Style and chronology: a stylometric investigation of Aphra Behn’s dramatic style
and the dating of *The Young King*

Abstract

Aphra Behn’s dramatic outputs are recognised for their diversity and responsiveness to trends in Restoration drama. A stylometric approach is used to investigate the linguistic dimension of Behn’s dramatic style, with a particular focus on evidence of chronological change. Quantitative analysis (most frequent words, function words, zeta) suggest that Behn’s drama falls into three periods. A qualitative analysis indicates that the periodisation may reflect a change in the construction of Behn’s dramatic worlds, from an abstract psychological focus to a more grounded, interactive and social representation. The study considers the problematic dating of Behn’s tragi-comedy *The Young King*. Although critical opinion holds that this play was the first that Behn wrote (i.e. pre-1670), the stylometric analysis suggests that Behn heavily revised, or indeed, penned, the drama in the mid-to-late 1670s, mid-way through her writing career. The paper demonstrates the potential for stylochronometric techniques to complement other linguistic approaches to style, and enhance our understanding of how literary writing evolves.

Keywords

Authorial style, chronological change, Aphra Behn, stylometry, *The Young King*
Introduction

In 1682, Aphra Behn’s tragi-comedy *The Young King* was published; the thirteenth of her sixteen plays. In the dedicatory epistle, Behn apologises for the poor quality of the play, which, she explains, is because she wrote it early in her career: it represents ‘the first Essay of my Infant-Poetry’. She fears the play’s simplistic and immature style will be criticised by the Restoration audience, as the play is ‘harmless and unadorned, unpracticed in the Arts to please’. Any positive elements are attributable (not without some self-congratulation) to the author’s ‘natural & unskill’d Innocence’ (Behn, 1682: 2).

Behn’s defense of the *The Young King* demonstrates a long-standing commonplace of chronological change in literary writing; namely that an author’s literary style changes over the course of their career (and this, at least in Behn’s eyes, is an ameliorative process). Modern literary critics frequently periodise authorial outputs on this basis (see, e.g. McMullan (2007) on Shakespeare). Yet, as van Hulle and Kestemont (2016) recently observe, literary periodisation is traditionally founded on expert intuition, rather than a systematic analysis of linguistic evidence.
This paper investigates evidence of chronological change in the literary style of Aphra Behn (b. c. 1640 – d. 1689) from a stylometric perspective. Focusing specifically on her dramatic works (1670-1690), I assess: the extent to which her proclaimed early ‘unpract’ic’d’ style changes over her two decade career as a playwright; the characteristic features of this change; and the possible relationship with these chronological shifts to other factors, such as genre. Following a general overview of stylometric methods and the dramatic writing of Aphra Behn, the first part of the analysis explores evidence of chronological changes in Behn’s dramatic works. Stylometric methods are supplemented with qualitative readings of the linguistic data, to contextualise and verify the statistical trends. The second part of analysis examines the possible composition date of *The Young King*, based on the stylometric evidence, assessing the legitimacy of Behn’s description of the play as her ‘Infant-Poetry’. I conclude with a reflection on the insights provided by stylometric analysis, and its potential application to literary and stylistic questions beyond the remit of authorship attribution.

**Background: Stylometry**

Stylometry uses computational techniques to identify patterns in linguistic features across literary (and non-literary) works not readily discernable to the human eye. Its
primary application has been for authorship attribution (see, e.g. Boyd and Pennebaker, 2015; Kestemont, Luyckx, Daelemans and Crombez, 2012; Stamatatos, 2009). Stylometry uses the concept of the “linguistic fingerprint”, the linguistic profile unique to a given speaker or writer, as a means of evaluating works of questionable authorship. The fingerprint, the authorial signal, is identified from linguistic patterns in the data, which are derived from features that are not necessarily salient to a reader. For English, most frequent function words (e.g. *in*, *and*, *to*), most frequent words (function and content words combined), mid-range frequency words (known as *zeta*; see below and Burrows 2007), and the most frequent 2- and 3-grams have proved particularly effective. These linguistic markers are relatively distinct from contextual factors, such as topic, and occur sufficiently frequently across different writings by the same author). New methods, such as Word Adjacency Networks (Eisen et al., 2016) continue to further our capacity to identify the author signal on the basis of lexical and co-lexical patterns.

Stylochronometry is a sub-area of stylometry that focusses specifically on quantifying changes in style over time. Time is one of the most significant factors for ‘the evolution of the literary lexicon’ (Rybicki, 2015: 759). Changes emerge both in studies of the language of a single author (i.e. across their writing career), and in multi-author, multi-generational corpora. Rybicki muses that ‘both phenomena cannot be blamed on the same mechanism of linguistic change’ (Rybicki, 2015: 759), reflecting the complex
intersection between style change at the authorial level and larger-level systemic changes in the language.

Stylochronometric analyses of literary writing encourages us to ‘reassess critical commonplaces’ (van Hulle and Kestemont, 2016: 197) concerning writers, works and periods (for a survey, see Stamou, 2008). For example, van Hulle and Kestemont test the traditional periodisation of Beckett’s French and English fictional prose, and find that stylometric methods affirm the more intuitive pronouncements of previous critics. Their findings also indicate that authorial style does not necessarily show a consistent trajectory of change, nor is it restricted to specific periods of an author’s literary career. Other recent applications have examined authorial claims about changing style (e.g. WB Yeats; (Forsyth, 1999)), as well as the relevance of stylochronometric methods for non-literary purposes, such as the diagnosis of Alzheimer’s Disease (Kernot et al., 2017, Hirst and Wei Feng, 2012).

Stylometric research has also investigated the impact that other factors, including genre, gender (e.g. Jockers, 2013), and characterization (e.g. Burrows and Craig, 2012; Delgado and Miguel, 2015) have on the distribution of linguistic markers in a literary corpus. However, studies repeatedly show that the presentation of these factors occurs ‘within a larger framework of persistent authorial proclivities and specialisations’ (Burrows and Craig, 2012: 292); that is, within the over-arching boundaries of the
authorial signal, or fingerprint. Any changes in authorial style thus encompass, but, it should be assumed, are also shaped by, the other factors.

As Stamou (2008) observes, we know relatively little about how changes over time intersect with other social and biographical factors. She argues that stylochronometry has ‘untapped potential’ to enrich our ‘broader understanding of the creativity process related to the written word’ (Stamou, 2008: 196). Style change in literary writing intersects with variation and change in the language system as a whole, and with the daily psychological and social experiences of a language user. Stylochronometric methods can provide a valuable perspective on this complex linguistic relationship. The present investigation uses stylochronometric research to explore late seventeenth-century dramatic style: specifically the style of Aphra Behn.

