Shakespeare's Language: Styles and Meanings via the Computer

Jonathan Culpeper
Lancaster University, UK

Dawn Archer
Manchester Metropolitan University, UK

1. Shakespeare's language and the digital revolution

We celebrate Shakespeare because of what he created in language. It is strange then that a trip to any library will find shelves groaning with the vast quantity of literary criticism, but relatively few volumes on his language. This is not to say that there is no scholarship of significance on Shakespeare's language. Aside from some major, and indeed recent, works on grammar (e.g., Blake 2002), lexis (Crystal and Crystal 2002), and phonology (e.g., Crystal 2016), as well as collections of note (e.g., Salmon and Burgess 1987; Adamson 2001) and a steady flow of articles, there is a considerable quantity of insightful linguistic commentary in the form of footnotes in edited editions. The problem with the latter is that it does not amount to a coherent research programme. Moreover, it can also be very difficult to retrieve a specific insight tucked away in a footnote. A more general problem with scholarship relating to Shakespeare's language is its split personality, encompassing as it does both linguistic and literary approaches. Consider some of the topics mentioned in the chapter titles of the recent volume, *The Cambridge Companion to Shakespeare's Language* (edited by Lynne Magnusson with David Schalkwyk, 2020): 'Shakespeare's creativity', 'performative power', 'figures of speech', 'rhetoric', 'literary experience', 'writing for actors' and 'popular culture'. Whilst all of these are relevant to Shakespeare's language, they would not be recognized by many linguists as having much to do with language. Conversely, some linguists, especially historical linguists, seem to deny that Shakespeare's language is a literary artefact. Instead, their focus is on the role of Shakespeare's language in the context of the study of the history of English. In other words, Shakespeare's language becomes a language dataset that can cast light on general linguistic issues pertaining to the English language. The work of Salmon (e.g., 1987a [1965], 1987b [1967]) is a landmark here. Of course, tensions between linguistic and literary approaches are all too familiar to the stylistics scholar. Modern literary stylistics is often traced back to the 1958 Indiana Style Conference, where Roman Jakobson concluded his intervention with a statement that was to become the unofficial manifesto of future stylisticians:

> […] a linguist deaf to the poetic function of language and a literary scholar indifferent to linguistic problems and unconversant with linguistic methods are equally flagrant anachronisms. (Jakobson, 1960: 377)

More than 60 years later, there are still scholars who are not hearing this point. A raison d'être of this very journal is to bring together those approaches, and indeed this is the line we pursue in this special issue. There are inevitably differences of emphasis across the papers, but we take the view that each approach brings something of value to the table, and where they can be successfully synthesized, the whole is greater than the sum of the parts.

A specific problem that Shakespeare's language encounters is that the bulk of his output is plays, and the language of plays is 'the neglected child of stylistics' (Culpeper et al., 1998: 3). The situation has not changed much since 1998, as a glance through the papers of this journal will attest. Culpeper et al. (1998: 3-4) argue that part of the reason for this neglect might be the fact that spoken language, which the language of plays resembles to an extent,

has been seen as an ill-informed, unstable form of language with little value. Some traditional 20th century literary-critical movements (e.g., New Criticism in the USA and Practical Criticism in Britain) were prepared to reprieve Shakespeare and some other Elizabethan playwrights on the basis that their plays were often written in verse – they constitute 'dramatic poems'. But note that the move here is simply to deny the speech-like, dialogic characteristics that the language of plays evidently has, rather than consider them as worthy of analysis. Linguists had traditionally shied away from the perceived 'messiness' of spoken language, preferring to deal in neater abstractions than what happens when language is used in spoken conversation. Chomsky (1965: 3), for example, thought actual language use represented 'irrelevant conditions'. However, for some years, the winds of change have been blowing through scholarship. Although not setting a new direction for all, they have produced a change in tack for at least some. The late 1970s and 1980s saw developments in discourse analysis, conversation analysis and pragmatics, methods of analysis developed by linguists to deal with face-to-face interaction.[1] These equipped stylisticians with tools to analyse the meanings of utterances in fictional dialogue, and galvanised the early work on the stylistics of drama (e.g., Burton, 1980, Short, 1981, Simpson, 1989, Herman, 1995, Culpeper et al., 1998). Scholarship on Shakespeare's language has also felt these effects – witness, for example, the book-length studies by Rudanko (1993), Magnusson (1999) and Kizelbach (2014).

