora of ESC: First Folio Plus created for keyword analysis

Gender	Genre	Social status (see Table 1 for social status categories)
Female Male	Comedy_Female	Female_status_0 Male_status_0
	Comedy_Male	Female_status_1 Male_status_1 HIGH (0-2)
	History_Female	Female_status_2 Male_status_2
	History_Male	Female_status_3 Male_status_3
	Tragedy_Female	Female_status_4 Male_status_4
	Tragedy_Male	Female_status_5 Male_status_5 LOW (3-6)
		Female_status_6 Male_status_6

We also created other subcorpora that combined all three variables, for example: Comedy_Female_HighSS; Comedy_Female_LowSS; HistoryTragedy_Female_HighSS; HistoryTragedy_Female_LowSS. These allowed us to further refine our investigations.

3.3. Semantic category analysis

Archer et al. (2009) show that semantic field or ‘domain’ analysis is of value in exploring the concept of love in Shakespeare’s comedies and tragedies. They use the USAS tool in Wmatrix, mentioned in Section 2.2, which assigns a semantic category label (in the form of an alphanumeric tag) to each word, using a taxonomy of 232 categories of meaning grouped into 21 main semantic fields (see further <http://ucrel.lancs.ac.uk/usas/>). As Archer et al. (2009) point out, semantic categorisation is not without problems, especially when used with historical data. The USAS taxonomy was developed for late 20th century English, so some words that tend to occur in Early Modern English will not be in the USAS lexicon, and will not be tagged in consequence. In addition, the meaning or main sense(s) of some words may have changed from the late 16th/early 17th century to the present day, resulting in inaccurate tagging. It is for such reasons that the ESC corpora have undergone spelling normalisation (see Section 2.2.), with a view to improving tagging accuracy (see also Archer and Findlay, in press).

In our study we focus particularly on a single semantic domain ‘People: Female’ and ‘People: Male’ (USAS categories S2.1 and S.2.2, respectively), which fall under the main area of ‘S: Social Actions, States & Processes’. We compare the words identified as belonging to these categories in the dialogue of female and male characters in the Shakespeare corpus. To avoid introducing another variable, we do not include in this study the dialogue of characters who assume a different gender identity by disguise/cross-dressing, though this would be worthwhile in a follow-up study. Our investigation takes in semantic collocates, that is, other semantic areas which co-occur relatively frequently (statistically speaking) with the ‘People: Female’ tag, in the dialogue of female and male characters. We used CQPsyntax to run the following queries, [semtag = ‘S2.1.*’] and [semtag = ‘S2.2.*’] to identify language tagged as belonging to the ‘People: Female’ and ‘People: Male’ semantic domains. Our results are discussed in Section 4.4.

4. Results and discussion

4.1. Gender comparison

We used the ESC: First Folio Plus in CQPweb to create subcorpora for ‘all female’ and ‘all male’ speech and then used the ‘Keyword’ function to compare the two. Table 3 shows positive and negative keywords for females.

Table 3. Keywords for females

+ (23 forms significantly overused)	- (6 forms significantly underused)
-------------------------------------	-------------------------------------

Pisano, Nerissa, Willow, les, Lysander, Malvolio, Je, Husband, Romeo, Nurse, alas, sister, prithee, Oh, O, pray, Love, Me, ?, you, my, not, I	king, we, our, the, of, and
---	-----------------------------

Two particularly noteworthy findings, relating to the 23 positive key items, are that three keywords, in particular ('alas', 'Oh', 'O') were representative of emotional states to do with sorrow or grief, and the most frequent collocate of 'alas' was 'poor'.¹⁵ These findings tie in with Culpeper and Oliver's (in press) ongoing ESL project research into pragmatic noise. They have found, for example, that, although both female and male characters tend to overuse pragmatic markers when each are compared to the other, they use them differently (Culpeper and Oliver, in press). Female characters tend to draw on pragmatic markers expressing grief or sorrow, as above. Male characters' pragmatic markers, in contrast, reveal them to be the architects of events. This latter result fits with the nature of the historical period we are dealing with, underwritten as it was by social hierarchy and patriarchy (Nevalainen and Brunberg, 2003: 32–38). This explains, in turn, the positive keyness of 'husband' and the negative keyness of 'king' in women's talk (see Table 3), if we assume that the former signals female characters' focus upon local (and arguably more intimate) figures of power and authority, and the latter signals male characters' focus on the national figure (the male monarch). As Section 4.2 will reveal, most instances of 'husband' are voiced by women of middling to high social rank. The kinship term 'sister' is also key in female dialogue, suggesting women talk to and about close female relations as well as about husbands ('sister' and 'husband' are both terms of address and terms of reference). Such findings suggest female characters' senses of identity are often socially constructed on the basis of family relations. Collocational analysis of females' statistically significant overuse of 'you' reveals a high number of instances of polite formulas such as '(I) pray you (93), 'Please you' (30), 'I (do) beseech you (28) and '(I) thank you' (26), which are discussed below. When considering the 38 plays as a whole, females do not statistically overuse or underuse 'thou' forms in comparison with males. The keyness of 'thou' forms in history plays by females is discussed in Section 4.2.2.

Additional noteworthy findings are female characters' statistical underuse of the plural form 'we' but overuse of other first-person singular pronouns ('I', 'me' and 'my'), as well as politeness markers ('prithee' and 'pray'). An overuse of 'I' and 'me' suggests that female characters are engaging in self-referentiality as part of their relational work. Their marked underuse of 'we' (females – 2,570 instances per million words; males – 4,013 instances per million words) can be accounted for by greater use of the royal 'we' by male monarchs and other characters with monarch-like status (e.g. Duke Vincentio in *Measure for Measure*). Statistical analysis supports this claim, as can be seen in Table 4.

