Shakespeare and his contemporaries: Designing a genre classification scheme for Early English Books Online 1560–1640

Sean Murphy, Lancaster University

Abstract

The language of Early Modern texts can potentially reveal a lot about Shakespeare's language. In this paper I describe the creation of a genre classification scheme for a segment of Early English Books Online – Text Creation Partnership (EEBO-TCP), covering the period 1560–1640. This categorisation permits meaningful comparison of the language of Shakespeare with that of his contemporaries and makes an integral contribution to The Encyclopaedia of Shakespeare's Language project at Lancaster University. I outline the rationale behind the scheme, describe preliminary automatic genre classification work and present the prototype approach adopted for this categorisation. I also provide specific examples of classification in practice and discuss internal and external factors which influenced genre selection. I finish by suggesting how a range of scholars might benefit from this research.

1   Introduction and aims

It is well known that Shakespeare was influenced by what he read. Particular works known to have been highly influential include Holinshed's Chronicles (https://www.bl.uk/collection-items/holinsheds-chronicles-1577), and Ovid's Metamorphoses,[1] both published during his lifetime. But what of all the other output of the period? Shakespeare must have come into contact with many published and unpublished texts. To what extent did his use of language reflect that of his contemporaries? Thanks to large-scale digital collections of Early Modern English texts, it is now possible to answer such questions.

The largest and most comprehensive such collection, and therefore ideal for comparison with Shakespeare, is *Early English Books Online – Text Creation Partnership*[2] (EEBO-TCP). This digital resource of more than one billion words in over 40,000 texts dating from 1473–1700 offers researchers numerous possi-

bilities. By comparing Shakespeare's language with a suitable selection from such an immense repository of language, we can address issues such as how Shakespeare's linguistic usage compares with that of his contemporaries, and what they thought the language meant. For example, we can look at what language forms reveal about attitudes towards health and disease; what it meant to be a schoolteacher; or the significance of drinking sherry as opposed to beer.

The project outlined in this paper forms part of a larger research project, *The Encyclopaedia of Shakespeare's Language*, a three-year, AHRC-funded project based at Lancaster University.[3] The *Encyclopaedia* offers the first comprehensive treatment of Shakespeare's language using state-of-the-art corpus-based methods to identify language patterns and meanings in electronic text collections (corpora). One of the ways in which this is achieved is by comparing a Shakespeare corpus with a suitable segment of the EEBO-TCP corpus to provide valuable comparative information on the frequency and use of linguistic forms, as well as the social and stylistic flavour of words as used in the period. The Shakespeare corpus is also compared with a corpus of contemporary playwrights,[4] containing plays by Marlowe, Jonson, Fletcher, and others.

EEBO-TCP consists of complete and fully searchable transcriptions of Early Modern texts derived from digitized page images in the original EEBO collection, together with markup information indicating text structure (e.g. chapter divisions). It also contains bibliographic metadata such as year and place of publication, listed authors, and keyword terms like 'Bible', 'comets' or 'marriage'. These keyword classifications offer some indication of genre, but can be vague, e.g. 'God -- Proof -- Early works to 1800; Immortality' or non-existent, and therefore unreliable. Given the acknowledged importance of genre and register in explaining language change, the absence of a more systematic and reliable genre classification scheme is striking. The aims of this genre classification project therefore were: 1) to select a segment of EEBO-TCP more closely contemporaneous to Shakespeare; 2) to organise the texts in the segment on a meaningful and sound theoretical basis. This involved the design and implementation of a principled classification scheme of texts by genres, domains and styles. This project takes an innovative step by applying the genre classification scheme to a far more comprehensive dataset than previously achieved by classifications based on representative but nonetheless selective data.

Interest in Shakespeare's language is blossoming, judging by the increasing number of books and websites on the subject. Notable recent examples include Ravassat and Culpeper (2011), Crystal (2012, 2018) and Johnson (2014, 2019). Less attention, however, has been paid to contextualising Shakespeare's language in the period in which he wrote. There are exceptions, such as important

work behind the New Oxford Shakespeare led by Gary Taylor (2017), contrasting Shakespeare's language with that of other playwrights such as Kyd and Marlowe, among others. There is also a fascinating volume by Jonathan Hope (2014) on Renaissance conceptions of language and their expression in Shakespeare's plays. However, systematic and comprehensive comparison of Shakespeare's language with the total surviving printed output of the period has not yet been undertaken. The availability of EEBO-TCP now provides a starting point for comparison, but covers too broad a period and lacks classification by text-type or genre. This paper therefore outlines a genre classification scheme for a suitable segment of EEBO-TCP, to enable an apposite comparison with Shakespeare.

2   Creation of the genre classification scheme

The first stage in the creation of the genre classification scheme was to delimit the EEBO segment to cover only texts dated 1560 to 1640. This 80-year period was chosen for a number of reasons. First, it encompasses Shakespeare's lifetime (1564–1616). Second, it covers two 40-year periods, or two generations, a generation being defined as 40 years in the well-known *Helsinki Corpus of English Texts* (for the manual, see Kytö 1996 [1991]; see also Culpeper and Kytö 2010). Further, the two generations conveniently span 40 years either side of the turn of the century, itself significant for the change of monarch (Elizabeth I died and James I succeeded to the throne in 1603) and arguably a turning point in Shakespeare's career (see Shapiro 2011). Finally, the periods are similar to those used in other historical corpora such as *A Corpus of English Dialogues 1560–1760* (CED),[5] thus facilitating cross-corpus comparison.

Naturally, a corpus of texts covering Shakespeare's lifetime also includes Shakespeare's works, so for meaningful comparison with the Shakespeare dataset, these had to be removed. Similarly, it was also necessary to remove texts by dramatists whose works appear in the above-mentioned corpus of contemporary playwrights, though other works by the same dramatists were left in. As the focus of comparison is linguistic forms in English, I also discarded texts written entirely in Latin, Welsh, French and Italian. Though the remaining texts are primarily in English, that does not preclude a few from containing a certain amount of text in another language. The total number of texts requiring classification amounted to 5,697. The overall size of the corpus is just over 300 million words.