We would point to a further change in scholarship that is producing positive results in general, and that is the advent of digital methods. Digital humanities, or more generally digital scholarship, has been revolutionising the way we store, retrieve, analyse and visualise texts. Corpus linguistics is part of this revolution. McEnery and Hardie (2012: 1) suggest that we can 'reasonably define corpus linguistics as dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions'. Corpus linguistics, like pragmatics and discourse analysis, is strongly concerned with authentic language use, not abstractions. The papers in this issue fit the specific area of corpus stylistics, a sub-field that unites corpus methods with literary and linguistic concerns. This area has its stylistics roots in the quantitative approach to the analysis of style, as elaborated in, for example, Leech and Short (1981). What is distinctive in corpus stylistics is the mechanical advantage afforded by computers, as a computer can do in moments what one scholar might struggle to do in a lifetime. Corpus stylistics is concerned with a quantitative methodology, computers, styles and – crucially – meanings. This area continues to grow within stylistics, recently having been given a boost by the book *Corpus Stylistics: Theory and Practice* (McIntyre and Walker, 2019).

Computational techniques are not unfamiliar in the context of Shakespearean scholarship, but they are strongly associated with studies of authorship attribution (and digital humanities more generally). There is no direct connection with corpus linguistics. Authorship attribution draws on a branch of literary computing which is sometimes referred to as 'stylometrics'. A pioneering work here is Burrows (1987), analysing Jane Austin's style. Burrows (1987) recognised that a key merit of computer-driven analyses is that they could encompass a whole work, thus avoiding the selection of extracts for qualitative analyses, as is typical in many stylistics studies. As he put it, 'It is a truth not generally acknowledged that, in most discussions of works of English fiction, we proceed as if a third, two-fifths, a half of our material were not really there' (Burrows, 1987: 1). Interestingly, he uses the same kind of statistical methods, cross tabulations and chi-square, that underpin the notion of 'keywords' deployed by corpus linguists today (see, e.g., Murphy et al., Archer and Demmen and Archer and Gillings, this volume). Numerous studies of Shakespeare have proceeded in this kind of tradition. Craig and Kinney (2009), for instance, is a collection of illuminating and sophisticated studies in which computer analyses are designed to identify and compare

Shakespeare's and other writers' styles. However, what is missing is the broader contribution to stylistics, as most studies in this area focus only on authorial style. Authorial style is of course an important issue within stylistics, but it is not the only issue. Moreover, and crucially, author attribution scholars do not treat language as something that has structure and meaning. For them, language affords the possibility of tracking patterns, strings of reoccurring items which can be used to identify authors. For the corpus linguist, as indeed the corpus stylistician, words pattern in such a way that they form chunks of language (expressions or constructions), or broader notions such as idiolects and registers, and these patterns construct meanings. These patterns and their meanings are the points of interest for the stylistics scholar.

2. The Encyclopedia of Shakespeare's Language project: Texts, tagging and tools

In spite of advances in digital methods, no study of Shakespeare has deployed the full array of methods used by corpus linguists to interrogate large electronic collections of text, that is, corpora.[2] The AHRC-funded (AH/N002415/1) *Encyclopedia of Shakespeare's Language* project (2016-2019) (hereafter *ESL*) aims to fill this gap and bring scholarship on Shakespeare's language fully into the 21st-century. This special issue is one step towards this goal. It is not intended as a traditional collection of studies on Shakespeare's language. What makes it distinctive is its corpus-based nature.

Being corpus-based implies both a particular method for revealing meanings, and a particular theoretical approach to meaning. The question 'what does X mean?' is pursued through another question: 'how is X used?' We use computers to identify patterns of use across Shakespeare's works, some of which would be difficult for the human reader to encompass on such a large scale. We can see, for example, that the uses of the word *bastard* were very different from those of today. Today, it appears in colloquial language as an expression of abuse, or more often mock abuse (i.e. banter), its most frequent co-occurring words (collocates) in the 100 million-word the British National Corpus being *you*, *lazy, fat* and *dirty*. 400 years ago, it occurred most densely in informational, instructional registers, and co-occurred with words for plants, especially flowers. It was primarily a technical term for hybridity. It could be used to offend (co-occurring with, for example, *brat* and *cockold*), but this was a minor use. So, what does this mean for the interpretation of a line like 'Thou cursed bastard' in *King Lear* (2.1). To interpret *bastard* as a colloquial generic term of abuse or banter is simply wrong. That would be importing today's notions. The more technical sense of being an unnatural hybrid is key in Shakespeare's period. From that powerful negative associations can flow, including that you are socially inferior, that because you are of unnatural birth you are not fit for the kingdom of heaven, and so on. Furthermore, the ability corpus methods give to track words across broader contexts of use allows us to, for example, reveal the linguistic construction of characters or genres (e.g., tragedy, comedy or history). For instance, Juliet, in *Romeo and Juliet*, differs, statistically speaking, from the other characters in the play in her overly frequent use of words such as *if*, *yet* and *or*. Such words articulate her continual anxieties: 'But if thou meanest not well' (2.2); 'Is thy news good, or bad?' (2.2) (see Culpeper, 2002).