**Background: Aphra Behn**

As one of the first English women to have her works performed on the London stage, Aphra Behn occupies a significant place in English literary heritage. Her dramatic writing spans a twenty-year period, 1670-1690, encompassing sixteen plays. This makes her one of the most prolific writers of her age, with more dramatic premieres between 1670 and 1689 than any of her contemporary male playwrights (Hughes, 2004: 30). As is typical of Restoration theatre (Love, 2005), Behn’s plays are predominantly comedies (twelve in total), with three tragi-comedies (*The Forced Marriage, The Young King* and
Widdow Ranter) and one tragedy (Abdelazer). There are also (at least) five other plays from the period with long-standing associations with Behn (see Hume and Milhous (1983)). The findings of the present study will assist in the future verification of their authorship attribution.

Critics (Hughes, 2004: 32; Owen, 2004) note that Behn’s dramatic writing is very responsive to changing audience tastes, and to the turbulent political events of her lifetime, all whilst catering for the talents of the repertory theatre company. Behn can be considered to be part of the reactive, social-literate sphere of the period, in which literary writing was a ‘social practice’ (Griffin, 2005: 38). The trends in genre and overall output suggests that Behn ‘was very much a writer [...] who had her finger on the cultural “pulse”’ (Owen, 2004: 69). These critical assessments indicate that changes in Behn’s style might be expected, although it is not clear to what extent, and how, her response to market sensibilities may manifest in linguistic features examined.

Stylochronometric analyses concerning the chronology of an author’s works (i.e. based on linguistic change) are most effective when external evidence is available to corroborate the linguistic results (Stamou 2008: 185). The external evidence for dating Behn’s plays is based on historical records of first performance (if known) and/or the documented publication date(s). Admittedly, these sources do not provide uncontestable proof. However, generally speaking, the performance and publication dates fall within
6-12 months, and Behn’s plays can be organised according to their theatrical season (autumn to the following summer) (see Table 1).

<table>
<thead>
<tr>
<th>PLAY</th>
<th>THEATRICAL SEASON</th>
<th>GENRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Forc’d Marriage</td>
<td>1670-1</td>
<td>Tragicomedy</td>
</tr>
<tr>
<td>The Amorous Prince</td>
<td>1670-1</td>
<td>Comedy</td>
</tr>
<tr>
<td>The Dutch Lover</td>
<td>1672-3</td>
<td>Comedy</td>
</tr>
<tr>
<td>Abdelazer</td>
<td>1676-7</td>
<td>Tragedy</td>
</tr>
<tr>
<td>The Town Fopp</td>
<td>1676-7</td>
<td>Comedy</td>
</tr>
<tr>
<td>The Rover</td>
<td>1676-7</td>
<td>Comedy</td>
</tr>
<tr>
<td>Sir Patient Fancy</td>
<td>1677-8</td>
<td>Comedy</td>
</tr>
<tr>
<td>The Feign’d Curtizans</td>
<td>1678-9</td>
<td>Comedy</td>
</tr>
<tr>
<td>The Young King</td>
<td>1679-80 (pub. 1682)</td>
<td>Tragicomedy</td>
</tr>
<tr>
<td>The Second Part of the Rover</td>
<td>1680-1</td>
<td>Comedy</td>
</tr>
<tr>
<td>The City-Heiress</td>
<td>1681-2</td>
<td>Comedy</td>
</tr>
<tr>
<td>The Roundheads</td>
<td>1681-2</td>
<td>Comedy</td>
</tr>
<tr>
<td>The False Count</td>
<td>1681-2</td>
<td>Comedy (Farce)</td>
</tr>
</tbody>
</table>
The discrepancy between the publication date of *The Young King*, and the reported earlier date of composition, thus presents an exception in the chronology of Behn’s dramatic works. The play was first published in 1682 (the title page post-dates it to 1683), but Behn’s apologetic preface points to its juvenile style. Consequently, it is often taken by critics to be her first play. Hayden suggests a date between 1664 and 1670, noting that the play is ‘rather unsophisticated and straightforward for such an experienced playwright’ (Hayden, 2010: 18, 20). She finds plot features and character
types, such as druids, to have equivalents in other 1660s plays. Narrative continuities with Behn’s plays from the early 1670s are also cited as evidence for an earlier (pre-1676) composition date (Owen, 2004; Pacheco, 2015). Attempting to explain the delay between composition and performance/publication, Todd and Hughes (2004: 84) suggest that Behn shelved the play after observing the failure of a similar work, Edward Howard’s *The Women’s Conquest* in c.1670, in-keeping with Behn’s sensitivity to the literary marketplace. However, there is no external evidence that indicates when Behn actually wrote the play. The preface apology could be disingenuous, with the profession of the play’s immaturity instead a trope of modesty and/or an excuse for the play’s perceived weaknesses.

If Behn’s claim for the earlier date of composition is accepted, then it might also be asked to what extent she later revised the play, prior to its performance and publication. No critical consensus has been reached. Likely areas of revision relate to themes and character traits (e.g. anti-religious sentiment), rather than specific scenes or acts (Pacheco, 2015: 322). Owen (1996: 304) finds no ‘signs of immaturity’, and suggests Behn may have edited the play, perhaps quite extensively, in the late 1670s to enhance its ‘topicality’ for current events. Owens’ proposed date of revision reflects the final piece of the chronological puzzle: the date of first performance. A reference in the play’s epilogue to the Duke of York’s second exile in Flanders, which occurred in 1679, has been taken to suggest it appeared on stage in either 1678-9 season, or the autumn of
the 1679-80 season (Hume and Milhous, 1974: 385; Owen, 1996: 303–4). However, it is not implausible that the play was performed, and thus revised, nearer to its publication in 1682. A stylochronometric analysis offers a means of evaluating the legitimacy of the composition-to-publication narrative of *The Young King*.

The present stylometric analysis of Behn’s dramatic work therefore has two objectives. The first goal is to identify any evidence of change over time in Behn’s dramatic style, in support of Behn’s own assertions (although the stylometric assessment does not seek to provide an evaluative measure of the purported sophistication or immaturity of her writings). The second objective is to use the stylometric evidence to offer insight into the possible date of composition of *The Young King*, and thus establish its chronological position in Behn’s dramatic works.