Table 4. Instances of 'we' by female and male characters (real frequency / relative frequency per million words)

Social status	Female	Male
0 Monarchy	114 / 3,048	738 / 5,809

1 Nobility	168 / 2,263	1,311 / 3,972
2 Gentry	53 / 2,117	550 / 3,148
3 Professional	34 / 3,718	101 / 3,016
4 Middling	30 / 3,375	111 / 4,324
5 Commoners	14 / 3,222	161 / 5,168
6 Lowest	7 / 2,317	225 / 4,694

Although male nobles have the highest real frequency of ‘we’ (1,311 instances), male monarchs have the highest relative frequency (5,809 instances per million words) of all social groups. The second highest relative frequency is for commoners (5,168 instances per million words), particularly in plays such as *Henry VI Part 3*, *Coriolanus*, *Julius Caesar*, *A Midsummer Night’s Dream* and *Pericles*, in which ordinary citizens and tradespeople act as collectives. The only instance in which females’ relative use of ‘we’ is higher than males’ is for professionals (3,718 instances per million words), and is mostly accounted for by Mistresses Ford and Page, the wives of well-to-do citizens, in *The Merry Wives of Windsor*, who act in tandem to ridicule Falstaff. A final point to note with regard to the higher relative frequency of ‘we’ among males may concern the factional nature of some male talk, particularly in history plays, as male characters represent or identify with groups. For example, Henry V tells the French Herald, Mountjoy ‘We would not seek a Battle as we are, / Nor as we are, we say we will not shun it.’ (3.6). His use of ‘we’ befits his kingly status, yet he is also speaking on behalf of his army. In contrast, female characters self-identify as belonging to a group or speaking on behalf of others less than their male counterparts, in line with the reality that women in the Early Modern period were excluded from the kinds of power politics exhibited by Henry V.

The overuse of ‘pray’ aligns with Lutzky and Demmen’s (2010) finding that male characters use the term about 25% less than female characters in their multi-authored Early Modern English comedy play-text samples (published between 1560 and 1760). Lutzky and Demmen suggest its use is triggered not by the gender of the speaker but the addressee, based on their finding that female (as well as male) characters use ‘pray’ forms to male addressees more than female addressees. An explanation for why ‘pray’ and ‘prithee’ are relatively over-used among our female characters may be simply because there are many more male than female characters in the plays. Hence, women are more likely to be speaking to men than to other women in our dataset.¹⁶ ‘Pray’ is also associated with the making of polite requests (Culpeper and Archer, 2008: 74-76), suggesting another reason for its relative overuse in our results might be due to women making more requests than men (relatively speaking).¹⁷

Readers will have noted that question marks were also found to be ‘key’ in the female dialogue dataset (see Table 3). This punctuation marker is clearly not a word form (as keywords tend to be), but nonetheless has a relative frequency of 13,339 per million words (compared to 10,841, for male characters). It still needs to be treated with some caution, however, as some instances may be compositorial choices. It is possible, too, that some questions in the plays may not be followed by question marks (as they would be if the plays were modern texts). Even so, the dispersion overview of its use among female characters provides us with some interesting findings (see Figure 1).

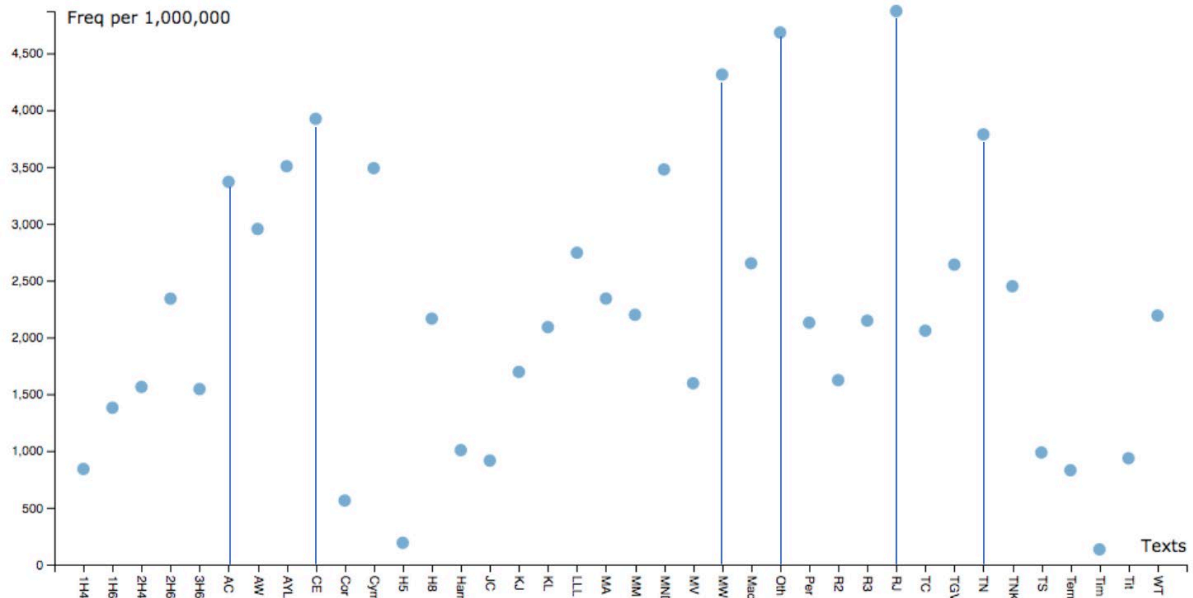


Figure 1. Dispersion overview of the normalised frequency of question marks in female speech in 38 Shakespeare plays

Note, for example, that frequencies per million words are particularly high in certain tragedies (*Romeo and Juliet*, *Othello* and *Antony and Cleopatra*), and many comedies, especially *The Merry Wives of Windsor*, *The Comedy of Errors* and *Twelfth Night* (dots representing these plays are connected by lines to the play abbreviation for ease of reference). Conversely, low relative frequencies of ‘?’ are mainly found in histories, where ‘[w]omen, who live to be the grievous survivors of wars men make and die in, stand in Shakespeare’s history plays for permanence and fidelity against shifting political sands’ but, with the notable exception of Queen Margaret and Joan La Pucelle in the first tetralogy, are essentially ‘impotent’ (Dusinberre 1996: 294, 297).

Our results suggest, further, that four characters – Juliet, Desdemona, Cleopatra and Viola – seem particularly prone to questioning. That they are speaking at all makes them more powerful than the silenced female character who was modelled in texts of the period and idealised in the patriarchal society of the time (Findlay 1999: 114). Their questions nonetheless reveal something about each of them specifically. Juliet’s are indicative of her anxious state of mind, for example. Hence turns such as ‘O thinkst thou we shall ever meet again?’ (3.5) when wondering about a possible future with Romeo. Culpeper’s (2009) analysis of Juliet’s specific keywords (i.e., ‘if’, ‘yet’, ‘be’, ‘would’), revealing of a syntactic style that helps to articulate her anxieties, provides us with further support for such a characterisation. Desdemona’s questioning in *Othello* is more assertive and challenging. She first teases Iago in 2.1 and then pursues Cassio’s return to favour on his behalf in 3.3. Indeed, she bombards Othello vociferously with seven questions in 5 lines: ‘Shall’t be shortly?... Shall’t be tonight at supper?... Tomorrow dinner then?... Why then tomorrow night? Or Tuesday morn? / Or Tuesday noon or night? On Wednesday morn?’.

becomes evidence of her infidelity and he plots to kill her. As he prepares to do so, Desdemona continues to ask questions, but this time they challenge Othello on the reason for and justice of his strangely emotional behaviour towards her:

Alas the heavy day: why do you weep?
 Am I the motive of these tears my Lord?
 ... I fear you ... when your eyes roll so.
 Why I should fear, I know not,
 Since guiltiness I know not: But yet I feel I fear
 [...]
 Alas, why gnaw you so your nether lip?
 Some bloody passion shakes your very Frame
 These are portents: but yet I hope, I hope,
 They do not point on me (*Othello*, 5.2).