It goes without saying that, in order to achieve the kinds of analyses thus described, we need texts in machine-readable form, enriched by tagging or annotation and computer programs to perform analyses. Part of the mission of the *ESL* project is to create such resources and make them freely and publicly available. We briefly describe them here (more detail can be found in the project's website http://wp.lancs.ac.uk/shakespearelang/) and associated publications. One may think that such Shakespeare resources exist already. However, much of what is readily available are edited modern editions of Shakespeare (and

even these may have copyright restrictions). Sometimes text in these editions has no historical antecedent, being created by a modern editor through the collation of a number of different historical texts. In addition, the original spelling, punctuation, and sometimes whole lines and speeches are freely changed, deleted or inserted, thus impacting upon meaning and interpretation. We wanted texts that were a faithful electronic transcription of an actual historical original. This is not to say that there is one historical text that is a true instance of Shakespeare's writing. Authors did not have the same kind of control over publication that they often do today. Many of Shakespeare's plays were published after he died in 1616 in the First Folio of 1623. Moreover, authorial collaboration was common, and the processes of printing promoted interpolations and adjustments by many other hands. Still, at least one can avoid the layer of modern editorial interference. Internet Shakespeare Editions (https://internetshakespeare.uvic.ca/), based at the University of Victoria (Canada), is perhaps the only place in the world to have accurately transcribed a comprehensive set of (original spelling) works attributed (at least in part) to Shakespeare, including the First Folio, the Quartos, and the poetry.[3] They were kindly made available to the project, and they became base texts of the *Enhanced Shakespearean Corpus,* a richly annotated set of three corpora built for the analysis and comparison of Shakespeare's language. The *Enhanced Shakespearean Corpus* consists of three sub-corpora:

Enhanced Shakespearean Corpus: First Folio Plus (hereafter, *ESC:Folio*)
Enhanced Shakespearean Corpus: Comparative Plays (hereafter, *ESC:Comp*)
Enhanced Shakespearean Corpus: EEBO-TCP Segment (hereafter, *ESC:EEBO*)

These are openly available (at present via Lancaster University's CQPweb server and in future as full downloads; see http://wp.lancs.ac.uk/shakespearelang) under a Creative Commons licence. At the time of writing two additional subsidiary components, one capturing the Quartos the other the Poetry, are under construction. *ESC:Folio* contains 38 plays – the 36 in the First Folio, with the addition of *The Two Noble Kinsmen* and *Pericles Prince of Tyre* – and amounts to approximately 1 million words. It represents the 'canon' as far as the plays are concerned (as already noted, we cannot assume that Shakespeare solely authored all of these) (for more detail, see Culpeper et al., in preparation). *ESC:Comp* contains 46 plays by 24 playwrights (22 of whom are named, two of whom are anonymous), with first production dates ranging from 1584-1626 (compared to Shakespeare's plays, written circa 1590-1613). It is similar to *ESC:Folio* both in size and in its proportions of comedy, history and tragedy, thereby facilitating comparisons (for more detail, see Demmen 2020). *ESC: EEBO* was compiled as a broad corpus to assist in contextualising Shakespeare's language. It comprises texts from *Early English Books Online* (Text Creation Partnership), amounting to some 300 million words spanning the 80-year period 1560-1639 and incorporates diverse genres. A key feature is that a genre classification scheme has been applied its 5,900 texts (for more detail, see Murphy, 2019).

