Between Authorship and Oral Transmission: Negotiating the Attribution of Authorial, Oral and Collective Style Markers in Early Modern Playtexts

Lene Buhl Petersen

VISL, University of Southern Denmark (<lenebuhlpetersen@gmail.com>)

Abstract

The production of playtexts in early modern England falls between two categories of artistic provenance: textual production in quill and print and oral transmission of the text committed to paper. Both categories are rightly speaking processes, and may be repeated several times over within the lifespan of a play. The former is the domain of authors, scribes and printers, the latter the responsibility of actors using their memories to verbally transmit the play in performance. An early modern playtext may thus be (co)written, probably performed and potentially printed, and possibly rewritten, reperformed and reprinted in almost any given combination. It is only to be expected that a number of stylistic 'complications' will ensue. The question remains how to determine which stylistic markers characterise which creative domain. This paper returns to the cross-roads between authorship attribution and the quantification of other (oral, collective) style markers in an attempt to offer discussion and a better overview of appropriate methodologies for determining which features may feasibly be attributed to which source(s).

Keywords: Authorship Attribution, Computational Linguistics, Early Modern Text Studies

1. *Introduction*

The play's the thing, wherein...

The following essay draws on the work of Walter Ong, Milman Parry and Albert Lord, John Miles Foley, and Thomas Pettitt on oral-formulaic theory. It also draws on advances in attribution studies based on computational stylistic analysis of linguistic features. This field is thankfully still honing its methods through the dedicated efforts of scholars like e.g. Brian Vickers, John Burrows, David Hoover and Hugh Craig, many of whom are currently

embracing the perspectives opened up by cognitive and neurolinguistic research. Within Shakespeare studies, Ian Lancashire remains an important advocate for understanding the mechanisms of language production behind the early modern playtexts and the application of cognitive linguistic theory to corpus-based attribution studies. His articles on the role of the working, or verbal short-term, memory, its relation to long-term lexical storage and the production and retention of words not as single items but as *chunks*[1] deserve credit for highlighting the significance of neural mechanisms also in determining authors' idiolects.[2] Hoover (2003), Craig and Kinney (2009), Burrows (2012), Vickers (2011a, 2011b, 2012), and Craig (2014) have all, in turn, published studies that apply this particular approach to Shakespearean attribution studies albeit from different vantage points.

The chunks Lancashire started quantifying back in the late 1990s[3] are variously also known as prefabricated units, strings, multi-word units, N-grams, fixed phrases, verbal sequences, phrasal repetends, formulas, and/or collocations. In a longer version of this contribution, I have taken the opportunity to argue at some length for a stricter definition of the phenomenon of the multi-word unit, focussing in particular on collocation.[4] The longer essay argues that there are limits to the *unmediated* application of collocation theory/collocation extraction in Shakespearean attribution studies, but also identifies promising uses of collocation extraction if we narrow the

[1] For an explanation of the function of the verbal short-term memory (the so-called phonological loop) and its connection to long-term memory and the learning of verbal sequences, see Burgess and Hitch 2006, 627-652.

[2] See e.g. Lancashire 1996, 1997, 171-185; 1999. In his 2004 contribution to *A Companion to Digital Humanities*, Lancashire elaborates on the phenomenon of 'convergence zones', the ostensibly unique 2-second long combinatorial *composita* of single speakers/authors: 'These combinations appear in repeated phrases or unfixed collocations that are not more than 5-9 units in length. This follows if, as scientists suspect, working memory constraints are associated with a deeper limitation existing at the level of neural networks. Indeed, one use of computer text analysis is to help determine the size and complexity of long-term-memory networks, how many things can converge on a convergence zone' (Lancashire 2004). I expect that this phenomenon will be more fully discussed and its applicability to attribution studies further investigated in future publications.

[3] In 10 early Shakespeare plays, totaling 197,000 words, Lancashire counted '12,600 different word forms which combine to form about 32,300 *fixed phrases*' (emphasis mine). Lancashire subsequently deducted that '[Shakespeare's] phrasal lexicon must exhibit traces of [his] networked associational memory and thus of his idiolect' (1999, 744). Also quoted in Vickers 2012, 27.