**Methodology**

Stylometric (and stylochronometric) analysis investigates linguistic trends in a text or corpus of texts, and extrapolates the plausible connection between these trends and extra-linguistic factors, such as authorial style, time period, genre and so on. The linguistic trends are metaphorically known as signals (authorial, temporal, generic). The identification of these signals involves two methodological steps. Firstly, the selection of the linguistic features for analysis, and, secondly, the statistical processes used to identify underlying trends in the quantitative data.
The present study uses three lexical features that have repeatedly shown reliability for attribution investigations. The first is the most frequent words (MFWs) in the dataset (i.e. a corpus of Behn’s dramatic writing). The MFWs are calculated using the stylometric software package, Intelligent Archive (Craig et al., 2015), which produces a word-list organised by frequency. The top 300 words are used in the statistical analysis; MFW measures appear most robust when using 100-500 words (e.g. van Hulle and Kestemont, 2016).

The second lexical measure calculates the distribution of most frequent function words in the dataset (MF-FWs). These words are extracted from the same word-list produced for the MFWs (following van Hulle and Kestemont (2016)). MF-FWs include pronouns, modal and auxiliary verbs. Although some authorship studies remove pronouns (e.g. Binongo, 2003; Hoover, 2007), van Hulle and Kestemont (2016: 182) argue persuasively that this category may reflect particular ‘stylistic shifts’ relevant for stylochronometric periodisation, hence their retention here.

The third lexical measure are mid-frequency words, items that occur ‘between the extremes of ubiquity and rarity’ (Burrows, 2007: 27), also known as zeta words on account of the statistical process used in their analysis. Zeta words are those that are most representative of the “typical” lexis of an author’s writing, but not necessarily the most frequent. The analysis considers up to 500 zeta words characteristic of a given
dataset; again, a threshold that has been found to be a reliable measure in previous studies (e.g. Burrows, 2007).

The values of the three lexical measures are calculated for every play-text, in order to identified continuities and differences that may reflect change in Behn’s chronological style. To counteract differences in text length, the plays are subdivided into “blocks” of e.g. 3000 words, and the profile of each block is calculated. For example, the items identified as the MFWs in Behn’s overall dramatic corpus are re-counted on a block-by-block, play-by-play basis. This offers a linguistic profile of MFWs in, for example, *The Forced Marriage* (blocks 1-6) which can be compared with the distribution of the same MFWs in *The Rover* (blocks 1-6).

The resultant vectors (that is, the grid of words and frequencies in blocks of play-text) are then statistically analysed, to establish whether there are underlying patterns in the dataset. In the analysis of MFWs and MF-FWs, two unsupervised tests are used: Cluster Analysis and Principal Components Analysis (PCA). Unsupervised statistical analysis identifies inherent similarities and associations within a dataset; that is the statistical procedure organises the data into sub-sets of correlated variables (for instance, features associated with the authorial signal of Author A, and another group that are associated with the authorial signal of Author B). It is well-suited to exploratory (rather than affirmative) examinations of a dataset (Tolmie et al., 2011: 173). Both processes are
data reduction techniques: they reduce a large matrix of variables (e.g. 300 MFWs across 50+ segments of play-texts) to two or three underlying dimensions (Tolmie et al., 2011: 158-9) that can be interpreted more easily. SPSS 24 is used for both statistical tests.

The two processes have slightly different exploratory emphases. Hierarchical cluster analysis “forces” the dataset to identify similarities between texts, grouping them together into clusters of “likeness”. This is typically visualised with a dendrogram, in which the length of connections between entities (e.g. play-texts) corresponds to their identified similarity. Cluster analysis will always find some similarity between factors, even if this is, relatively speaking, very tenuous. By comparison, Principal Components Analysis (PCA) identifies factors that can best account for the majority of the variance within the dataset along multiple dimensions (e.g. x-, y- and sometimes z-axes). It thus reveals greater nuances within a dataset, such as outliers, along more than one dimension (Antonia and Jordan, 2014: 312). Cluster analysis and PCA provide an account of the underlying correlations in the linguistic data extracted from Behn’s play-texts. The role of the analyst is thus to identify whether these sub-sets can be attributed to specific factors, such as time-period.

The zeta analysis, used for the mid-frequency vocabulary, is a supervised method. This relies on the analyst configuring the dataset, and is used to confirm or reject a pre-
existing hypothesis (such as the dating of a play as being either one decade or another). Zeta was developed by Burrows (2007) and has been used in a range of investigations comparing authorial styles for attribution (e.g. Antonia et al., 2014), as well to investigate chronological style changes in the work of a single author, such as Hoover’s study of Henry James (Hoover, 2007). In the present study, zeta is used to confirm the temporal groupings suggested by the exploratory results obtained from the PCA and cluster analyses. Zeta is performed using Intelligent Archive (Craig et al., 2015).

Zeta calculates the frequency and proportion of zeta words in (segments of) a dataset, such as Behn’s dramatic writing (known as the base-set). The proportion of the zeta words is also calculated in another group of texts, the counter-set, which represents, for example, another author’s writings. The process is then run again, but with the opposing configuration, so that the zeta words of counter-set texts is identified and the proportional frequencies of those words in the base-set are calculated. A questionable text can then be evaluated against the profiles of each text set (Burrows, 2007: 32–4).

The reliability of the zeta tests are verified using randomly-extracted segments, comprising 10% of each text-set. If zeta is calibrated accurately, the test segments should fall alongside their associated text-set segments (e.g. plays from the same decade). Pilot tests found that sharply contrasting frequencies of zeta items (i.e. 80% occurrence in segments in the base-set with 20% occurrence in segments in the counter-
set) did not sufficiently differentiate the text-sets, despite their effectiveness in previous studies (e.g. Hoover, 2007). Instead, a threshold of < 60% segments, and up to 100% segments, in both text-sets, was found to be most reliable for Behn’s dramatic works.

The stylometric analyses require electronic texts, which were obtained from EEBO-TCP. Each text file is formatted as follows:¹

1. Each play is encoded using TEI (Lite) XML conventions for drama (speaker IDs, stage directions), demarcating songs and foreign-language sections.
2. Paratextual material is removed, such as dedicatory epistles, prologues, epilogues, and songs not documented as Behn’s.
3. The seventeenth-century spelling is normalised using VARD 2.5.4 (Baron 2013). Contractions were expanded and verb-endings e.g. knowest, modernized. However, second-person pronoun forms you and thou are retained.
4. All texts are manually proofed to correct errors arising from tagging or spelling automation (e.g. I for aye).