All play-texts in the *ESC* are marked up and annotated with XML tags (see Bray et al. 2008; Hardie 2014). Each utterance is marked with an opening speaker ID tag and a close tag. One attribute of the speaker ID tag is the speaker label in its original format in the text. Act and scene boundaries, stage directions, front matter, end matter and paratext, e.g., prologues and epilogues, are also marked with XML tags. The play-texts have regularisation of their spelling variation (without this, a search on, say, *sweet*, would fail to find *sweete*). This was done with the help of the software tool VARD 2 (i.e. VARiant Detector; the echo of 'bard' is intended) (see http://ucrel.lancs.ac.uk/vard/about/). This tool – using a lexicon of spellings, rules about spelling and statistics – regularises spellings in manual or automatic mode. The *ESC:Folio* was regularised to a high level of accuracy by running VARD 2 in manual (word-

by-word) mode (it can on most occasions suggest regularisation options in order of likelihood, from which the human operator approves one). *ESC:Comp* and *ESC:EEBO* have undergone some regularisation through PHP scripts (notably to join open compounds which are now typically closed, e.g., *it self > itself*) and VARD 2 in automatic mode at a 70% confidence level.

All play-texts have been enriched with grammatical part-of-speech tags using a customised version of the Constituent Likelihood Automatic Word-tagging System (CLAWS) (see Leech et al., 1994; http://ucrel.lancs.ac.uk/claws/). CLAWS tags are alphanumerical 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 addition, for the *ESC:Folio* alone, every word was manually checked for accuracy at the highest level of the tag (e.g., a word tagged NN1 and another NN2 were both checked that the initial 'N' (i.e. noun) is correct). The play-texts have also been annotated for semantic meaning, using the UCREL 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 fields (see further http://ucrel.lancs.ac.uk/usas/). Although USAS has been successfully used for semantic analysis of historical texts, it needs to be treated with caution, as the USAS semantic classification system was developed for late 20th century English and is prone to errors. Such tagging is clearly an area for further development. Finally, the play-texts in the *ESC: Folio* and *ESC:Comp* have been annotated with XML tags for social categories. The social categories are listed in Table 1. The categories relating to a character's status/ social rank draw upon the scheme developed by Archer and Culpeper (2003) which reflects 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 reworked to capture particular Shakespearean features (e.g., the category Supernatural Beings was added to account for the ghosts, gods, fairies, etc.).

Table 1. Social annotation categories

Field	Feature marked	Possible values
1	Speaker(s)	Singular (s) or multiple (m)
2	Speaker ID tag	Character's name
3	Gender of speaker	Male (m), female (f), assumed male (am), assumed female (af), problematic (p)
4	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 (8)

The principal tool used in the *ESL* project, and indeed the papers of this volume, is CQPweb, a web-based tool which was originally designed to replicate the user-interface of the popular BNCweb tool, but can be used with *any* corpus (see in particular Hardie, 2012). It is reasonably user-friendly (it is widely used with students) and is exceptionally powerful. Moreover, it is able to handle huge corpora and is stable. Andrew Hardie is continually updating CQPweb. In due course, some of the specific tools designed for the project (and which sit on top of CQPweb) will be integrated into it.

3. The upcoming papers

Of the six papers in this special issue, four are overtly corpus linguistic in their approach, that is, they draw upon statistical measures as a means of demonstrating what Shakespeare's grammatical choices and/or word choices tell us about 'his' authorial style and/or his representations of nationality, gender and deception. The two remaining papers do something different. The fifth paper, which is more pedagogical in orientation, examines the difficulties associated with Shakespeare's language and discusses a possible role for corpora. The sixth paper, which is more literary in orientation, shows that even when phenomena are so rare as to preclude the inferential statistics that underpin many standard corpus linguistic techniques (e.g., collocations, keywords), corpus methods can make a positive contribution.

The first paper, by Culpeper and Findlay, explores early modern understandings of Celtic (specifically, Welsh, Scottish and Irish) identity terms across an 80 year period (1560-1639). The paper focuses on the Celtic characters in *Henry V*: the Welshman, Captain Fluellen; the Scotsman, Captain Jamy; and the Irishman, Captain Macmorris. The authors' desire to uncover contemporary understandings of such Celtic identities – how Shakespeare's contemporaries generally viewed the Welsh, Scots and Irish – is part of their aim to discover how the stereotypes held by contemporary readers differed from those of modern readers, with consequent interpretative consequences for the cognitive construction of characters. The authors achieve this by studying the collocates of Celtic-related terms, that is, the words that regularly co-occurred with *Scottish*, *Irish* and *Welsh*, thereby helping to colour their meanings. Collocation patterns are extracted from the 300 million word *ESC:EEBO*. Results flowing from the analyses of collocates, reveal, for example, the fact that the Irish were considered *wild* and *savage*, but also that the word *Irish* had one particular positive use – when modifying the word *rug*. In discussing their findings, the authors incorporate literary critical discussion, notably on 'nationhood'.