2.1   General approach to genres and genre classification

Literature and genres are directly related, and the study of literary genres is also related to the study of nonliterary genres (Steen 1999). Seminal genre studies such as Fowler's theoretical work, *Kinds of literature* (1982), and Swales' lin-guistically-oriented study, *Genre analysis* (1990), among many others, raise important issues related to genre, such as notions of purpose, prototypicality, fuzziness, hybridity, mixed modes and mutability. I shall briefly deal with each in turn. According to Fowler, the purpose of genre has to do with identification and communication, as opposed to definition or classification: "We identify the genre to interpret the exemplar" (1982:38). He proposes Wittgenstein's theory of family resemblance as a way of accounting for genres, which are seen as relating to one another in different ways. In this view, genres are based on liter-ary tradition, influence and imitation, inherited codes and literary relations. An alternative view is to 'identify the extent to which an exemplar is prototypical of a particular genre' (Swales 1990:52). Prototypicality carries with it the idea that some members may be regarded as more or less typical of a genre. In the latter case, this inevitably implies a certain fuzziness (e.g. Steen 1999; Lee 2001) between boundaries and even fuzzy categories not defined by necessary and suf-ficient elements. Genres tend to overlap so that "membership of one [genre] by no means rules out membership of others" (Fowler 1982:37). Hybridity of genres may produce types in which "two or more complete repertoires are present in such proportions that no one of them dominates" (Fowler 1982:183). Mixed modes (originally a Platonic description of literary representation) are forms of discourse which employ more than one channel of communication, such as narrative and dialogue. With regard to mutability, Fowler notes that "Every literary work changes the genres it relates to" (Fowler 1982:23), imply-ing that genres are in a constant state of flux. These notions of purpose, proto-typicality, fuzziness, hybridity, mixed mode and mutability inform the genre classification scheme described in this paper.

With this in mind, I propose definitions for some basic labels such as 'genre', 'domain' and 'style' (see Lee 2001 for an extended discussion of these and related terms). When we refer to 'genre', we understand a category of texts grouped according to culturally, conventionally and consensually recognised criteria which change over time and which allow for division into sub-genres (e.g. letters, essays). When we talk about 'domain', we refer to a broad subject field (e.g. religion), but not overly broad (e.g. belief), which encompasses a number of genres. When we discuss 'style' (a term which may overlap with 'register'), we refer to conventional, functional and situational uses of language (e.g. formal, literary).

Researchers classifying genres generally employ either: a text-external approach, based on characteristics such as the target audience, the author, the relationship between the two, the purpose of the text, etc.; or, a text-internal approach, based on linguistic features within the text, such as sentence length, word classes, verb tenses, etc. (see Biber 1988 and Lee 2001 for further discussion of these two approaches). The approach adopted in this project does not easily fit the internal/external dichotomy. It uses EEBO's bibliographic metadata to identify generic labels. The metadata may include linguistic items or features in the title, dedication, address to the reader or other linguistic features (internal) that the author, publisher, dedicatee or other contributor (external) uses to classify the text, such as 'A sermon preached at Pauls Crosse'. The advantage of this combined internal and external approach is that it aims to mitigate possible researcher bias when classifying texts into genres.[6] However, it is important to bear in mind that labels such as 'sermon' are not unproblematic owing to the mutabilility and differential use of generic terminology (Fowler 1982). Claridge and Wilson (2002) for example point out that sermons can be oral, written or a mixture of both, and there are unconventional forms such as satiric sermons (Fowler 1982). An interesting recent development with regard to metadata are metapragmatic studies such as Taavitsainen's (2017), diachronic study of the essay using metacommunicative expression analysis, the analysis of speech act expressions.

Inevitably, there are problems with grouping texts into genres and genres into a limited number of domains based on the internal and external criteria outlined in the previous paragraph. First, as our labels are based on a diverse set of criteria, it can be difficult to ensure heterogeneity of the genre or domain (Meurman-Solin 2001). Secondly, linguistic features often overlap and boundaries between genres can be fuzzy. Thirdly, hybrids or mixed-mode forms may cover two or more genres or even domains. In such cases, analysis needs to take into consideration other factors such as purpose and overall structure (Taavitsainen 2001a).

2.2 *Preliminary work on classifying genres*

Automatic categorisation work on identifying broad genres for EEBO documents (e.g. sermons, broadsides, etc.) has been carried out by Smith (2017), based on an analysis of category labels in the English Short Title Catalogue (ESTC) metadata. Two major problems identified included large uninformative categories, such as 'Early works to 1800' and a long list of categories containing only one item.

Preliminary work on genre identification was carried out by Tony McEnery at Lancaster University.[7] This involved computer-based identification of indicators of genre in titles (see also Schmied and Claridge 1997; Sullivan 2007; Orr 2011). McEnery decided to categorize the texts by working on their titles, which he decided had the quality of self-descriptions. The main title was used first, then the secondary title if necessary and, in some cases the third. Where these failed to suggest a categorization, he then consulted library catalogues to see how librarians had categorized the texts and found that often they had placed the book in different categories in different libraries. Using this approach McEnery's aim was to avoid applying modern classifications to early modern texts. What he did instead was put himself in the place of a person who might be browsing in a bookseller's shop in the period, working principally from the title. Table 1 shows McEnery's results sorted into primary, secondary and tertiary categories:

Table 1: Indicators of genre in EEBO-TCP texts (McEnery)

Example indicators of genre	Primary categories (10)	Secondary categories (44)	Tertiary categories (42)
acted, masque, comedie/comedy, tragedie/tragedy	plays	comedy, masque, etc.	
Virginia, colonial, plantation, America	colonial		
letter(s), pamphlet	ephemera	letters, pamphlets	
tale, fiction, legend	fiction		
life of, discourse, dialogue, defence, essaies/essay, death	non-fiction	biography, essay, factual, etc.	argumentative, dialogue, death, etc.
By the Queene/King, proclamation(s), parliament, experiments	government	royal, parliamentary, legal, science, etc.	proceedings, trials and disputes, etc.
tune, verse, song(s), lute, musicke, poems	poetry, verse and song	ballads, devotional, etc.	
God, Christian, church, epistle(s), Christ, religion, Bible	religion	Bible, epistles, etc.	
sermon(s), preached, lecture(s)	spoken	sermons, lectures, etc.	
Latin, French, Welsh	bin	unclassified, Latin, etc.	