¹ The author received invaluable assistance in this preparatory process from Georgia Priestley, formerly an undergraduate student at the University of Birmingham. Prof. Elaine Hobby and Dr. Claire Bowditch checked the EEBO-TCP against originals held in U.K and US libraries.
When calculating the MFWs, MF-FWs and zeta words, character and place names are excluded, so that these content-specific, high-frequency items (e.g. plays with characters named Antonio) do not distort the results.

**Results: Stylochronometric Exploratory Analysis**

The preface to *The Young King* suggests that Aphra Behn believed that (her audience would believe) changes in authorial style were not only possible, but readily perceptible. The first stage of analysis is to ascertain the validity of Behn’s assumption. This exploratory analysis uses MFWs and MF-FWs, processed using cluster analysis and PCA. To avoid potential interference from other stylistic factors, such as genre, the analysis focusses on Behn’s twelves comedies. These plays comprise the majority of her dramatic writings and span both decades of her literary career.

The results support Behn’s claim for change in style over time. The arrangement of plays suggested by the statistical tests broadly correlates with their dates of performance and publication, indicating a temporal signal is present in Behn’s comedies. Using the MF-FWs data, which includes 133 function words extracted from the 300 MFWs wordlist, the cluster analysis (Figure 1) produces three groups of plays that are interpretable as chronologically-organised sets. The first cluster includes plays dating from 1681-2 (*Roundheads, False Count*) and 1685-7 (*Emperor of the Moon, Lucky Chance*). The second cluster comprises two sub-groups. The first includes plays
performed between 1676 and 1682, including Behn’s best-known comedy, *The Rover*. The second sub-group includes plays performed in the earliest years of Behn’s career (1670 to 1677). Whilst some plays from the same season appear in different clusters e.g. the 1677-8 plays *Town Fopp* and *The City-Heiress*, no play occurs out of its chronological sequence.

Figure 1: Hierarchical cluster analysis. 133MF-FWs. Behn’s comedies.

Figure 2: PCA results. 133MF-FWs. Behn's comedies.

The PCA results (Figure 2) reduce the MF-FWs data to two dimensions. Chronology is a plausible explanatory factor for both the x- and y-axes. On the y-axis, which accounts for 27.9% of the variance in the dataset, plays with a higher score are generally earlier in date than those with a lower (negative) score, providing a scalar value that corresponds with time-period. The x-axis, which accounts for 28.4% variance, also suggests temporal organisation, although in a slightly different way. Plays from the period 1677-82 score higher on this component than plays written either before or after this period, with one exception: *The Dutch Lover* (1672-3), has a higher value than the later *Town Fopp* (1676/7). Collectively, the exploratory results from MF-FWs suggest that there is a chronological style shift in Behn’s dramatic writing, and this can differentiates her plays in three broad groups, roughly equivalent to early, middle and late stages of her twenty-year career.
The relationship between the statistical results and a higher-level factor, such as time period, is an interpretative, and therefore not infallible, analytic step. However, the MFWs support the preceding assessment of the MF-FWs results. The MFWs for Behn’s plays include a mixture of function and content words, including address forms (e.g. *sir*), interjections (e.g. *oh*) and lexical verbs (e.g. *love*, *come*), and are therefore more susceptible to interference from other factors, such as topic. However, the groupings identified in both statistical tests are comparable with those obtained in the MF-FWs analysis (see Figure 3 and Figure 4). Cluster analysis identifies three temporal clusters, although there are small differences with the MF-FWs results (Figure 1). For example, *Sir Patient Fancy* is situated on its own branch within the middle play cluster (1676-82), suggesting that the proportion of MFWs differs to that of the contemporary comedies. *Emperor of the Moon* is also now a marginal outlier in the late play group cluster. Such deviations are perhaps to be expected. Style change is not necessarily consistent, nor rectilinear (e.g. van Hulle and Kestemont, 2016). The permeation of topic into the MFWs data is also a potential factor.

The PCA results are also comparable to those obtained using MF-FWs (Figure 4 combines the results for the MF-FWs and MFWs analysis). Time-period again provides the most persuasive interpretation of the two dimensions, with only *Dutch Lover* sitting outside the expected chronological sequence.
The exploratory analysis suggests that Behn’s works can be chronologically differentiated on linguistic grounds. Moreover, her plays appear to sub-divide into three groups, which broadly correlate with the early, middle and late years of her career. The posited groupings are shown in Table 2.

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Plays (Cluster analysis and PCA results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early (1670-77)</td>
<td><em>Amorous Prince, Dutch Lover and Town Fopp</em></td>
</tr>
<tr>
<td>Middle (1677-1682)</td>
<td><em>The Rover, Sir Patient Fancy, The Feign'ed Curtizans, Second Part of the Rover and City Heiress</em></td>
</tr>
<tr>
<td>Late (1682-90)</td>
<td><em>Lucky Chance, False Count, The Roundheads and Emperor of the Moon</em></td>
</tr>
</tbody>
</table>

Table 2: Periodization of Behn's comedies, based on exploratory stylometric analysis.

Of course, the periodization of these works is not explained solely by the passing of time. In the next section, the language data underpinning these results is examined in more detail, and considers the possible forces shaping the chronological changes in Behn’s dramatic style.

**Results: Exploring Stylochronometric Data**
Word Plots

The stylometric exploratory analysis indicates that a temporal signal is identifiable in Behn’s dramatic writing. In this section, the statistical data, and the language that underlies it, is scrutinized to better understand the chronological groupings of early, middle and late plays. The analysis continues to focus on Behn’s comedies, rather than combining genres, as the twelve plays provide a more consistent and coherent body of work in which to track and interpret chronological change.

Figure 5: Word plot of 300MFWs in Behn's comedies.