The second paper, by Murphy, Archer and Demmen, focuses on the social aspects of characterisation. It is an example of how the addition of annotation to the resources created by the *ESL* project enable users to examine characters' gender and/or social status. This paper is one of two papers that draw upon the keyword method, which involves generating lists of words that are overused or underused, statistically speaking, in one dataset in comparison with another. Specifically, the authors examine keywords characterising female and male speech across the 38 plays of *ESC:Folio*, and then broaden their scope by examining how this speech varies according to characters' social status (high or low) and different genres (comedy, history and tragedy). They achieved this through semantic category analysis and collocational analysis of the gendered forms identified (*girl*, *woman*, *lady*, etc.). This innovative approach enabled the authors to show that the keywords of middling to high status female characters are more focussed on local figures of power and authority, whilst male are more focussed on the national figure (the male monarch) – a finding that fits the hierarchical and patriarchal nature of the historical period they deal with (Nevalainen and Brunberg, 2003). Female characters (of all statuses) were also found to be more 'relational' in their use of first-person pronouns than male characters. Regarding genre, high-ranking male characters in comedy were found to speak significantly more about women than their counterparts in histories and tragedies. When it comes to gendered language forms, the authors found one exception to the fact that there are more similarities than differences between female and male characters: male characters use gendered terms that female characters do not (e.g., *womanish*), and they tend to be used to denigrate women.

The third paper in this special issue, by Archer and Gillings, explores five specific Shakespearean characters (Aaron, Tamora, Iago, Lady Macbeth and Falstaff) as a means of demonstrating how a particular issue, the depiction of deception, can be profitably pursued through corpus methods. The authors combine the keyword methodology with another

quantitative method that involves determining the extent to which the five characters use language associated with real-world deception. They also engage in qualitative analysis, as a final step, paying particular attention to where keywords and/or potential deceptive indicators cluster (or co-occur) in the interaction(s) of two characters: Aaron and Falstaff. The authors emphasise the importance of context, throughout their paper. Indeed, rather than making claims and predictions based on the different frequencies of a collective set of features (as some deception related papers do), they argue that 'deception is best investigated by careful examination of features *alongside* their wider context within a scene (or, even, a turn)'. They also identify how dramatic devices, such as soliloquies and asides, can be used by Shakespeare (and other playwrights) 'to keep their audiences *in the know* (when other characters remain in the dark)'. In addition, Archer and Gillings' work provides insights to the playwright's ability to depict traits associated with deception.

The fourth paper, by Hardie and van Dorst, investigates grammatical style in Shakespeare's plays by focussing on 15 grammatical features of stylistic interest. This study is possible only because of the project's creation of the first grammatically tagged and manually checked version of Shakespeare's plays. Using the *ESC:Folio* and the *ESC:Comp*, the authors present a method that 'steers between the narrow focus of close reading and the naïvely quantitative metrics of authorship analysis'. This method involves retrieving all instances of the fifteen features in each play via complex corpus searches, and then considering them, in aggregate and at the text level, in order to reveal their dispersion across plays and dramatic genre, and between Shakespeare and the other dramatists. The results are explored via both statistical summary and visual representations of variability. They argue that it is only in understanding the extent of variability that exists generally in the language that researchers can better appreciate the extent to which Shakespeare's grammatical style is distinct from or similar to that of his contemporaries. Some of the more significant findings include Shakespeare's grammatical style tending (especially in comedies and tragedies) towards a dis-preference for informationally-dense noun phrases (relative to other, contemporary) playwrights and a preference for tense, aspect and pronoun features, all of which are taken to suggest a greater degree of narrative focus in Shakespeare's style. The authors also found that Shakespeare tends to be markedly distinct regarding verb complement subordinate clause types. Such findings point the way not only to methodological innovation but also to further research questions concerning Shakespeare's grammatical style.