2.3   A prototype approach

The genre classification provided by McEnery was a useful starting point, but for the purposes of *The Encyclopaedia of Shakespeare's Language*, I needed to provide a smaller number of domains and styles which would be easily under-standable for a general reader. Furthermore, the genre classification scheme needed to be relatively quick and easy to implement as the overall project was time-limited. Table 2 shows the prototype approach I adopted (based on Lee 2001):

Table 2: A prototype approach to genre

Super-ordinate	Domain	Literary
Basic level	Genre	Poetry, Play
Sub-ordinate	Sub-genre	Ode, Comedy

I decided on five super-ordinate domains, each with between four and six basic-level genres, most of which have a number of sub-genres. These essentially cor-respond to the Primary, Secondary and Tertiary categories in the McEnery scheme. However, I made some changes to the original scheme. I maintained 'Religion' and 'Government' as domains, changing the latter name to 'Administrative', as I felt it better represented all of the genres included in the domain. I promoted 'Factual' to the domain level, thereby aiming to encompass all genres and sub-genres whose content dealt in some sense with real life and events. I renamed it 'Informational', and added 'Literary' and 'Instructional' as domains (these choices were influenced by the *Helsinki Corpus*). Table 3 shows the com-plete genre classification of EEBO texts for the period 1560–1640.

Table 3: Genre classification of EEBO texts for the period 1560–1640

Domain	Genre	Sub-genres	Style
Literary	Plays Poetry, Verse & Song Fiction Miscellaneous	Comedy, History, Tragedy, Masque Ballads, Songs	Literary
Religious	Bible Catholicism Protestantism Doctrine, Theology & Governance Miscellaneous	Anti-Catholicism Church of England, Church of Scot-land, Non-Conformism Heresy, Prayer, Sin and Repentance, Christians, Devotional, Epistles, Ser-mons	Formal Spiritual
Adminis-trative	Royal Parliamentary Legal Miscellaneous	Communications and Orders, Pro-ceedings General, Proceedings and Reports Legislation and Orders, Trials and Disputes Articles, Declarations, Military, Peti-tion, Proceedings, Speeches	Formal Statutory
Instruc-tional	Philosophy Science Mathematics Medicine Miscellaneous	Experiments Anatomy Alchemy, Almanack, Analysis And Instruction, Astrology and Predic-tions, Lecture	Formal Didactic
Informa-tional	Biography Colonial Essay Letters Pamphlets Miscellaneous	Admonition, Advisory, Apologia, Argumentative, Commentary On People And Places, Death, Obituaries and Epigraphs, Dialogue, Exhorta-tion, General Lamentations Chronology, Directory, Finance and Trade, Food and Cookery, History, Language, Travel, Treatise, London, Petitions, Reportage, Satire, Wit and Humour	Informational

At the level of genres, I reduced the number of genres per domain to between four and six to facilitate subsequent analysis (mostly respecting McEnery's original Secondary category); I moved some of the original primary categories to the genre level (e.g. 'Colonial'). The Tertiary category became the sub-genres, except in the case of Religion, where I re-categorised the original Secondary level items as either genres or sub-genres to avoid having an excessively large number of genres under one domain. The aim here was, as far as possible, to achieve proportionality among the domains.

Finally, I mapped style labels (influenced by Biber 1988 and Hünig 2003) onto each domain, thus facilitating stylistic analysis. One example of such analysis has led us to a greater appreciation of the stylistic range and usage of the word *bastard*. As we might expect, the highest number of occurrences (2,002) can be found in informational texts, particularly in historical works, collocating most frequently with *son* and meaning 'illegitimate'), with a relative frequency of 21 instances per million words. However, the highest relative frequency of *bastard* with 53 instances per million (605 instances) is surprisingly found in formal didactic texts. Examples of specialised reference include: 'bastard pellitory' (a herbaceous flowering plant), "bastard saffron" (a type of tree) and 'bastard ribs', an anatomical term found in an anatomy manual by Alexander Read, a 'Doctor of Physick':[8] "The bastard ribs are in number five, so called because they are short less arched, touch not the sternum, and are softer".

For the purposes of *The Encyclopaedia of Shakespeare's Language*, we use the genre Plays as a source of comparative data for stylistic information on colloquial language.

3   *Assigning genres to unclassified texts*

The automatic classification conducted by McEnery labelled 418 (7.3%) of the texts as 'Unclassified'. These required manual classification. Below, I present a number of examples to demonstrate a range of possible criteria that may be taken into account when assigning genre to texts unclassified by automatic means.

Example 1
Domain: Literary Genre: Poetry, Verse, Song Style: Literary
In some cases, there is no indication of genre on the title page or in the dedication, though genre may be hinted at by certain words as in the following (my emphasis):

MAruaile you may, at the bolde approach of these my vnblushing **lines**, the first borne of my barren **inuention**, who begotten in my anticke age, now steps into the world to seeke some worthie Godfather: but certes if you weigh the inducing motiues, whose alliust (respect) perswading **oratorie**, tempted my triuiall **muse** to this presumption.
Acolastus his after-witte. By S.N. Nicholson, Samuel, fl. 1600–1602.
(http://name.umdl.umich.edu/A08212.0001.001)

The cumulative effect of the words suggests a fictional literary work of a poetic nature. Poetry, verse and song of the period are easily identifiable by characteristics such as stanzas, rhyme and specific metrical patterns (generally speaking). The dedication is from *Acolastus, his after-witte*, has 250 six-line stanzas and an ABABCC rhyme scheme and is largely written in iambic pentameter.