Viola and Cleopatra are likewise strong characters. Viola uses questions, in *Twelfth Night*, to relocate herself as a stranger in the geographical and emotional world of Illyria and manages (by so doing) to position herself in the very hearts of its two ruling figures. Cleopatra is a ruler, in *Antony and Cleopatra*, albeit one who fails to rule effectively, to the point of being portrayed as a ‘despotic figure’ (Kemp 2010: 97). Her questions are thus a curious mixture of assertive demands for information (‘Wherefore’s this noise?’ (5.2)) and insecurity and jealousy about Antony’s whereabouts and his love for her (‘Why is my Lord enraged against his Love?’ (4.12)).

4.2. Social status

Table 5 shows key language forms and their most frequent collocates, compared by gender and social status.

Table 5. Key language forms (in bold) with common collocates, compared by gender and social status

Social status	Female	Male
Monarch (0)	my good lord	we ’ll/we will/we shall
Nobility (1)	my husband I pr ithee, I pr ay you O true love	my lords the king we ’ll/we will/we shall our swords/hands/hearts the (noun) of (noun) and (esp. after a comma)

Gentry (2)	my husband alas	the (noun) of (noun)
Professional (3)	my husband	-
Other Middling Groups (4)	<i>parenthetical comment</i> : (good heart), (says he)	the (noun) of (noun)
Ordinary Commoners (5)	-	-
Lowest Groups (6)	hail Macbeth thou art/hadst/shalt	my (good) lord

These findings both confirm and amplify the points made in Section 4.1. We see females of the highest status showing deference to males. Higher status (0 and 1) males' use of 'we' may be royal or collective and often occurs with a future-oriented auxiliary implying decision or intention expressed with 'we'll', 'we will' and 'we shall'. These findings correlate more with histories and tragedies than with comedies as characters affirm their group identity: 'we'll spill the blood', says Antony, speaking of his forces (*Antony and Cleopatra*, 4.8). The keyness of 'our swords', 'our hands', and 'our hearts' for the nobility is most evident in tragedy, and particularly *Julius Caesar*: 'let us bathe our hands in Caesar's blood / Up to the Elbows, and besmear our Swords (3.1)', says Brutus.

For high and middle-status women, 'my husband' is a common point of reference, with polite requests ('prithee' and 'pray') occurring most amongst the nobility. There appear to be two important stylistic distinctions as regards phrase and sentence structure between the genders. Male characters from the nobility to middling groups appear to use more structures based around 'the (noun) of (noun)' including 'the Duke of York', 'the name of God' and 'the House of Lancaster'. This arguably represents a 'report' speech style which focuses primarily on the exchange of information between participants (Tannen 2005). In addition, male nobles significantly overuse 'and', most often after a comma: 'He is already named, and gone to Scone (*Macbeth*, 2.4), 'whom they doted on, / And blessed, and graced, and did more than the King' (*Henry IV, Part 2*, 4.1). This is often referred to as 'narrative AND' (Culpeper and Kytö 2010: 173-5). Male nobles play an important role in constructing the narrative and are therefore characterised as having an additive style, consistent with reporting. Similarly, it could be argued that female keywords and their collocates are consistent with a 'rapport' style of speech (Tannen 2005) emphasising the building and maintaining of social relations. Hence, we find features such as polite requests ('I prithee'), expressions of emotion ('alas'), including surge features ('O'), references to relationships ('my husband') and parenthetical comments ('good heart'). We might note, here, Mistress Quickly's style as she talks to Falstaff about Mistress Ford in the *Merry Wives of Windsor*: 'Alas the day, (good heart) that was not her / fault' (3.5). By using 'Alas', a keyword for female gentry (status 2), Mistress Quickly may be imitating a higher-status speech style than her own (status 4). This said, 'good heart' is key for her status, suggesting a mixing of styles as Mistress Quickly expresses sympathy for Mistress Ford. Such style mixing would be consistent with other aspects of mixing in Mistress Quickly's speech, including malapropisms, as well as being a contributing factor to her comic presence.

4.3. Genre

The freedom to create subcorpora in CQPweb allows us to explore the intersections between genre, gender and social status, something not previously attempted. The results can then be viewed as graphical wordclouds (colour or monochrome). Our explorations involved comparisons among the 24 subcorpora discussed in Section 3.2. Although we can only present a selection of our findings here, we aim to show the types of comparisons that can be made. As Table 6 indicates, sections 4.3.1 to 4.3.3 describe the results of three comparisons, each with one of the variables of genre, status and gender as the independent variable, with the other two as dependent variables. The particular variables for each comparison are selected on the basis of what we consider to be notable results, which is not to say that other comparisons did not also produce noteworthy findings. For this reason, we would encourage readers to create their own subcorpora and conduct investigations according to their particular interests.

Table 6. Independent and dependent variables (genre, status, gender) compared

Section	Independent variable: choice of variable	Dependent variables: choices of variables
4.2.1	Genre: Comedy	status/gender: high-status/males
4.2.2	Status: high-low	gender/genre: female/History
4.2.3	Gender: female-male	genre/status: Tragedy/high-status

4.3.1. Comedy

Figure 2 presents the keywords for high-status males in comedies (when compared with characters of the same status in histories and tragedies).