The fifth paper in this special issue, by Murphy, Culpeper, Gillings and Pace-Sigge, takes a somewhat different tack, leaning towards pedagogical stylistics. Corpus-based methods are better geared toward some areas of linguistics than others. For example, most corpus analyses rely on the 'word' as the basic focal point of analysis. This begs the question of whether this is the specific area which students generally find difficult. Following a review of Shakespeare and pedagogy, they highlight a lack of an empirical investigation as to what exactly students find problematic when they read the language of Shakespeare's plays. Their study was conducted with three groups of Shakespeare undergraduate students, in Lancaster (UK), Barcelona (Spain) and Joensuu (Finland), some of whom spoke English as a first language and some of whom spoke English as an additional language. All were asked to identify any difficulties they experienced when reading certain play extracts, by rating specific linguistic forms according to difficulty as well as discussing what they think of Shakespeare's language. Having established 'archaic words, borrowings from other languages, coinages and false friends' as the most common areas of difficulty for these particular students, the authors go on to discuss some corpus-related pedagogical solutions. The authors believe that the type of approach afforded by pedagogical corpus stylistics addresses at least some of the problems associated with traditional approaches to Shakespearean textual teaching, because it requires the active involvement of learners, and

because the main focus is on the language itself, not on the context alone, yet the language is treated in a contextualised fashion.

The sixth paper, by Findlay, concentrates upon the last words of plays and in particular epilogues, a specific kind of paratext. As these epilogues, and indeed last words, are a rare phenomenon, not even occurring in all 38 plays, Findlay is careful to note the difficulties in using corpus methods when the number of words are few, before then outlining a possible way forward. This involves using the <epilogue>-tag to search for and then create a sub-corpus of epilogues within CQPweb, from which absolute (raw) frequencies, as well as dispersion scores for particular words can be generated (using #LancsBox, v.5). Findlay explores, for example, how the future markers *will* and *shall* and how pronouns used in epilogues differ from their use in the rest of plays, discussing, for example, the importance of inclusivity in terms of address. Single actors spoke epilogues, it is pointed out, on behalf of the company, rather than on behalf of a single author, and this might account for the fact that plural pronouns like *our* and *we* rank higher in the last words of plays than they do in the rest. Findlay also engages in a specific case study of the language of *As You Like It,* exploring the relationship between the last words and the play which precedes them. Findlay argues, for instance, that the actor's use of certain terms such as *good* point to values which are connoted throughout the play – they are common threads stitching the body of a play and its epilogue together.

When taken as a whole, the six papers are designed to represent some of the broad array of the opportunities afforded by the new corpus resources created as part of the *ESL* project. For example, users are able to access a tagged/annotated version (*ESC:Folio*) of 38 of Shakespeare's plays, and – where relevant – restrict their searches in order to focus on specific words, grammar, semantic fields, status, gender, etc. They also have access to a matched reference corpus of contemporaneous plays (*ESC:Comp*) and wider language context of Early Modern English in the period 1560-1640 (*ESC:EEBO*). Users will have the opportunity not only to research Shakespeare's language usage, as all of the papers in this special volume do, but also to place it in its linguistic context. As Murphy, Archer and Demmen note, in closing their paper for this special issue, 'the possibilities of extending our understanding of Shakespeare's work are [thus] limited only by our imaginations'. We hope that future users make use of these resources and, more generally, corpus methods, and are inspired by the papers so that our understanding of Shakespeare's language usage grows exponentially.

Funding

The research presented in this entire special issue was supported by the UK's Arts and Humanities Research Council (AHRC), grant reference AH/N002415/1.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

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.

Blake N (2002) *A Grammar of Shakespeare's Language*. Basingstoke: Palgrave.

Bray T, Paoli J, Sperberg-McQueen CM, Maler E and Yergeau F (2008) (eds) *Extensible Markup Language (XML) 1.0.* Fifth edition. W3C Recommendation 26 November 2008. https://www.w3.org/XML/ (accessed 01.06.2019).

Burrows, JF (1987) *Computation into Criticism: Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.

Burton, D (1980) *Dialogue and Discourse*. London: Routledge and Kegan Paul.

Chomsky NA (1965) *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Craig H and Kinney AF (2009) (eds) *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.

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

Crystal D (2016) *The Oxford Dictionary of Original Shakespearean Pronunciation*. Oxford: Oxford University Press.

Crystal D and Davy D (1969) *Investigating English style*. London: Longman.