Example 2
Domain: Literary Genre: Fiction Style: Literary
Contemporary descriptions may lead us to expect a different genre to what we actually find. In *Eromena, or Love and Revenge* by Giovanni Francesco Biondi, translated from Italian by James Hayward, the latter refers to Biondi's work as "Helidorian Poesie (I mean an historicall way of Poetizing, or Poeticall manner of historizing, or displaying of the fained seeming unfained adventures and actions of persons reall, masked under the vizard of invented names:)", suggesting a work of historical non-fiction (though the names have been changed) written in a poetic style. He finishes with a short verse in which he refers to "The Plot" (pages unnumbered). The beginning of the book displays many of the features one would expect to find in a narrative mode: ancient setting (pre-Christian Berber kingdoms), identification and relationship of characters, narration using simple past and third-person forms and character description (Polimero). In addition, we find examples of the kind of textual signalling of narrative prospection which Toolan (2008) describes. These include, among others, sentences with narrative verbs of mental processing, with clause-projection: 'he knew already what it imported him to be last borne'; and sentences containing a character's represented thought: 'Metaneone the eldest brother (observing this their inclination, became iealous thereof)'.

CAttalampo King of Mauritania, had by his wife Algidosia (the Numidian Kings daughter) a goodly, though too too numerous a progenie; Because the law of birthright (in that kingdome inviolable) was not liberall to yonger brothers, of other patrimony, than sword and horse. The youngest among them, was a babe of exquisite beauty, to whom together with perfections of body, nature was bountifull of those vertues of the minde, which best beseemed a wellmeriting Prince. Polimero (for so was he called) in growth preuented his

yeares, and in knowledge and discretion the age of discretion; so as (though a child) he knew already what it imported him to be last borne, and what it was to bee a younger Brother […]. Metaneone the eldest brother (observing this their inclination, became iealous thereof) began so deeply to hate Polimero, that he (borne to endure rather any misfortune, than domestike hatred) purposed (though not as then arived to sixteene yeeres of age) to leave the countrey; and the effect had accompanied his deliberation, had not the reverence which he owed his father, perswaded him to a discreet dissimulation thereof.

Eromena, or, Love and revenge. Written originally in the Thoscan tongue, by Cavalier Gio. Francesco Biondi , Gentleman extraordinary of his Majesties Privie Chamber. Divided into six books. And now faithfully Englished, by Ia. Hayvvard, of Graies-Inne Gent (http:// name.umdl.umich.edu/A16157.0001.001)

Example 3
Domain: Religious Genre: Miscellaneous Style: Formal Spiritual
The 'General' genre of religious texts include the sub-genres of 'Christians', 'epistles' and 'sermons'. The author may describe a text in general terms which might suggest a more specific genre. In the case of Example 3 below, William Perkins describes his work in his address to the reader as a 'treatise', *How to liue, and that well in all estates and times, specially when helps and comforts faile*. A wide variety of treatises can be found in this period: religious, scientific, medical, academic, educational and so on. Gotti (2001) describes treatises as preferred for lengthy argumentation of new theories or debate of older ones compared to the shorter experimental essay. The content of Perkin's work is clearly religious: *faith, God, life euerlasting,* etc. The pronouns *vs* and *we* suggest the writer identifies with the reader, yet the final imperative 'see thou be a doer of them' indicates advice and hence a sense of authority. Together, we might reasonably assume this is a sermon, particularly if we know that Perkins was a theologian.

Secondly, when a man at his wits ende, knowes not what in the world to doe, being (as it were) plunged into a sea of miseries, faith giues direction and staies the minde. For when all temporall things faile vs euen to the very skinne and life, faith preserues within vs an affiance of the grace and mercie of God, and the hope of life euerlasting. Faith shewes vs hidden things not to be discerned by sense and reason. Life euerlasting is promised vs, but we die for al that: we heare of the resurrection, but in the meane season we rot in our graues: we are pronounced blessed, but yet we are ouerwhelmed with infinite miseries: bundance of all things is promised, but for all this we often huger and thirst: God promiseth to heare vs and to be present with vs, but he seemes oft times to be deafe (as it were) at our cries. Now then comes faith, which is the substance of things hoped for, and makes vs lift our minds aboue the whole world, to apprehend the inuisible and vnspeakable things of God which he hath reuealed and promised vnto vs. These things I shewe more at large in this small treatise following; read it at thy leisure, vse it for thy good, and see

thou be a doer of them.
How to liue, and that well in all estates and times, specially when helps and comforts faile. Perkins, William, 1601. (http://name.umdl.umich.edu/A09439.0001.001)

Example 4
Domain: Religious Genre: Catholicism Style: Formal Spiritual
Categorising unclassified texts can sometimes require a combination of text-internal and text-external approaches to genre categorization. The text *Marie Magdalens funeral teares* is a case in point. It is first important to know that the author, Robert Southwell was a Roman Catholic Jesuit priest, and that his work was published in 1591, four years before his execution for treason. It focuses on Mary Magdalene's distress at discovering Christ's empty tomb. Such a purpose is evident in the following extract:

EMONGST other mourneful accidents of the passion of Christ, that loue presenteth it selfe to my memory, with which the blessed Mary Magdelen louing our Lord more then her life, followed him in his iourney to his death, attending vppon him when his Disciples fledde, and being more willing to die with him, then they to liue without him. But not finding the fauour to accompany him in death, and loathing after him to remaine in life, the fire of her true affection enflamed her heart, and her enflamed hart resolued into vncessant teares, so that burning and bathing betwéen loue and griefe, shee led a life euer dying, and felt a death neuer ending. And when hee by whome shée liued was dead, and shée for whom he died enforcedly left aliue, shée praised the dead more then the liuing, and hauing lost that light of her life, shee desired to dwell in darkenesse, and in the shadow of death, choosing Christs Tombe for her best home, and his corse for her chiefe comfort. For Mary (as the Euangelist saith) Stoode*without at the Tombe weeping. (pages unnumbered)
Marie Magdalens funeral teares. Robert Southwell, 1591. (http://name.umdl.umich.edu/A12628.0001.001)

The opening of this dedication "To the worshipfull and vertuous Gentlewo-man, Mistres D. A." contains phrases such as "blessed Mary Magdelen" and "vncessant teares" which, while not explicitly identifying the text as Catholic, give a Catholic flavour to the piece. However, bibliographic searches for texts can sometimes produce a relevant academic article elucidating practices around the time of publication, as the following extract illustrates:

> In early modern Protestant England, traditional Catholic worship and sacraments, particularly the Mass, declined, and many Catholics feared for their salvation. At the same time, an increased veneration of Mary Magdalene focused no longer on penance and redemption [a

more Protestant interpretation], but on Mary's discovery of Christ's empty tomb. Magdalene's distress at losing the corporeal body of Christ mirrored English Catholic anxiety over losing the body of Christ as contained in the Eucharist in the absence of regular Mass. English Catholics chose to revive and adapt this form of Magdalene symbolism to best meet their spiritual needs. (McClain 2007:77)

The article goes on to mention Southwell's text as an example of such Catholic practice, thus affirming its genre categorisation.