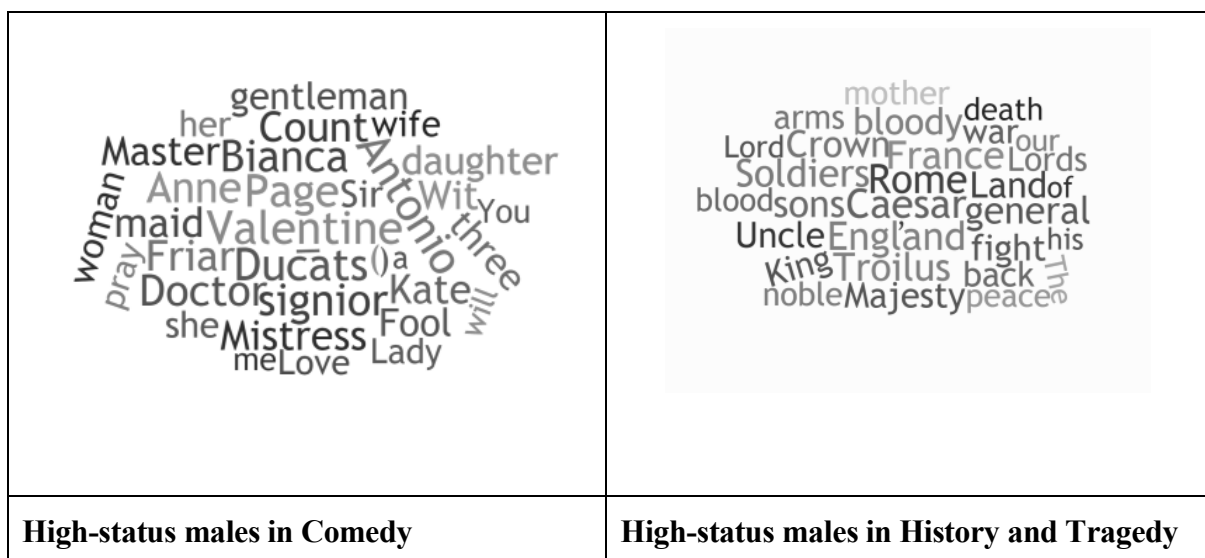


Figure 2. Language forms of high-status males compared according to genre

High-ranking male characters in comedy clearly talk significantly more about women than their counterparts in histories and tragedies, evidenced by keywords such as ‘woman’, ‘maid’, ‘her’, ‘wife’, ‘daughter’, ‘lady’, ‘Mistress’, ‘Kate’, ‘Bianca’ and ‘Anne’. These results, together with the keyword ‘love’, are predictable, and in line with the findings of Archer et al. (2009). Male characters reference and address other men frequently, key forms being ‘gentleman’, ‘Sir’, ‘Count’, ‘Master’, ‘Friar’, ‘Doctor’, ‘Signior’, ‘Antonio’ and ‘Valentine’. The keyness of ‘pray’ contrasts with the findings in Section 4.1, which indicated ‘pray’ was used more by female than male characters overall in the plays (relatively speaking) and, moreover, was triggered by the addressee’s (rather than the speaker’s) gender. It may therefore be that the use of ‘pray’ is related to its use in speech acts, which tend to occur more in the dialogue of comedy plays, notably requests (see Culpeper and Archer 2008), and that such requests are made more by high-ranking male characters in the form ‘I pray thee/you’ (accounting for 31 of the 112 instances): ‘I pray thee stay?’ says Troilus to Ulysses (*Troilus and Cressida*, 5.2). Among this group of characters, it is also common to find similar polite requests such as ‘please you’ (38) (‘Please you read’), ‘(I) beseech you’ (29) (‘I beseech you Sir, pardon me’), and ‘(I) entreat you’ (13) (‘I entreat you with me home to dinner’). It would seem, then, that high-status males in comedies appear well-bred because of their use of politeness. Another notable feature of high-status, comedy male characters’ language is the statistical overuse of ‘wit’ (collocation, ‘thy wit’), emphasising the importance of the demonstration of sharp-mindedness as an element of higher-ranking characters in comedy (‘Sir, your wit ambles well, it goes easily’: *Much Ado About Nothing* 5.1). Corpus methods even reveal the keyness of the indefinite article ‘a’ in this high-status grouping, most commonly in *Love’s Labour’s Lost* (248 instances). For example, Biron (who has just been writing poetry to Rosaline) denies that he would praise (a woman’s) physical attributes: ‘when shall you hear that I will praise a / hand, a foot, a face, an eye: a gait, a state, a brow, a breast, / a waist, a leg, a limb’ (4.3). In linguistic terms, his long list contravenes the Gricean Co-operative Principle of conversation (Grice 1989: 26-7), by flouting the maxim of quantity (providing more information than is required), and the maxim of manner (by not being brief). By flouting these maxims, the dramatist creates the implicature, at least to the audience, that the lord doth protest too much and is actually in love, thereby creating a comic effect. One example should not be generalised to the whole genre, but it does perhaps suggest one of several reasons as to why ‘a’ may be key.

Our focus in this Section has been on Comedy, rather than History and Tragedy. Nevertheless, we might make some brief remarks on these genres on the basis of the keyword evidence in Figure 2. First, high-status males in these genres use different kinship terms (‘mother/sons’) to those in comedy (‘wife/daughter’) as histories and tragedies are more oriented to sovereignty and dynasty (hence also ‘king’, ‘crown’, ‘majesty’). Other major concerns for high-status males are place or territory (‘Rome’, ‘England’, ‘France’, ‘land’) and conflict (‘blood(y)’, ‘death’, ‘war’, ‘fight’). We do not have space to explore each of these themes in greater detail here, but they would be worthy of further investigation.

4.3.2. History

Figure 3 shows keywords for high and low-status females in histories compared with females of the same status in comedies and tragedies.

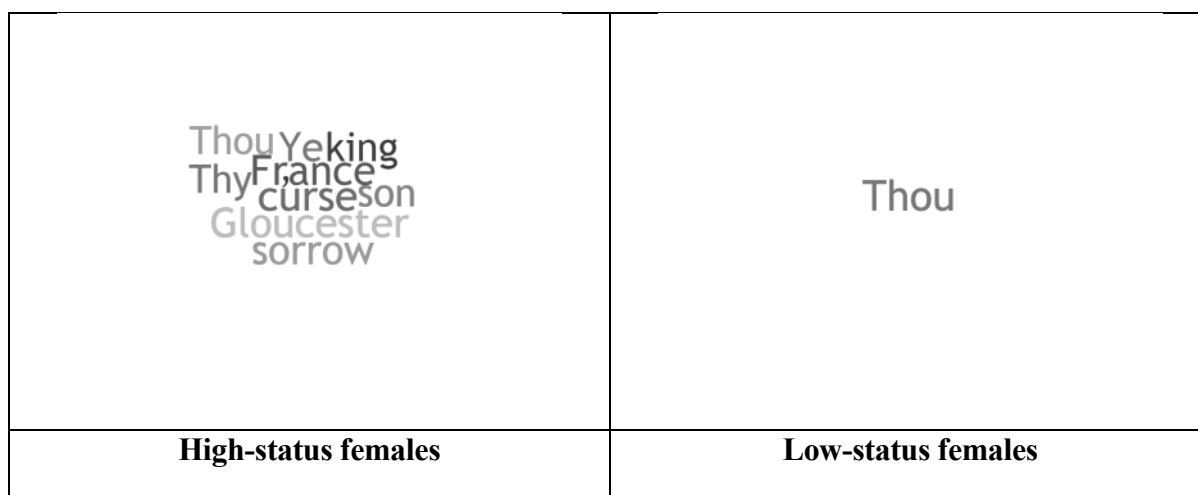


Figure 3. Language forms of females in histories compared on the basis of high or low status

Notice that ‘thou’ is overused by both social ranks, with high-status females using the second person pronoun 288 times (9,949 times per million words) and low-status females using it 53 times (14,593 times per million words).