The word plot in Figure 5 shows the MFWs used for the PCA analysis. The majority of words score high on the first component, with a more variable range on the second component. This suggests that the linguistic style of Behn’s comedies is most coherent for the plays produced in the middle of her career (e.g. The Rover, The City-Heiress). Words scoring highly on the first component include interjections such as pox, nouns referring to social roles (friend), entities (devil), and women (lady, women, mistress), and evaluative adjectives, both positive and negative (damned, honest, poor, dull). For the second component, the highest scoring items include the second-person pronouns (thou, thee), epistemic verbs (must, believe), references to human features (eyes, face) and abstract (metaphysical) concepts, including reason, and passion. These items are most strongly associated with Behn’s early plays, such as The Amorous Prince. Finally,
lexical items in the space associated with the late plays comprise a rather disparate collection (e.g. arms, honour, lord, great, their, world, heard). Read collectively, the MFWs associated with the early plays foreground subjectivity, and the cognitive and emotional experiences of the characters. As Frederick muses in *The Amorous Prince*:

> But I love pleasures which thou can not prize
> Beyond dull gazing on thy Mistress eyes,
> The lovely object which enslaves my heart (1.2)

The wordplot indicates that Behn’s earliest plays focus on the psychology and self-experience of her characters and their dramatic world. The MFWs associated with post-1677 comedies show a greater focus towards the external, social world, and the interaction and interpersonal lives of its participants. In *Sir Patient Fancy*, for instance, interjections, social labels and epithets indicate the speaker’s stance towards her topic, and its significance for her relationship with her interlocutor, Isabella:

> Isabella: You will grow very expert in the arts of love, Fanny?
> Fanny: When I am big enough I shall do my endeavour, for I have heard you say, Women were born to no other end than to love: And it is fit I should learn to live and die in my calling. Come open the Gate or you will repent it, we shall
have my Father marry you within a day or two to that ugly man that speaks hard words. Agad I cannot abide him (3.4)

The dispersion of the wordplay therefore suggests that the chronological change in Behn’s style could be attributed to a decrease in the representation of characters’ inward, psychological and abstract experiences, and a shift towards characters shaped by their interactions, in a dramatic world more expressly focussed on human relationships. To explore this hypothesis in more detail, the MF-FWs and MFWs data is supplemented with another kind of lexical evidence: keywords.

**Keyword Analysis**

The exploratory stylometric analysis indicates that Behn’s plays can be grouped into three sub-periods. Keyword analysis offers another perspective on the potential contrasts between these groups of plays. Keywords are lexical items that have an unexpected frequency in a corpus than is predicted based on a reference (target) corpus, using a statistical measure e.g. log-likelihood ratio (Weisser, 2015: 183). Positive keywords are those that occur more frequently than expected, and negative keywords are items that occur less frequently. Unlike the high-frequency words used in the exploratory analysis, keywords are closely bound to the topic of the texts in which they
appear, and their frequencies are therefore relative to the texts being compared. Content keywords (e.g. love) reveal the “aboutness” of a text, whereas function keywords (e.g. he) provide more stylistic information. Keyword analysis thus produce a semi-autonomous dataset without researcher bias (Murphy, 2015: 341, see also Archer, 2009), that may indicate how the stylometric periodizations intersect with change in other area, such as subject-matter or genre. The keyword analysis considers contrasts between the three period groupings of Behn’s comedies.

The keywords generated for each period grouping of Behn’s plays uses the plays from the other periods as the reference corpus e.g. early comedy keywords were obtained through comparison with the middle and late comedies. This ensures the focus is on temporal style shifts rather than Behn’s style as an author (as might result from using a reference corpus of contemporary Restoration drama). Up to 200 keywords (positive and negative) were calculated in AntConc, with only those with a log-likelihood value > 15.13 (p > 0.0001) considered. The analysis disregards character names, as these are not considered indicative of period-specific style.

Given the connection between keywords and the subject-matter of a corpus, the word-lists were classified into different semantic fields. Gender-related keywords (split into male and female) offer a useful and relevant starting point, gender relations being an
area for which Behn’s work has received considerable attention (see Hayden, 2010: 12-16) (Table 2; log-likelihood scores are shown in brackets).

Female keywords for Behn’s early comedies suggest a narrow sphere of social roles, comprising young women and their companions (maid, nurse), and sororal kinship. The presence of pronouns (she, her) suggests that women are “talked about” more often in these early plays than in later comedies. The negative keywords reinforce this female focus, as they include mainly masculine items (including his). The middle play keywords suggest a more diverse representation of female social roles, acknowledging their sexual activities and positioning with regards to their male interlocutors (heiress, whore, child). In the late comedies, women are defined according to their marital and/or social role (viscountess, wife, landlady). The female keywords show a change over the three periods from youthfulness to middle-age, loosely correlating with Behn’s own age and life experiences.
<table>
<thead>
<tr>
<th>Keyword</th>
<th>Early: F</th>
<th>Early: M</th>
<th>Middle: F</th>
<th>Middle: M</th>
<th>Late: F</th>
<th>Late: M</th>
</tr>
</thead>
</table>

<p>| Positive | sister (129) | nurse (87.8) | maid (41.3) | she (21.9) | her (16.6) | prince (83.6) | brother (48.1) | lord (17) | father (16.5) | child (58) | courtezan (44.8) | whore (36.2) | woman (19.8) | heiress (16.5) | seignior (41) | uncle (30.2) | lieutenant (18.2) | seigniors (18.2) | nephew (16.6) | landlady (77.5) | wife (76.1) | wives (21.3) | viscountess (20.2) | lord (90) | lords (64) | husband (55.1) | cuckold (52.2) | sirs (48) | highness (44.6) | king (40.8) | captain (36.9) | general (31.7) | farmer (30.3) | majesty (29.4) | page (27.7) | officer (27.3) | seamen (2) | knave (17.7) | governor (15.3) | fellow (15.3) |</p>
<table>
<thead>
<tr>
<th>Negative</th>
<th>child (17.8)</th>
<th>cuckold (36.5)</th>
<th>governor (28.7)</th>
<th>gentlemen (28.3)</th>
<th>king (28)</th>
<th>his (25.7)</th>
<th>lords (16.4)</th>
<th>wife (76.9)</th>
<th>landlady (35.3)</th>
<th>nurse (26.3)</th>
<th>sister (20.7)</th>
<th>lord (176.3)</th>
<th>doctor (156.9)</th>
<th>prince (61.9)</th>
<th>page (40.3)</th>
<th>highness (37.7)</th>
<th>father (29.8)</th>
<th>lords (26.5)</th>
<th>husband (17.7)</th>
<th>brother (16.8)</th>
<th>sister (64.9)</th>
<th>her (60.8)</th>
<th>she (34.3)</th>
<th>nurse (30.8)</th>
<th>child (26.5)</th>
<th>whore (23.5)</th>
<th>maid (16.3)</th>
<th>boy (32.3)</th>
<th>knight (16.8)</th>
</tr>
</thead>
</table>

Table 3: Gendered keywords for Behn’s “early”, “middle” and “late” comedies. Extracted from top 200 keywords; log-likelihood > 15.13.
In the male keywords, the early comedies refer to young men of high status (*prince*), and kinship roles (*father, brother*). In the middle comedies, male roles diversify to include professional roles (e.g. *lieutenant*). In the late comedies, male keywords are more numerous and diverse, both compared to the male keywords for earlier plays and the late period female keywords. This suggests that Behn’s later comedies are more concerned with patriarchal affairs and institutions, an interpretation further supported by the greater quantity of female negative keywords, and an increase in terms relating to marital relationships (*husband, cuckold*).