Example 5
Domain: Administrative Genre: Royal Sub-genre: Communication and Order Style: Formal Statutory
The Royal genre includes the sub-genres of Communications and Orders, and Proceedings. *By the King. A proclamation against the disorderly transporting His Maiesties subiects to the plantations within the parts of America* orders officials to prevent people from travelling to America without a licence from His Majesty's Commissioners for Plantations. The diversity of linguistic features, content and visual appearance of this text can best be accounted for by a 'pragmatics on the page' methodological approach as proposed by Carroll, et.al. (2013). Using a combination of etic description of visual features in one utterance, comparison with others and emic analysis of the relation between visual form and function, (un)conventional usages and pragmatic effects can be identified. In this case, the title clearly indicates the royal genre, while the presentation and typeface also suggest formal conventions for royal proclamations. The royal coat of arms and mottos, "HONI SOIT QVI MAL Y PENSE" and "DIEV ET MON DROIT" precede the text. The first letter is ornately embellished, the first nine words and the place names *America*, *England*, *Wales* and *Barwick* are in an unelaborate typeface (e.g. that), while the main body of the proclamation is in the highly formal English Gothic font (e.g. 𝔱𝔥𝔞𝔱). It may be that the intention was to convey the essence via the simple typeface, while conveying royal authority via more elaborate typography, including capitalisation ("MAIESTIE"). Regal authority is also conveyed by honorific and legalistic language "His Maiestie hauing taken the premisses into consideration, is minded to restraine", "That they doe not hereafter permit or suffer any persons", etc.

THe Kings most Excellent MAIESTIE being informed that great numbers of His Subiects haue bin, and are euery yeare transported into those parts of America, which haue been granted by Patent to seuerall persons, and there settle themselues, some of them with their families and whole estates: amongst which numbers there are also many idle

and refractory humors, whose onely or principall end is to liue as much as they can without the reach of authority: His Maiestie hauing taken the premisses into consideration, is minded to restraine for the time to come such promiscuous and disorderly departing out of the Realme; And doth therefore straitly charge and command all and euery the Officers and Ministers of his severall Ports in England, Wales, and Barwick, That they doe not hereafter permit or suffer any persons, being Subsidie men, or of the value of Subsidie men, to embarque themselues in any of the said Ports, or the members thereof, for any of the said Plantations, without Licence from His Maiesties Commissioners for Plantations first had and obtained in that behalf.

By the King. A proclamation against the disorderly transporting His Maiesties subiects to the plantations within the parts of America. Charles I, King of England, 1637. (http://name.umdl.umich.edu/A73992.0001.001)

Example 6

Domain: Instructional Genre: Miscellaneous Style: Formal Instructional

Texts in the instructional domain can cover a wide variety of content, such as alchemy, almanacs, astrology and predictions, lectures and general manuals. Example 6 is an extract from *The courtier*, Thomas Hoby's (1561) translation of Baldassarre Castiglione's (1528) *Il cortegiano,* an early example of courtly courtesy or politeness literature for "*yonge gentilmen and gentilwomen abiding in court, palaice or place*". Culpeper (2017) points out that Hoby's 1561 translation of Castiglione's courtly guide was soon superseded by Robert Peterson's (1576) translation of Giovanni della Casa's highly influential Galateo (1558), aimed at a broader audience of city dwellers. In Hoby's translation of Castiglione's manual, the instruction largely, though not exclusively, takes the form of precepts beginning with infinitive forms, here expressing the necessary qualities required to be a waiting gentlewoman.

Of the chief conditions and qualityes in a vvaytyng gentylvvoman.

TO be well born and of a good house.

To flee affectation or curiositie.

To haue a good grace in all her doinges.

To be of good condicios & wel brought vp.

To be wittie and foreseing, not heady and of a renning witt.

Not to be haughtie, enuious, yltunged, lyght, contentious nor vntowardlye.

To win and keepe her in her Ladies fauour & all others.

To do the exercises meete for women, comlye and with a good grace.

To take hede that she giue none accasion to bee yll reported of.

To comit no vice, nor yet to be had in suspitio of any vice. *The courtyer of Count Baldessar Castilio diuided into foure bookes. Very necessary and profitable for yonge gentilmen and*

gentilwomen abiding in court, palaice or place, done into English by Thomas Hoby. 1561 (http://name.umdl.umich.edu/A18135.0001.001)

Example 7
Domain: Informational Genre: Miscellaneous Style: Informational
The Informational domain contains a diverse range of publications, ranging in content and extension from finance to food, and biographies to letters. As such, it captures a broad range of genres, some not named but coming under the category 'Miscellaneous'. The text from which Example 7 is taken, *The prayse of nothing*, has been termed a paradoxical encomium, a "rhetorical jest or display piece which involves the praise of unworthy, unexpected, or trifling objects" (Miller 1956). Indeed, a potential genre label might be 'jest book', a kind of rogue literature often featuring beatings and tricksters, written by the educated and propertied of the period (Woodbridge 2003). The jest here, however, appears more philosophical than physical. Alternatively, the word *praise* in the title might also suggest a label in line with Austin's theory of speech acts (Austin 1962), in which the author performs the illocutionary act of praising (see section 2.1. and Taavitsainen's application of metacommunicative expression analysis). He also describes his work as "a lenetiue [palliative] of the mind", and makes a case for the superiority of Nothing, it being the beginning and end of everything. Ultimately, it seems reasonable to classify this piece as a hybrid genre (Fowler 1982). Stylistically, the text appears to be both informational and instructional, though the authorial comment in the Example below, "tarying so long on nothing" suggests a more reflective rather than didactic purpose.