In the case of the high-status women, the play in which ‘thou’ occurs most often and the female characters with which it is most associated is *Richard III* (Queen Anne, Queen Margaret and Queen Elizabeth). These higher-status women mainly use ‘thou’ deprecatingly towards Richard (note also the keyness of ‘curse’). To a far lesser extent in the same play, ‘thou’ can also signal intimacy. The keyness of ‘thou’ among lower-status women is mostly down to the Hostess (and to a small extent, Falstaff’s lover, Doll Tearsheet) in *Henry IV, Part 1* and *Henry IV, Part 2*. As before, we find ‘thou’ used as both a marker of intimacy and an insult, though in the latter case, with considerably less invective and more humour. Hostess Quickly in *The First Part of Henry IV*, for example, lewdly tells Falstaff (her sometime bed-partner) ‘thou, or any / man knows where to have me, thou knave thou’ after he tells Prince Hal ‘She’s neither fish nor flesh, a man knows not where to have her’ (3.3). The keyness of ‘sorrow’ among higher-ranking women highlights female lamentation and impotency in the face of male power. As noted in Section 4.1, the emotion label is not the only keyword which is statistically relevant, when it comes to female characters. ‘Alas’, ‘Oh’, ‘O’ and ‘poor’ were found to be key too, in line with findings by Culpeper and Oliver (in press) that female characters tend to draw on pragmatic markers to express sorrow, as well as grief.

4.3.3. Tragedy

Figure 4 highlights the keywords for the high-status females and males found in the tragedies (when compared against each other).

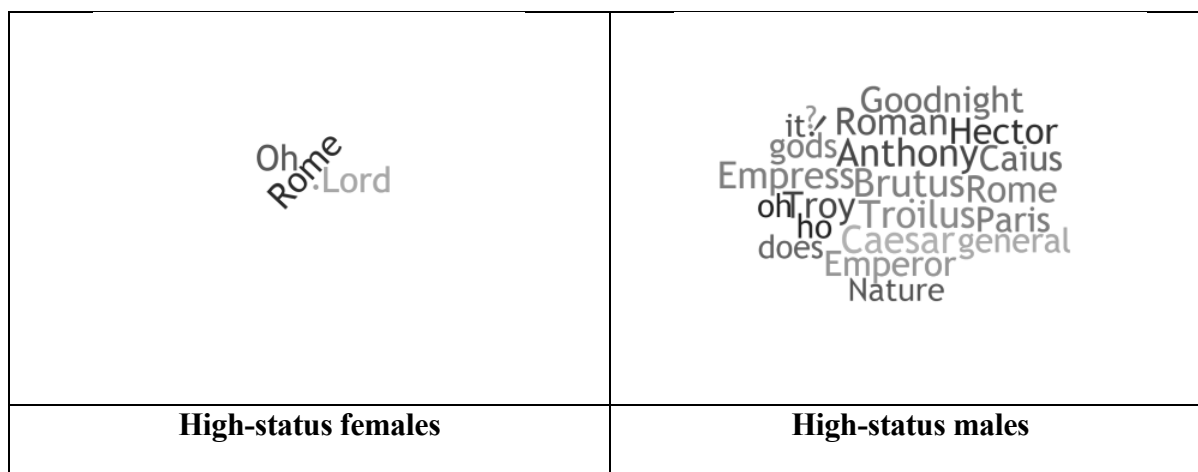


Figure 4. Language forms of high-status characters in tragedies compared according to gender

Notice that high status females use exclamatory ‘Oh’ more than twice as much as their male counterparts, relatively speaking (2,822 instances per million words compared to 1,315 instances per million words; see also Section 4.1). High status females also overuse ‘Lord’, many instances of which collocate with ‘my’ (242, equating to 4,948 instances per million words). This may point to female deference towards male characters, which in itself may not be unusual as there are many more male characters than female, so it may be that women are more often addressing men than other women (in line with our suggestion regarding forms of ‘pray’ in Section 4.1). Keywords for high status males mostly emphasise the importance of other men (especially ‘Mark Antony’, but also ‘Brutus’, ‘Julius Caesar’, ‘Troilus’, and ‘great Hector’). The exception is ‘Empress’ (in *Titus Andronicus*). This powerful female is atypical of Shakespearean female characters, however, having gone from slave to Empress in the first scene and then exacting a bloody and terrible revenge on the man who enslaved her (Titus) throughout the rest of the play.

4.4. Semantic category analysis: use of gendered language by female and male characters

The results presented in Tables 7 and 8 show the similarities and differences between female and male characters’ use of gendered language forms. Note that we have combined singular and plural forms and slight variations in spelling in both tables (see, e.g. ‘madam/e/s’ in Table 7).

Table 7. Gendered language forms concerning women used by females and males in Shakespeare’s plays

Forms used by both female and male characters (instances f/m)	
Beldam/s (1/3)	maidenly (1/1)
female/females (2/16)	maidhood/s (1/1)
girl/s (18/45)	mistress (70/308) ¹⁸
Housewife/ves (6/7)	sisterhood (3/2)
Lady/ladies (141/522)	unmanly (/4)
Lass/es (1/6)	wench/wenches (9/69)
Madam/e/s (130/362)	woman/women (149/398)
Maiden/s (8/18)	Womanhood (4/3)
maidenhood/s (2/1)	womanly (1/2)
Forms used only by female characters	Forms used only by male characters
	crone (1) effeminate (6) feminine (1) hag/s (12) womanish (7) womankind (2)