Culpeper J (2002) Computers, language and characterisation: an Analysis of six characters in Romeo and Juliet. In: Melander-Marttala U, Ostman C and Kytö M (eds) *Conversation in Life and in Literature: Papers from the ASLA Symposium.* Association Suedoise de Linguistique Appliquee (ASLA), 15. Universitetstryckeriet: Uppsala, pp.11-30. (Also available at: http://www.lexically.net/wordsmith/corpus_linguistics_links/papers_using_wordsmith.htm)

Culpeper J, Short M and Verdonk P (1998) *Exploring the Language of Drama: From Text to Context.* London: Routledge.

Culpeper J (1998) Introduction. In: Culpeper J, Short M and Verdonk P (eds) *Exploring the Language of Drama: From Text to Context*. London: Routledge, pp. 1-5.

Culpeper J, Andrew H, Demmen J, Timperley M and Hughes J (in preparation). Supporting studying Shakespeare: Structural markup and grammatical annotation enhancing a corpus of the First Folio.

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 (2014) Modest XML for Corpora: Not a standard, but a suggestion. *ICAME Journal* 38: 73-103.

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

Jakobson R (1960) Closing statement: Linguistics and poetics. In: Sebeok TA (ed.) *Style in Language*. Massachusetts: MIT Press, pp.350-377.

Kizelbach U (2014) *The Pragmatics of Early Modern Politics: Power and Kingship in Shakespeare's History Plays*. New York/Amsterdam: Rodopi.

Leech GN and Short M (1981) *Style in Fiction*. Harlow: Longman.

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.

Magnusson L (1999) *Shakespeare and Social dialogue. Dramatic Language and Elizabethan Letters*. Cambridge: Cambridge University Press.

Magnusson L with Schalkwyk D (2020) *The Cambridge Companion to Shakespeare's Language*. Cambridge: Cambridge University Press

McEnery T and Hardie A (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McIntyre D and Walker B (2019) *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press.

Neuhaus HJ (1991) Integrating Database, Expert System, and Hypermedia: The Shakespeare CD-ROM Project. *Literary and Linguistic Computing* 6 (3): 187-191.
Nevalainen T and Raumolin-Brunberg H (2003) *Historical Sociolinguistics: Language Change in Tudor and Stuart English*. Harlow: Pearson Education Limited.
Rayson P (2008) From key words to key semantic domains. *International Journal of Corpus Linguistics* 13 (4): 519-549.
Rayson P, Archer D, Piao SL and McEnery, T (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.
Rudanko M (1993) *Pragmatic approaches to Shakespeare: Essays on Othello, Coriolanus, and Timon of Athens*. Lanham: University Press of America.
Salmon V (1987a [1965]) Sentence structures in colloquial Shakespearian English. In: Salmon V and Burnes E (eds) *A Reader in the Language of Shakespearean Drama* (Studies in the History of the Language Sciences 35). Amsterdam: John Benjamins, pp.265–300.
Salmon V (1987b [1967]) Elizabethan colloquial English in the Falstaff plays. In: Salmon V and Burnes E (eds) *A Reader in the Language of Shakespearean Drama* (Studies in the History of the Language Sciences 35). Amsterdam: John Benjamins, pp.37–70.
Short M (1981) Discourse analysis and the analysis of drama. *Applied Linguistics* 11 (2): 180-202.
Simpson P (1989) Politeness phenomena in Ionescu's The Lesson. In: Carter R and Simpson P (eds) *Language, Discourse and Literature*. London: Routledge, pp. 170-193.

[1] Not all early works in stylistics eschewed spoken language, of course. To take one example, Crystal and Davy (1969), although not riding the wave of discourse analysis and pragmatics, saw styles as varieties of language that correlate with particular contexts, and accorded equal value to all varieties of language, including spoken.
[2] Although, not a full corpus linguistic treatment, there is a small degree of overlap between the *Encyclopedia of Shakespeare's Language* project and *The Shakespeare Database Project*, created by H. Joachim Neuhaus and Marvin Spevack, which catalogues Shakespeare's texts in a relational database in diverse ways (see Neuhaus 1991). However, unfortunately, the promised public output of that project, a CD-ROM, does not exist, and neither does the web version. The remnants of this project now seem to have been removed from the internet. Both Neuhaus and Spevack retired long ago (Spevack died in 2013).
[3] Digital projects in progress at the Folger Shakespeare Library producing transcriptions of Shakespeare's plays and those of other writers are warmly recommended, not least because they are of high quality.