I feare I haue béene offensiue to many, for tarying so long on nothing: who therfore should not (if they duely regarde their particuler) remaine seuere censors of this matter: they hauing either feared, loued, or estéemed nothing, with greater partiallitie then I haue addicted my selfe to this labor.
The prayse of nothing. By E. D. Dyer, Edward, Sir, 1543–1607, attributed name, Daunce, Edward, attributed name. (http://name.umdl.umich.edu/A21010.0001.001)

Paradoxical plays on 'nothing' are also common in Shakespeare: "And nothing brings me all things" (*Timon of Athens* 5.1); "Nor I, nor any man that but man is, /With nothing shall be pleas'd till he be eas'd / With being nothing" (*Richard II* 5.5).[9]

Example 8
Domain: Informational Genre: Essay Style: Informational
The *Essais* of Montaigne (1580) and the *Essayes* of Bacon (1597) helped to establish the essay as a genre. The form attained increasing importance across a variety of fields in the Early Modern period (Taavitsainen 2017). The word *essay* derives from the French for a *trial* or *attempt* and, according to Montaigne, was the fruit of one's natural faculties of mind as opposed to acquired knowledge ("C'est icy purement l'essay de mes facultez naturelles, et nullement des acquise"[10]). As Taavitsainen suggests, the label may have served as a form of modesty, particularly in professional medical texts. Diverse types of essay characterise the informational domain. A sub-genre of essays is Commentaries on People and Places, of which *Asylum veneris, or A sanctuary for ladies* is a good example. It is a composition of moderate length explaining (or perhaps 'mansplaining') women's virtues and moral values, as the contents page in Example 8 shows.

CHAP. 2. Of their Beauty.
CHAP. 3. Of their Chastity.
CHAP. 4. Of their outward modestie.
CHAP. 5. Of their Humility, and supposed Pride.
CHAP. 6. Of their Silence and falsly obiected Talkatiuenesse.
CHAP. 7. Of the Constancie of their Affections.
CHAP. 8. Of their Learning and Knowledge.
CHAP. 9. Of their Wisedome and Discretion.
CHAP. 10. Of their Valour and Courage.
Asylum veneris, or A sanctuary for ladies . D. T. (Daniel Tuvill) 1616.
(http://name.umdl.umich.edu/A14083.0001.001).

4   Summary of classification criteria and limitations

As the discussion of the examples above shows, I have been able to draw on a wide range of both internal (beyond only lexical and grammatical) and external features of texts when assigning genres. Table 4 summarises these criteria:

Table 4: Criteria for manual genre classification

Text-internal criteria	Text-internal and external
Formal properties of text • layout (e.g. stanzas, continuous text, list) • typography (e.g. font choice, embellishment, capitalization) Textual content • lexis (e.g. lines, muse) • syntax (e.g. verb tense) • function (e.g. *to*-infinitive describing a necessary condition, implying advice) • register (e.g. honorific) • setting (e.g. ancient kingdom) Page pragmatic analysis (Carrol et.al. 2013) • etic and emic description	EEBO metadata • e.g. year and place of publication, listed authors, keyword terms • contributor's description of text • author • translator
	Text external criteria
	• bibliographic search on author/text • academic books, book chapters and articles • websites – Wikisource, Wikipedia
	Further considerations
	• fuzzy boundaries, hybrids, mixed modes, purpose, overall structure, etc.

Qualitative analysis clearly aids the classification of texts which have not easily been categorised by automatic means. However, certain limitations should be recognised. While such analysis may bring rewards in terms of accuracy by attempting to mitigate researcher bias, it can prove to be not only time-consuming, but also susceptible to error, particularly when categorising or checking the categorisation of a large number of texts (5,697 in this case). A further difficulty concerns the categorisation of texts which seem to cover a variety of genres. As one literary critic has observed: "Sidney's Arcadia is perhaps the supreme Elizabethan example of what I shall call the mixed mode, that strange conjunction of literary kinds which Polonius might have termed 'tragical-comical-historical-pastoral.'" (Greenblatt 1973: 269). Such 'mixed mode' texts invariably prove difficult to classify and often end up in a 'Miscellaneous' category.

In terms of meeting key desiderata of corpus design (see Kytö and Smitterberg 2015), the 1560–1640 corpus conforms to criteria of representativity insofar as it contains all existing and available texts in the EEPO-TCP dataset. That said, it can never be completely representative of all language of the period, as it evidently does not contain spoken language, handwritten texts and non-extant works. As regards comparability, given that genres can and do change with the

passing of time, the degree to which any texts can be said to belong to the same genre over an 80-year period will always involve a certain degree of doubt.

5   Benefits of the genre classification scheme

The *Encyclopaedia of Shakespeare's Language* project makes use of the genre classification scheme to show the distribution of linguistic forms found in Shakespeare across the different domains. This will give users a sense of whether certain forms were widely used across a variety of domains or restricted to specific contexts. It will also give a sense of the stylistic flavour of particular words of forms – be they formal, informational, literary, and so on.

5.1   For Shakespeare studies

More generally, the genre classification presented here will allow Shakespeare scholars to conduct a variety of corpus linguistic studies. Investigations might attempt to answer questions relating to:

- Shakespearean and other poetry of the same period;
- particular themes, such as the concept of wine in Shakespeare's plays and its contemporary associations:

  > Every time wine appears in Shakespeare's plays, I suggest, it activates 'England' and 'Englishness' as concepts – at once newly revivified yet still fluid – that rely on a body/state analogy. When Falstaff lauds the powers of his sherris-sack in *2 Henry IV*, audiences would recognize that his personal addiction is implicated in a national failing: it is a legacy of foreign wars and an ongoing danger for England's corpus economicum (Raber 2016: 75).