Table 7 shows that the most frequently-used terms by both men and women about women are ‘lady’ or ‘ladies’, ‘woman’ or ‘women’ and ‘Madam/e’, followed by ‘wench/es’, ‘girl/s’, ‘maiden/s’ and ‘lass/es’. Further investigations show this to be the case across all social ranks of speaker. However, the words most closely associated with these terms are not all the same. For male speakers, the collocates associated with gendered forms concerning women, taken as a group, are ‘sweet’, ‘gracious’, ‘fair’, ‘mad’, ‘gentle’, ‘dear’ and ‘good’. All except ‘mad’ are used in terms of address, though ‘fair’ is often a descriptive term as well as an address form. Collocates of gendered forms concerning women used by female speakers are ‘poor’ and ‘good’, the former being a descriptive term and the latter a term of address. ‘Good’ is a highly frequent term of address used by men and women to women in, for example, ‘Good Madam’, ‘(my) good Lady’, ‘good wench’, and so on. The collocate ‘poor’ occurs in expressions such as ‘Alas poor Lady/Woman/Women’, ‘poor weak woman’ and ‘poor wench’, and fits in with the relative over-use of expressions of sorrow and grief in women’s speech which we noted in the keyword results in Section 4.1. Apart from ‘mad’, the collocates in male speech describing women tend to focus on positive qualities, though this is a general trend in the group of female gendered terms. Some individual words, such as ‘wench’, features some positively-connotated terms (‘good’, ‘sweet’) and some negatively-connotated terms (‘light’, meaning ‘promiscuous’) in male dialogue.

Table 7 also shows that, for the most part female and male characters share the same terms to describe and to address women, as is also the case in their talk about men. We also see that there are no gendered terms about women used exclusively by women, but there are some terms used only by men. These are mainly derogatory. Only male characters address women as ‘hag/s’, as Plantagenet does Pucelle in *Henry VI, Part 1*, calling her a ‘Fell

banning hag' before advising the 'enchantress' to 'hold thy tongue' (5.3). Male characters also use the terms 'effeminate' and 'womanish', typically in pejorative ways. The term 'effeminate' is associated with female qualities and slanted as being undesirable in a man. In *Richard II* (5:3), Bolingbroke disparages his son as a 'young wanton, and effeminate Boy. In *Richard III* (3:7), Buckingham taunts Richard with accusations of 'gentle, kind, effeminate remorse' over his apparent reluctance to seize the throne from his young nephew. On being entreated to relent from his killing purpose, Clarence's murderer in the same play responds, 'Relent? / No: Tis cowardly and womanish.' (1:4). In *Romeo and Juliet*, Friar Laurence compares Romeo unfavourably to a woman using the same term: 'Art thou a man ... / Thy tears are womanish ... / Unseemly woman, in a seeming man,' (3:3).

Common collocates, some examples of which are given in Table 8, also highlight differences between how female and male characters use gendered words.

Table 8. Common collocates of gendered words used by females and males

Collocates (instances f/m)	
wench (13/56)	f: good (2) m: good (4), light (3), sweet (3)
housewife (6/5)	f: Fortune (2) m: Bianca, / A Housewife that by selling her desires (Oth)
unmanly (1/4)	f: Sure he can not / Be so unmanly, as to leave me (TNK) m: deed, customs, grief, melancholy

Notice that only male characters use collocates such as 'light' and 'sweet', some uses of which might suggest wantonness. Consider, for example, Dromio of Syracuse speaking of the Courtesan: 'she is the devil's dam: / And here she comes in the habit of a light wench' (*The Comedy of Errors*, 4.3). Female characters tend to use the conventional metaphor of Fortune as a housewife at a spinning wheel, as in Celia's 'Let us sit and mock the good housewife Fortune / from her wheel' (*As You Like It*, 1.2). Male usage of 'housewife', in contrast, can imply promiscuity, as when Iago states 'A Housewife that by selling her desires / Buys herself Bread, and clothes' (*Othello*, 4.1). or very bawdily by Sir Toby Belch to Sir Andrew Aguecheek where a 'good housewife take thee between her legs and spin' his hair off (1.3). Interestingly, the use of 'unmanly', by the Jailer's daughter in *The Two Noble Kinsmen* implies that real men do not abandon women, but male collocates of 'unmanly' refer to actions such as relinquishing royal succession (*Henry VI, Part 3*), wearing French fashions (*Henry VIII*), grieving for a father (*Hamlet*) and the state of being depressed (*Timon of Athens*), all unbecoming of a 'man'.

Table 9 captures those words that are used by both genders to talk about males. Notice that the majority of these are used by both female and male characters.

Table 9. Gendered language forms concerning men used by females and males in Shakespeare's plays

Forms used by female and male characters (instances f/m)	
Bachelor/s (5/16)	Man-child/Men-children (1/1)
boy/s (71/344)	manhood/s (4/22)
Eunuch (4/10)	manly (3/15)
fellow/s (34/67)	mannish (1/2)
Gentleman/Gentlemen (63/639)	masculine (1/2)
Lad/s (1/48)	Monsieur/s (9/41)
male/s (3/13)	Signior/s (14/114)
man/men (415/2314)	Sirrah (17/124)
Forms used only by female characters	Forms used only by male characters
	Esquire/s (9) manlike (1) Tomboys (1)

There is also an absence of derogatory forms used by females of males (in contrast to the derogatory terms used by males of females we encountered in Table 7). 'Esquire' denotes the social rank of a candidate for knighthood or a country gentleman. The other two forms used only by male characters, 'manlike' and 'tomboys', are used, respectively, to criticize a man for being like a woman, and to refer to female prostitutes: Caesar says of Antony: 'Is not more manlike / Than Cleopatra: nor the Queen of Ptolemy / More Womanly than he.' (*Antony and Cleopatra*, 1.4); Iachimo talks to Imogen of 'Tomboys [whores] hired' by Posthumus (*Cymbeline*, 1.7). Thus we can see that even words implying maleness can be used to demean women.

5. Conclusion and future study

This paper has reported on the Encyclopedia of Shakespeare's Language (ESL) project, and some of the resources it has produced. Users are now able to explore the plays in a variety of (semi-automatic) ways via CQPweb. We have demonstrated, for example, how users can:

- interrogate Shakespeare's 38 plays using queries restricted by dramatic genre, gender and/or social status of characters, and

- target and explore such language use not only at the word level, but also at the semantic level (via semantic categories).

We have shown the benefits of being able to perform such searches, through an innovative exploration of language and gender in Shakespeare's drama. Our approach involved, first, an examination of the keywords characterising female and male speech across the 38 plays. We then broadened our scope by examining how female and male language varies in different genres (comedy, history and tragedy), and according to characters' social rank (high or low). Finally, we engaged in semantic category analysis as a means of exploring the collocational patterns of gendered forms of language usage ('girl', 'woman', 'lady', etc.) and thereby identified some similarities and differences in language behaviour between our female and male characters.