Viewed collectively, the distribution of gendered keywords suggests that Behn’s early works focus on youth and marital availability, whereas the later comedies show a more diverse set of social, political and marital categories. This can be seen to fit the preceding assessments of the stylochronometric and word-plot data, which proposed a shift from subjectivity and psychological experiences, towards greater interactivity and inter-personal language markers.

Other keywords support this interpretation. Content keywords in the early comedies include those relating to high emotion and physical action (e.g. *vow, kill, punish, wronged, promised, unhappy, revenge, sin, innocence*). These words could be taken to indicate a dramatic world of heightened and contrastive experience, more in-keeping with youthful experiences of the world. Function keywords, which have a less overt
connection to topic perhaps offer more persuasive evidence. For the early plays, keywords are connected to the self and expressions of intimacy, including first-person pronouns (*I, my*) and second-person pronouns (*you, thou, thy, thee*). This supports the interpretation, based on the MFWs wordplot, of the early plays’ characters’ greater subjectivity compared to Behn’s later writing.

Also supportive of this interpretation are the keywords for the middle comedies, which comprise ten different interjections (Table 4). *Whe* occurs in four middle-period comedies: *The Rover, The Second Part of the Rover, The Feign'ed Curtizans* and *The City-Heiress*, and its usage fits the OED definitions of an ‘exclamation [...] used for emphasis, or to demand attention’ or a discourse marker ‘used to introduce an anxious question or a statement of something regrettable’ (OED Online, 2017) e.g.: ‘Whe my Friend here, and I have sent and paid our Fine for a small Tenement of pleasure’ (*The Feign'ed Curtizans*).

<table>
<thead>
<tr>
<th>Discourse Marker</th>
<th>LL Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>whe</td>
<td>185.5</td>
</tr>
<tr>
<td>godsheartlikins</td>
<td>60.8</td>
</tr>
<tr>
<td>agad</td>
<td>53.9</td>
</tr>
<tr>
<td>zoz</td>
<td>48.6</td>
</tr>
<tr>
<td>gad</td>
<td>31.52</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>pox</td>
<td>24</td>
</tr>
<tr>
<td>ha</td>
<td>22.8</td>
</tr>
<tr>
<td>harkee</td>
<td>18.2</td>
</tr>
<tr>
<td>prithee</td>
<td>16.7</td>
</tr>
<tr>
<td>damn</td>
<td>16.2</td>
</tr>
</tbody>
</table>

Table 4: Exclamatives, oaths and other discourse markers in the middle comedy keywords.

Curiously the OED lists whe as obsolete from the early sixteenth century, and a search of 83 Restoration plays finds no other instances of whe. Behn may have resurrected an archaic term for dramatic purposes, perhaps to achieve specific objectives relevant to characterisation.³ Godsheartlikins is a regularisation of a range of spelling variants for a now-obsolete euphemistic oath. In Behn’s comedies, it is an idiolectal marker of Ned Blunt, the foolish associate of Willmore in The Rover and its sequel. Another interjection with a narrow range of use is zoz. It has no entry in the OED, although it may be a corruption of other euphemistic oaths. Zoz occurs only in Sir Patient Fancy, and frames expressions of negative emotions.

The keyness of interjections for Behn’s mid-career comic writing may have socio-biographical significance. The practice of witty conversation among literary circles in the period provided a “model” for the comedic dramatic dialogue in the same authors’

---

² Pox has non-exclamative uses in Behn's plays but the exclamative function is dominant.
³ It is plausible that the form survived within vernacular contexts, and was not preserved in writing.
works (Griffin, 2005: 38). The ability to participate in, and lampoon, word-play was a marker of membership within this community. Behn’s lexical creativity and increased focus on the interactive dimensions of characters’ dialogue may be interpreted as a demonstration of the legitimacy of her position as a playwright in the late 1670s and early 1680s.

The final group of comedies do not contain similarly clear-cut themes in the keyword lists compared with Behn’s earlier works. However, it is worth noting that the overall list of statistically significant (positive) keywords is greater for the late plays (93 keywords) compared to the early (71) and middle play lists (53). This potentially supports the distinctiveness of these particular works observed in the cluster and PCA results, and suggests Behn’s later writing may have been more experimental and diverse in subject-matter and style.

**Second-Person Pronouns**

The wordplot and keyword analyses provides insight into the lexical features underpinning the exploratory stylometric analyses. In this section, some of these lexical items, second-person pronouns, are examined in more detail, to evaluate their apparent association with particular plays and particular time periods. In the PCA results, *you* and *thou* distribute differently. *Thou* (*thou, thee* and *thy*) occur in the top-left quarter of Figure 5, suggesting a link with Behn’s early plays. *You* (*you, ye and your*) occurs in the
bottom-right, a space associated with Behn’s middle and late plays. The keyword analysis corroborates this distribution.

The chronological distribution of the second-person pronouns in Behn’s comedies parallels their historical evolution in the English language. By the 1650s, *thou* was used with interpersonal topics, ‘within the ‘charmed circle’’ of a relationship, linked to proximity, heightened emotions and intimacy (Lass, 1999: 153). *You* was the neutral, distal form, used for less emotionally-charged interactions, and was the dominant variant in formal writing. By the end of the century, many speakers had a non-variable grammar, with a generalized *you* (Lass, 1999: 153).

Their affective, pragmatic meanings make the second-person pronouns a productive resource for dramatic writing. However, Behn uses *thou* less frequently over time, suggesting that pragmatic affordances are less appropriate in her later works (Figure 6) (there is also the possibility that her idiolectal grammar shifted towards the dominant *you* form, although she does not cease using *thou* completely).

Figure 6: Frequency of YOU and THOU forms (per 1000 words) in Behn’s comedies (presented in chronological order). Trendlines added.

Collocate analysis, using AntConc, sheds light on the interpersonal functions of the pronoun forms in her plays. Looking at the top-20 most frequent collocates, both pronoun forms co-occur with other high frequency function words, including *art/are,*
has/have, would and shall. Thou also occurs with a 3-gram cluster: thou art a [noun]. The cluster occurs across Behn’s comedies, although it is most frequent in The Rover and Town Fopp (both 1576-7). The most frequent complements of the 3-gram are fool, fellow, man, woman, rogue and coward, indicating its pragmatic function as an affective marker of intimacy and/or proximity; e.g.:

Charles: I do not wonder thou can lie, for thou art a Coward! Did not thou draw upon him? Is not thy Sword yet out? Did I not see thee fierce, and active too, as if thou had dared?