- a particular character and a commonly associated theme represented by a genre, e.g. Malvolio in *Twelfth Night* and Protestantism;
- notions of 'commodity' in Shakespeare's *King John* and the wider world.

5.3   For genre research

Historical studies of genre have often focused on carefully constructed corpora of genres or text type, such as medical recipes (Taavitsainen 2001b), letters (Bergs 2004) or law reports (Fanego et al. 2017). Modern corpus query tools such as CQPweb[11] facilitate searches based on different parameters and with a large corpus, allow researchers to build their own sub-corpora.

The genre classification scheme described in this paper is ideal for scholars engaged in researching specific genres or sub-genres as they now have access to ready-made text collections. Using CQPweb, sampling frames and queries based on them can be restricted to specific (sub-)genres, domains, styles, or combinations of some or all of these. Within the period 1560–1640, searches can be restricted to decades or specific years. An additional possibility is to create one's own sub-corpora based on particular research interests.

5.3   For historical linguistic research

This genre classification scheme should facilitate progress in historical linguistic research. Researchers such as Kytö and Smitterberg (2015) have shown that, particularly since the advent of corpus linguistics, genre has become an essential parameter in the field and predict a continued growth in its significance. Specifically, as Biber and Gray (2013) have shown, grammatical change can be mediated by specific genres[12] such as news reportage and academic writing, at least in twentieth-century texts. The scheme proposed here opens up numerous possibilities for comparing historical genres, domains and styles. Potential studies could investigate historical change by:

- exploring lexical innovation – research within the Science genre could reveal the extent of seventeenth-century attempts to create scientific forms of language (Slaughter 1982);
- studying semantic change in the Administrative domain, for example, by searching within the 'Royal' and 'Parliamentary' genres, the researcher could deploy corpus linguistic techniques to test a claim that post-French Revolution political and historical denominations such as 'Left', 'Right', 'radical' and 'conservative' misrepresent understanding of the terms in the early modern period (Condren 2016);
- comparing styles to investigate whether language became more 'secularized' over the period as has been claimed (De Grazia 1980).

6   Conclusion

In this paper, I have made a strong case for a principled genre classification scheme for a segment of EEBO-TCP texts covering the period 1560–1640. This scheme draws from and builds on a considerable body of genre theory and research over the last forty years. This scheme will not only serve as a valuable basis for comparison of Shakespeare's language with that of his contemporaries, but will also provide multiple opportunities for research in literary corpus stylis-

tics, genre studies and historical linguistics. An innovative feature of this classification scheme is that the texts included are not merely representative of the period, but comprehensively cover all published genres in the period, although excluding texts in the Shakespeare or Contemporary Playwrights corpora as previously mentioned. Plans for future extensions and refinements to the scheme include the inclusion of digitalized versions of handwritten texts, and expansion to include texts which represent more colloquial styles. Researchers, teachers and students with an interest in Shakespeare and/or the history of English should therefore find this work particularly profitable.

Acknowledgments

Notes

1. Probably Arthur Golding's English translation published in 1567. See Bate (1994) for a discussion on Shakespeare's relationship with Ovid.
2. https://quod.lib.umich.edu/e/eebogroup/
3. *The Encyclopaedia of Shakespeare's Language project* (2016–2019) is funded by the UK's Arts and Humanities Research Council (grant reference AH/N002415/1); see http://wp.lancs.ac.uk/shakespearelang/.
4. Demmen, Jane. (in preparation) *A comparative corpus for Shakespeare's plays: Compiling a corpus of Early Modern English play-texts by other contemporaneous playwrights for the Encyclopaedia of Shakespeare's Language project.*
5. See http://www.helsinki.fi/varieng/CoRD/corpora/CED/
6. It is also important to recognise that some texts may have been categorised differently than they would be today and that genre categories at the time were not stable (see Schmied and Claridge 1997).
7. I would like to express my sincere gratitude to Tony McEnery for his original work on genre classification of EEBO texts, which forms the basis of this work, for providing the information in this section, and for reviewing

and providing valuable feedback on the genre classification scheme presented here.

8. In *The manuall of the anatomy or dissection of the body of man containing the enumeration, and description of the parts of the same, which usually are shewed in the publike anatomicall exercises. Enlarged and more methodically digested into 6. books*. By Alexander Read, Doctor of Physick, a fellow of the Physitians College of London, and a brother of the Worshipfull Company of the Barber-Chirurgeons. Available at http://name.umdl.umich.edu/A10510.0001.001

9. See Jorgensen (1954) for a discussion of Shakespeare's use of 'nothing' and the word's use in contemporary genres.

10. See De Montaigne, Michel. (1802).

11. Available at https://cqpweb.lancs.ac.uk/ (Hardie, 2012).

12. Biber and Gray (2013) use the term 'registers'.

Corpus tool

CQPweb. Created by Andrew Hardie (Lancaster University). https://cqpweb.lancs.ac.uk/

References

Austin, John. 1962. *How to do things with words*. Oxford: Oxford University Press.

Bate, Jonathan. 1994. *Shakespeare and Ovid*. Oxford: Oxford University Press.

Bergs, Alexander. 2004. Letters: A new approach to text typology. *Journal of Historical Pragmatics* 5(2): 207–227.

Biber, Douglas and Bethany Gray. 2013. Being specific about historical change: The influence of sub-register. *Journal of English Linguistics* 41(2): 104–134.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Carroll, Ruth, Matti Peikola, Hanna Salmi, Mari-Liisa Varila, Janne Skaffari and Risto Hiltunen. 2013. Pragmatics on the page. *European Journal of English Studies* 17(1): 54–71.

Claridge, Claudia and Andrew Wilson. 2002. Style evolution in the English sermon. In T. Fanego, B. Méndez-Naya and E. Seoane (eds.). *Sounds, words, texts and change: Selected papers from 11 ICEHL*, 25–44. Amsterdam: John Benjamins.

Condren, Conal. 2016. *The language of politics in seventeenth-century England*. Basingstoke: Palgrave Macmillan.