Amongst our myriad findings, we show that female characters use pragmatic markers indicative of grief or sorrow more than their male counterparts, and that high-status female characters, in particular, also have the emotion label 'sorrow' as one of their keywords. In contrast, male characters' pragmatic markers reveal them to be the architects of events: a finding that fits with the nature of the historical period we are dealing with, underwritten as it was by social hierarchy and patriarchy (Nevalainen and Brunberg, 2003: 32–38). Related to this, the keywords of middling to high status female characters highlight a focus upon local (and arguably more intimate) figures of power and authority, whilst male characters' keywords tend to focus more on the national figure (the male monarch). Female characters (of all statuses) were also found to be more relational in their use of first-person pronouns than their male counterparts. One of several genre-specific findings is that high-ranking male characters in comedy talk significantly more about women than their counterparts in histories and tragedies. These well-bred males in the comedy plays generally make requests more politely than their male counterparts in the histories and tragedies. They are characterised, in addition, by 'wit' and demonstrable sharp-mindedness as well as a greater focus on self (see, especially, their overuse of the singular first-person pronouns, 'I' and 'me', when compared with male characters from the histories and tragedies). When it comes to their use of gendered language forms, we found there to be more similarities than differences between female and male characters across all 28 plays: with one exception. Male characters use some terms that female characters do not, and which tend to be derogatory (of women especially).

The fine-grained analyses presented in this paper are representative of opportunities the ESL resources afford when it comes to a focus on Shakespeare plays only. It is worth noting, however, that this AHRC-funded project also gives scholars access to reference corpora of contemporaneous plays (see the ESC: Comparative Plays; see Demmen, 2020) and wider Early Modern English from 1560-1640, based on a portion of *Early English Books Online* (see the ESC: EEBO-TCP Segment; see Murphy, 2019). We believe that these particular resources, when combined with the annotated Shakespeare resource, will enable future users to better research not only Shakespeare's language usage as we do here, but importantly, to compare Shakespeare's depictions with those of other playwrights of the period. That is, future researchers will be able to engage in corpus-based investigation of Shakespeare's plays in comparison with other Early Modern English plays and playwrights (see, e.g., Culpeper et al. 2018) and/or Early Modern English more widely (see, e.g., Culpeper 2011; Hope and Witmore 2010: 387-390). As such, the possibilities of extending our understanding of Shakespeare's work are limited only by our imaginations.

Acknowledgments

We are grateful to the anonymous reviewers at Language and Literature for their comments, and also to Professor Alison Findlay at Lancaster University for her comments and advice on this article.

Declaration of Conflicting Interests

The author(s) declared 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

Sean Murphy: <https://orcid.org/0000-0002-1421-1605>
Dawn Archer: <https://orcid.org/0000-0002-4547-6518>
Jane Demmen: <https://orcid.org/0000-0001-6432-4942>

References

- Archer D and Bousfield D (2010) 'See better, Lear'? See Lear better! A corpus-based pragma-stylistic investigation of Shakespeare's King Lear. In: McIntyre D and Busse B (eds) *Language and Style*. Basingtoke/New York: Palgrave Macmillan, pp.183-203.
- Archer D and Culpeper J (2003) Sociopragmatic annotation: New directions and possibilities in historical corpus linguistics. In: Wilson A, Rayson P and McEnery AM (eds) *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Frankfurt/Main: Peter Lang, pp.37-58.
- Archer 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, pp.xx-xx.
- Archer D and Gillings M (this volume) Depictions of deception, focussing on five Shakespearean characters. *Language and Literature*.
- 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*. Farnham, UK/Burlington, USA: Ashgate, pp. 137-57.
- Baker P (2004) Querying keywords. Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics* 32(4): 346-59.
- Baron A and Rayson P (2008) VARD 2: A tool for dealing with the spelling variation in historical corpora. In: *Proceedings of the Postgraduate Conference in Corpus Linguistics*, Aston University, Birmingham, UK, 22 May 2008.
<http://ucrel.lancs.ac.uk/vard/about/> (accessed 23 June 2020).

- Bray T, Paoli J, Sperberg-McQueen C, Maler E, Yergeau, F and Cowan J (2006) *Extensible Markup Language (XML) 1.0* (2nd Edition). [Online] <https://www.w3.org/TR/xml11/> (accessed 23 June 2020).
- Busse B (2006) *Vocative Constructions in the Language of Shakespeare*. Amsterdam/Philadelphia: John Benjamins.
- Busse U (2002) *Linguistic Variation in the Shakespeare Corpus: Morpho-Syntactic Variability of Second Person Pronouns*. Amsterdam/Philadelphia: John Benjamins.
- Callaghan D (ed) (2016) *A Feminist Companion to Shakespeare*. 2nd Edition. Chichester: Blackwell.
- Culpeper J (2001) *Language and Characterisation: People in Plays and Other Texts*. Harlow: Pearson Education.
- 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 (2011) A new kind of dictionary for Shakespeare's plays: An immodest proposal. In: Ravassat M and Culpeper J (eds) *Stylistics and Shakespeare's Language. Transdisciplinary Approaches*. London/New York: Continuum, pp.58-83.
- Culpeper J and Archer D (2008) Requests and directness in Early Modern English trial proceedings and play-texts, 1640–1760. In: Jucker A and Taavitsainen I (eds) *Speech Acts in the History of English*. Amsterdam/Philadelphia: John Benjamins, pp.45-84.
- Culpeper J and Demmen J (2015) Keywords. In: Biber, D and Reppen, R (eds) *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, pp.90-105.
- Culpeper, J and M Kytö (2010) *Early Modern English Dialogues: Spoken Interaction in Writing*. Cambridge: Cambridge University Press.
- Culpeper J and McIntyre D (2006) Drama: Stylistic aspects. In Brown, K (ed) *Encyclopedia of language and linguistics*. Volume 3. 2nd Edition. Oxford: Elsevier, pp.772-785.
- Culpeper J, Archer D, Findlay A and Thelwall M (2018) John Webster, the dark and violent playwright? *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews*. DOI: 10.1080/0895769X.2018.1445515.
- Culpeper J, Hardie A, Demmen, J, Timperley M and Hughes J (in preparation) Supporting studying Shakespeare: Structural markup and grammatical annotation enhancing a corpus of the First Folio.
- Culpeper J and Oliver S (in press) Pragmatic noise in Shakespeare's plays. In: Jonsson E and Larsson T (eds.) *Voices of English: Tapping into Records Past and Present*. Amsterdam/Philadelphia: John Benjamins, pp.xx-xx.
- 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.
- Dusinberre J (1996) *Shakespeare and the Nature of Women*. 2nd Edition. Basingstoke: Macmillan Press.
- Findlay A (1999) *A Feminist Perspective on Renaissance Drama*. Oxford, UK/Malden, USA: Blackwell.
- Findlay A (2010) *Women in Shakespeare. A Dictionary*. London/New York: Bloomsbury.
- Grice HP (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hardie A (2012) CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3): 380-409.
- Hardie A (2014a) Modest XML for corpora: Not a standard, but a suggestion. *ICAME Journal* 38: 73-103.