Sir Timothy Tawdry: Why he is gone Sir; a Pox of all Mistakes (Town Fopp, 3.1)

Collocates of you indicate its formal and distal meaning, such as the negative politeness address form sir e.g. ‘I humbly thank you, Sir’ (Dutch Lover), which indicates a more polite and deferential relationship between speaker and address, in line with seventeenth-century norms of use. The results indicate that Behn uses second-person pronouns according to their contemporary pragmatic functions. The decline of the highly intimate and emotive thou and the increase in the decorous and polite norm, you, may be interpreted against the larger shift in which characters’ change from inward-looking, emotively-charged subjectivity to ones concerned with social interactions and positioning (i.e. linked to social role, status and so forth). The changing distribution of
you/thou is only one facet of the stylometric periodization of Behn’s dramatic writing. Nevertheless, the analysis offers an impression of how large-scale changes are linked to, and intersect with, the formal construction of Behn’s dramatic worlds.

Stylochronometric perspectives on The Young King

The stylochronometric analysis strongly indicates evidence of change in the language of Behn’s twelve comedies. Before the analysis applies these findings to the questionable dating of The Young King, it is necessary to corroborate the chronological change in Behn’s four non-comedic works. The Young King is a tragi-comedy, and thus the relationship between time-period and genre needs to be established. The tests using MF-FWs and MFWs were recalculated, this time using wordlists derived from the whole Behn drama corpus. The results of the cluster and PCA analyses are shown in Figure 7 and Figure 8, respectively.

Figure 7: Hierarchical cluster analysis; 135MF-FWs. All Behn plays excluding The Young King.

Figure 8: Hierarchical cluster analysis; 300MFWs. All Behn plays excluding The Young King.

Despite the additional of genre, chronology is still an organising factor. In the MF-FWs and MFWs cluster analyses, three main clusters broadly correlate with the early, middle and late periodisation of Behn’s comedies in each dendrogram. The first cluster (reading top-to-bottom) incorporates The Widdow Ranter (1689-90) into the group of late
comedies. The second cluster adds the tragicomedy *Forced Marriage* (1670-1) and, albeit at a slight remove, the tragedy *Abdelazer* (1676-7) to the early plays. The one exception to the anticipated temporal distribution is *The Dutch Lover*, which in the MFWs analysis effectively swaps places with *The Town Fopp*.

Figure 9: PCA results; 135 MF-FWs. All Behn plays excluding The Young King.

Figure 10: PCA results; 300MFWs. All Behn plays excluding The Young King.

The distribution of the plays in the PCA analysis is similar to the comedy-only analysis. The second component, in particular, can be interpreted as reflecting time-period: the earlier plays score higher on this axis than their later counterparts. Genre (or sub-genre) offers a plausible explanation for the first component (x-axis). The more overtly comic plays, including *The Rover*, have a high first component score, whereas tragedy *Abdelazer* and tragi-comedies *The Forc'd Marriage* and *The Widdow Ranter* score low for this component. Overall, the cluster analysis and PCA results suggest that time is a strong signal in Behn’s writings, even in a mixed-genre dataset. On this basis, a stylochronometric analysis of *The Young King* may provide a valuable perspective on its possible date of composition.

**Results: The Young King**

The analysis treats *The Young King* as a questionable text, comparing it with Behn’s other plays. The working hypothesis is that, if the play has an early composition date it
will cluster with Behn’s early plays. A positioning elsewhere may indicate a later date of composition or, at least, of revision. Figures 11-14 show the stylistic properties of *The Young King* based on the MF-FWs and MFWs derived from all of Behn’s plays.

Figure 11: Hierarchical cluster analysis of all Behn's plays. MF-FWs. The Young King added.

Figure 12: PCA of all Behn's plays. MF-FWs. The Young King added.

Figure 13: Hierarchical cluster analysis of all Behn's plays. MFWs. The Young King added.

Figure 14: PCA of all Behn's plays. MFWs. The Young King added.

The results for the MF-FWs and MFWs suggest that *The Young King* may not be as early as Behn’s dedicatory epistle declares it to be. The play distributes similarly to *Abdelazer*, Behn’s tragedy dating from 1676/7. In the cluster analyses, *The Young King* groups firstly with this play, and then with the other pre-1677 plays. In the PCA analyses, *The Young King* and *Abdelazer* are aligned on both components; that is, with genre (x-axis) and chronology (y-axis).

*The Young King* does not, however, cluster as closely with the early tragicomedy *The Forc'ed Marriage* (1670-1), as might be expected if *The Young King* was written around the same time (other examples of Behn’s plays do group in this way; e.g. the comedies *The Rover* and *The Second Part of Rover*). Thus, the similarity with *Abdelazer* may arise from similarities in time-period, not just genre, entailing that *The Young King* was written, or at least revised, contemporaneously with this play (1676-9).
A finer-grained perspective can be achieved by analysing the plays from the 1670s only. The PCA results align *The Young King* with the early plays along the x-axis (genre), but situate it with the comedies from 1676-9 on the y-axis (time), and not with Behn’s plays written at the beginning of the decade (Figure 15 and Figure 16).

Figure 15: PCA of Behn’s 1670-79 plays; 135MF-FWs. The Young King added.

Figure 16: PCA of Behn’s 1670-79 plays; 300MFWs. The Young King added.

The results offer little evidence to support the view that *The Young King* – in its extant, published form – was the earliest of Behn’s plays. Instead, the stylochronometric analysis suggest the play is most similar to *Abdelazer*, which points to a date of composition, or at least revision, of the mid-to-late 1670s.

**Results: Zeta analysis**

To conclude the analysis, the purported periodizations of Behn’s plays, and in particular the dating of *The Young King*, are tested using a supervised analytic method, zeta, which uses words of mid-range frequency. The first zeta test compares all plays from the early and middle groups. As shown in Figure 17, the test segments distribute more-or-less as expected, although some of the early segments are quite widely dispersed.
These segments are from *Town Fopp*, a play from 1676-7, that sits at the cusp of the early and middle period, and therefore might be seen as a transitional work.

Figure 17: Zeta Analysis of Behn's early and middle plays (all genres); at least 1 occurrence in > 60% in base-set and no more than 0 in < 60% in counter-set. 3000-word segments. The Young King added.