CoRD | Helsinki Corpus (HC). 1991. *Helsinki.fi*. http://www.helsinki.fi/varieng/ CoRD/corpora/HelsinkiCorpus/textcategories.html,

Crystal, David. 2018. *Shakespeare's words*. https://www.shakespeare-swords.com,

Crystal, David. 2012. *Think on my words: Exploring Shakespeare's language*. Cambridge: Cambridge University Press.

Culpeper, Jonathan and Merja Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.

Culpeper, Jonathan. 2017. The influence of Italian manners on politeness in England, 1550–1620. *Journal of Historical Pragmatics* 18(2): 195–213.

De Grazia, Margreta. 1980. The secularization of language in the seventeenth century. *Journal of the History of Ideas* 41(2): 319–329.

De Montaigne, Michel. 1802. *Essais*. (Vol. 2: 10). Didot. (Original work published 1580).

Fanego Teresa, Paula Rodríguez-Puente, María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo, Cristina Blanco-García and Iván Tamaredo. 2017. *The Corpus of Historical English Law Reports 1535–1999 (CHELAR)*: A resource for analysing the development of English legal discourse. *ICAME Journal* 41(1): 53–82.

Fowler, Alastair. 1982. *Kinds of literature: An introduction to the theory of genres and modes*. Oxford: Clarendon Press.

Gotti, Maurizio. 2001. The experimental essay in Early Modern English. *European Journal of English Studies* 5(2): 221–239.

Greenblatt, Stephen J. 1973. Sidney's 'Arcadia' and the mixed mode. *Studies in Philology* 70(3): 269–278.

Hardie, Andrew. 2012. CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17 (3): 380–409.

Hope, Jonathan. 2014. *Shakespeare and language: Reason, eloquence and artifice in the Renaissance*. London: Bloomsbury.

Hünig, Wolfgang K. 2003. Style labels in monolingual learner's dictionaries. In G. Radden, H. Cuyckens, R. Dirven and K. Panther (eds.). *Motivation in language: Studies in honor of Günter Radden*, 367–390. Amsterdam: John Benjamins.

Johnson, Keith. 2014. *Shakespeare's English: A practical linguistic guide*. Harlow: Pearson.

Johnson, Keith. 2019. *Shakespeare's language: Perspectives past and present.* London: Routledge.

Jorgensen, Paul A. 1954. Much ado about nothing. *Shakespeare Quarterly* 5(3): 287–295.

Kytö, Merja. 1996 [1991]. 3rd edition. *Manual to the diachronic part of the Helsinki Corpus of English Texts: Coding conventions and lists of source texts.* Helsinki: Department of English, University of Helsinki.

Kytö, Merja and Erik Smitterberg. 2015. English genres in diachronic corpus linguistics. In P. Shaw, B. Erman, G. Melchers and P. Sundkvist. (eds.). *From clerks to corpora: Essays on the English language yesterday and today*, 117–133. Stockholm: Stockholm University Press.

Lee, David Y.W. 2001. Genres, registers, text types, domain and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning & Technology: A Refereed Journal for Second and Foreign Language Educators* 5(3): 37–72.

McClain, Lisa. 2007. "They have taken away my Lord": Mary Magdalene, Christ's missing body, and the mass in Reformation England. *The Sixteenth Century Journal* 38(1): 77–96.

Meurman-Solin, Anneli. 2001. Genre as a variable in sociohistorical linguistics. *European Journal of English Studies* 5(2): 241–256.

Miller, Henry K. 1956. The paradoxical encomium with special reference to its vogue in England, 1600–1800. *Modern Philology* 53(3): 145–178.

Orr, Leah. 2011. Genre labels on the title pages of English fiction, 1660–1800. *Philological Quarterly* 90(1): 65–95.

Raber, Karen. 2016. Fluid mechanics: Shakespeare's subversive liquors. In D.B. Goldstein, and A.L. Tigner (eds.). *Culinary Shakespeare: Staging food and drink in early modern England*, 75–96. Pittsburgh: Duquesne University Press.

Ravassat, Mireille and Jonathan Culpeper (eds.). 2011. *Stylistics and Shakespeare: Transdisciplinary approaches*. London: Continuum,

Schmied, Josef and Claudia Claridge. 1997. Classifying text- or genre-variation in the Lampeter Corpus of Early Modern English Texts'. In R. Hickey, M. Kytö, I. Lancashire and M. Rissanen (ed.). *Tracing the trail of time: Proceedings from the Second Diachronic Corpora Workshop*, 119–136. Amsterdam: Rodopi.

Shapiro, James. 2011. *1599: A year in the life of William Shakespeare*. London: Faber & Faber.

Slaughter, Mary M. 1982. *Universal languages and scientific taxonomy in the seventeenth century.* Cambridge: Cambridge University Press.

Smith, Winnie. 2017. Documenting categories in EEBO-TCP data [Blog post: May]. https://www.linguisticdna.org.

Steen, Gerard. 1999. Genres of discourse and the definition of literature. *Discourse Processes* 28(2): 109–120.

Sullivan, Ceri. 2007. Disposable elements? Indications of genre in early modern titles. *The Modern Language Review* 102(3): 641–653.

Swales, John. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taavitsainen, Irma. 2001a. Changing conventions of writing: The dynamics of genres, text types, and text traditions. *European Journal of English Studies* 5(2): 139–150.

Taavitsainen, Irma. 2001b. Middle English recipes: Genre characteristics, text type features and underlying traditions of writing. *Journal of Historical Pragmatics* 2(1): 85–113.

Taavitsainen, Irma. 2017. The essay in Early Modern and Late Modern English medical writing. *Recherches anglaises et nord américaines (RANAM)* 50: 15–30.

Taylor, Gary, John Jowett, Terri Bourus and Gabriel Egan. 2017. *The new Oxford Shakespeare: Critical reference edition*. Oxford: Oxford University Press.

Toolan, Michael. 2008. Narrative progression in the short story: First steps in a corpus stylistic approach. *Narrative* 16(2): 105–120.

Woodbridge, Linda. 2003. Jest books, the literature of roguery, and the vagrant poor in Renaissance England. *English Literary Renaissance* 33: 201–210.