- Hardie A (2014b) 'Log Ratio – an informal introduction'. ESRC Centre for Corpus Approaches to Social Science (CASS), 28 April, 2014. Available at: <http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/> (accessed 23 June 2020).
- Hope J and Witmore M (2010) The hundredth psalm to the tune of 'Green Sleeves': Digital approaches to the language of genre. *Shakespeare Quarterly* 61(3): 357-90.
- Jameson AM (2005) *Shakespeare's Heroines: Characteristics of Women, Moral, Poetical and Historical*, ed by Larsen Hoeckley, CL. Peterborough, ON: Broadview.
- Jardine L (1983) *Still Harping on Daughters: Women and Drama in the Age of Shakespeare*. 2nd Edition. Sussex, UK: Harvester Press/New Jersey, USA: Barnes and Noble.
- Jardine L (1996) *Reading Shakespeare Historically*. London: Routledge.
- Kemp TD (2010) *Women in the Age of Shakespeare*. Oxford, England: Greenwood Press.
- Leech G, Garside R and Bryant M (1994) CLAWS 4: The tagging of the *British National Corpus*. In: *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*, Kyoto, Japan, pp.622-628.
- Lenz CRS, Greene G and Neely CT (1983) *The Woman's Part: Feminist Criticism of Shakespeare*. Urbana and Chicago, USA: University of Illinois Press.
- Lutzky U and Demmen J (2013) Pray in Early Modern English drama. *Journal of Historical Pragmatics* 14(2): 263-284.
- McIntyre D and Walker B (2019) *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press.
- Murphy S (2015) I will proclaim myself what I am: Corpus stylistics and the language of Shakespeare's soliloquies. *Language and Literature* 24(4): 338-354.
- Murphy S (2017) 'Shakespeare and social status'. Encyclopedia of Shakespeare's Language Blog post 5 June 2017. Available at: <http://wp.lancs.ac.uk/shakespearelang/2017/06/05/shakespeare-and-social-status/> (accessed 23 June 2020).
- Murphy S (2019) Designing a genre classification scheme for *Early English Books Online* 1560-1640. *ICAME Journal* 43(1): 59-82.
- Nevalainen T (2006) *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.
- Nevalainen T and Raumolin-Brunberg H (2003) *Historical Sociolinguistics: Language Change in Tudor and Stuart English*. Harlow: Pearson Education Limited.
- Rayson P, Archer D, Piao SL and McEnery AM (2004) The UCREL semantic analysis system. In: *Proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 25 May 2004, Lisbon, Portugal. Paris: European Language Resources Association, pp.7-12.
- Rayson P (2008) From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519-549.
- Scott M and Tribble C (2006) *Textual Patterns. Key Words and Corpus Analysis in Language Education*. Amsterdam/Philadelphia: John Benjamins.
- Tannen D (2005). *Conversational style: Analyzing talk among friends*. Oxford: Oxford University Press.
- Thompson A and Roberts S (eds) (1997) *Women Reading Shakespeare 1660-1900: An Anthology of Criticism*. Manchester: Manchester University Press.

Notes

-
- ¹ This type of corpus linguistic study can usefully complement the large and longstanding body of existing feminist literary critical research (e.g., Dusiñberre, 1996; Findlay 1999, 2010; Jameson 2005 and Jardine 1983, 1996; for a wider overview see also Callaghan 2016; Lenz et al 1983; Thompson and Roberts 1997).
- ² We follow Nevalainen (2006: 1) in considering the ‘Early Modern’ period to be circa 1500-1700.
- ³ See <http://wp.lancs.ac.uk/shakespearelang/> (accessed 23 June 2020).
- ⁴ ‘Play-text’ refers to the written form of a play under consideration (Culpeper and McIntyre 2006: 775), serving as a reminder that the focus is upon written, textual versions of plays, in contrast to performances.
- ⁵ See <http://internetshakespeare.uvic.ca/> (accessed 23 June 2020).
- ⁶ The ESL project takes an inclusive approach in retaining plays which are now recognised to be collaborations between Shakespeare and other writers, but which are historically recognised as part of his body of work.
- ⁷ See <https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online/> (accessed 23 June 2020). The four EEBO source collections are The *English Short-Title Catalogue* (1475-1640), compiled by A.W. Pollard and G.R. Redgrave (1927); the *Short-Title Catalogue* (1641-1700), compiled by Donald Wing (1945-1951); the Thomason Tracts (1640-1661) and the Early English Books Tract Supplement. See further <http://eebo.chadwyck.com/about/about.htm> (accessed 23 June 2020).
- ⁸ UCREL = University Centre for Computer Corpus Research on Language.
- ⁹ See <http://ucrel.lancs.ac.uk/usas/> (accessed 23 June 2020).
- ¹⁰ See Hardie 2012. See also <https://cqpweb.lancs.ac.uk/> (accessed 23 June 2020).
- ¹¹ See <https://cqpweb.lancs.ac.uk/doc/cqpweb-simple-syntax-help.pdf> (accessed 23 June 2020).
- ¹² See http://cwb.sourceforge.net/files/CQP_Tutorial/ (accessed 23 June 2020).
- ¹³ See Hardie (2014b) and McIntyre and Walker (2019) for more detailed descriptions and use of Log Likelihood and Log Ratio.
- ¹⁴ Murphy (2019) details the prototype approach to genre taken in the ESL project; see also Demmen (2020: 39).
- ¹⁵ We have set aside results that are play-specific character names (‘Pisanio’, ‘Willow’, ‘Nerissa’, ‘Lysander’, ‘Malvolio’, ‘Romeo’ and ‘Nurse’), given they are topical, and ‘les’ and ‘Je’, which are play-specific (being French words occurring relatively frequently in *Henry V*).
- ¹⁶ As space does not allow for a full analysis of the gender of addressees to whom ‘pray’ forms are used, we can only suggest this as a worthwhile future study.
- ¹⁷ A detailed investigation of the numbers and types of requests made by female and male characters constitutes a second worthwhile future study.
- ¹⁸ ‘Mistress’ tends to be used in titles, made up of honorific + surname (e.g., ‘Page’, ‘Ford’, ‘Anne’, ‘Quickly’).