Zeta suggests *The Young King* is stylistically similar to those written earlier in Behn’s career, supporting the MFWs and MF-FWs analyses. To produce a more fine-grained picture, a second zeta analysis was run using plays written in the 1670s only. As zeta requires discrete play groups, the works were divided according to their date of first performance: *The Forc’d Marriage, The Amorous Prince* and *The Dutch Lover* for the first-half of the 1670s, and *The Rover, Sir Patient Fancy* and *The Feign'ed Curtizans* for the second half. Due to its apparent similarity with *The Young King* (on the basis of genre and/or time-period), *Abdelazer* is not included in the zeta-word calculations to avoid undue skewing the data. Instead it is added alongside *The Young King* as a “questionable” text. *The Town Fopp* is excluded due to its ambiguous placement in the previous zeta test.

Figure 18: Zeta analysis of Behn's plays 1670-1679. The Young King added. At least 1 in up to 100% segments in base-set; no greater than 0 in up to 100% counter-set. 2000-word segments. 500 zeta words.

Figure 18 suggests that the test provides a fairly robust measure of the stylistic differences between the two temporal sub-groups. The test segments for each group of
plays occur in the expected area. In this analysis, *The Young King* groups with the 1670-4 plays, rather than with those from the end of the decade. Several segments cluster tightly together and overlap with those from the 1670-4 plays, and others occur more towards the periphery. This suggests, on the face of it, that *The Young King* may indeed date from a period early in Behn’s career, supporting her claims in the dedicatory epistle.

However, *Abdelazer* – a play that was performed and thus purportedly written – in 1676-7, the same year as *The Rover*, represents the proverbial spanner, as the play also groups with the 1670-4 plays and, hence, with *The Young King*. The difficulty is therefore one of interpretation, in terms of the factors shaping the distribution of these texts. Without the complication of *Abdelazer*, the dating of *The Young King* might be fairly confidently attributed to the early 1670s on the basis of the zeta results. Indeed, a zeta test including *Abdelazer* in the 1677-9 text-set places *The Young King* with the earlier plays, although the test segments suggest this test was not quite as robust as Figure 18. Another test included *Abdelazer* in the 1670-4 text-set, which makes for a more definitive comparison of genre than time-period. *The Young King* again positions firmly among the 1670-4 group (Figure 19).

Figure 19: Zeta analysis of Behn's 1670-4 (plus Abdelazer) and 1677-9 plays (ex. The Town Fopp). At least 1 in up to 100% segments base-set; no greater than 0 in up to 100% counter-set. 2000-word segments. 500 zeta words.
The zeta analysis thus supports the MFWs analysis, which identifies stylistic similarities between *The Young King* and *Abdelazer*. However, the zeta tests also associates the play with the 1670-4 text-set. This could be attributable to genre as much as time-period. The zeta findings contrast with the cluster analysis and PCA results, to some extent, as these identified similarities between *The Young King* and the post-1677 plays.

What these results show, perhaps, is the highly significant relationship between time-period and genre within the larger context of an author’s style, and the difficulty of differentiating these within a small pool of data. The highest scoring items in the zeta word list overlap with those identified in the keyword analysis. Early zeta words (excepting *Abdelazer*) include prince, revenge and kill, and the 1677-9 zeta words include whe, whore, aye, agad, pox, and women. The overlap attests to the intricate relationship between subject-matter, genre, and time-period in the chronology of Behn’s style.

On balance, the stylochronometric results suggest that *The Young King* was written early in Behn’s career, perhaps contemporary with the 1670-4 plays. Its similarity with *Abdelazer* may arise from Behn re-visiting her early style when writing that play, half a decade later. After all, stylochronometric studies attest to the bi-directionality of authorial style change (van Hulle and Kestemont 2016). (There is, of course, the other
possibility that *Abdelazer* was in fact written much earlier in Behn’s career, which would make the grouping a two-fold association of factors).

However, the similarities between *The Young King* and *Abdelazer* also permit an inverse interpretation: namely that *The Young King* was written or revised in 1676-7, contemporaneous with the tragedy. Given that the exploratory PCA results do not suggest a strong similarity between *The Young King* and any of her earliest plays, combined with the fact that *Abdelazer* persistently groups with *The Young King* in all the tests, the later date of composition, or revision, appears the more plausible interpretation of the two.

**Conclusion**

Stylochronometry provides an empirical means to explore and evaluate the linguistic validity of critical assessments concerning changes in authorial style. The findings of this exploratory study of the dramatic style of Aphra Behn suggest that, over the course of her twenty-year career, Behn’s writing developed in three identifiable broad phases. The characteristics of these stages are linked, in part, to changes in genre and the associated construction of the dramatic world. Her earliest plays are characterised by abstract, emotive lexis, with female characters described in terms of their chastity and marital availability. Her middle plays shift towards lexis associated with the social world, reflected in derogatory labels (e.g. *rogue*) as well as a greater attention to
interactivity, reflected in spoken language features such as interjections. Some of these features are maintained in Behn’s late plays, but others shift again, with an increase in male social roles, and a focus on patriarchal activities and technologies. The changes articulate a broadening dramatic scope from domestic and interior concerns to more diverse, and socially stratified interactions and events. The identification of three stylistic periods provides a baseline that can be incorporated into future attribution work. As a first indicator of its profitability, the stylistic analysis suggests that *The Young King*, in its known form, was revised or written later than Behn acknowledges in her dedicatory epistle.

The present investigation has concentrated on the style changes of an individual writer, and therefore it is not possible to evaluate whether the extent and nature of these changes are typical of Restoration drama more broadly. The findings highlight other uncertainties. For example, it is unclear if comparable changes occur in Behn’s writing in other genres. Moreover, whilst the qualitative analysis provides a sense of the linguistic basis of the quantitative patternings of her plays, it is not clear what factors moved her work along these specific trajectories (cf. Stamou 2008). There is scope for much greater scrutiny of chronological style change, both for specific authors and literary communities, using stylometric techniques.
Funding

This work was supported by the Arts and Humanities Research Council [AH/N007573/1]. The textual data used for the present analysis will be available through the public-facing project website.

Acknowledgements

I wish to thank project colleagues (Prof. Elaine Hobby, Dr Gillian Wright, Dr Claire Bowditch and Prof. Hugh Craig) and the anonymous reviewers for their generous and constructive feedback on earlier versions of this paper.

References


Available from:


https://doi.org/10.1080/09296174.2017.1405719