[4] The full paper is available upon request. The sections exploring the nature and general applicability of multi-word units in early modern attribution studies form part of a forthcoming research project with Marcus Dahl and Darren Freebury-Jones.

definition and acknowledge the special linguistic properties of the extant sample material. The following section summarises the longer discussion.

According to Ellis (1997, 128), chunks subdivide into collocations. As such, collocations are expressions of the way the brain habitually computes language through the rehearsal and repetition of words in the short-term memory in association with long-term-memory lexical storage.[5] J.R. Firth and John Sinclair (who elaborated on Firth's initial theory) both apply a contextual/lexical (though not semantic) perspective, and notably define collocation as *words that co-occur habitually, by mutual expectancy*, and *more frequently* than would be expected by chance. Hence, the Firth-Sinclair approach is sometimes also described as frequency-based (Firth 1957; Halliday and Robins 1966; Sinclair 1966, 1996; Hoey 1991). But another, so-called phraseological, approach to the collocation phenomenon also exists, represented by e.g. Mel'čuk (1998), Cowie (1978, 1981) and Gitsaki (1999). This to me more compelling definition combines the semantic and structural properties of language and states that the *meaning* and *function* of a collocation matter at least as much as frequent co-occurrence. Unfortunately, we have yet to arrive at a single precise or non-controversial definition of the concept of collocation. Nor has it been ascertained which of the above approaches best serves the purposes of attribution studies. Only very few scholars (and only one attribution scholar that I know of) seem willing to discuss the problems involved in making a still so loosely defined concept the basis of corpus-driven stylistic research (Seretan and Wehrli 2006, 2011; Burrows 2012, 380). Surely, if the aim is to (re)allocate authorship and perhaps even reorganise canons, we can expect the linguistic markers used to do so to be more scrupulously described?

Extraction of collocations in early modern sample texts, as exemplified below, is clearly no longer much of a problem, while describing the processes that underpin the multi-word units clearly is. One place to start would be to consider some questions pertinent to the multi-word unit as a style marker with specific reference to Shakespeare studies: firstly, what is the relation of collocations to phraseology, to individual phraseognomy[6] – and ultimately to idiolect? How is the concept best applied in the present *local* context of Shakespearean authorship studies where the sample material consists of

[5] See further Pawley and Syder: 'In the store of familiar collocations there are expressions for a wide range of familiar concepts and speech acts, and the speaker is able to retrieve these as wholes or as automatic chains from the long-term memory; by doing this he minimises the amount of clause-internal encoding work to be done and frees himself to attend to other tasks in talk-exchange, including the planning of larger units of discourse' (1983, 192).

[6] The term phraseognomy, meaning individual phraseology, was coined by John Sinclair (1991, 92).

playtexts written under some time pressure by professional playwrights for live, oral rendition? What *do* and *don't* we know about the connection between collocations and the verbal short-term memory, and between collocation and the syntactically binding combinatorial neuronal assemblies recently explored by neurolinguists like Friedemann Pulvermüller (Pulvermüller and Shtyrov 2003, Pulvermüller and Knoblauch 2009, Pulvermüller 2010)? How do collocations relate to lexical priming within a discourse community ('the early modern stage' being one such community) and to the functional 'verbal formulas' identified by oral composition scholars Lord and Parry and their followers; stock phrases which probably involve the long-term memory? And finally, is a multi-word unit significant simply because it is frequent enough to be statistically significant within a given text sample (words co-occur more frequently than by chance) or is it significant because of what it *is* or *does* in that text (e.g. a lexical vs. a grammatical collocation)?[7] If in future attribution studies we aim to look for the author through the identification of idiosyncratic multi-word units in surviving playtexts, I believe we need to know much more about what those units are and which functions they perform in those texts. This essay offers to at least begin this discussion.

2. *Local Definitions for Local Texts: Using Multi-word Units in Shakespearean Attribution Studies*

In attribution studies we have to assume that at least some linguistic markers are individual and unique to one author - be that Shakespeare, Kyd, or any other candidate under investigation. Based on the review of theories provided in the longer unpublished essay quoted above (cf. note 4), I believe we can – and should – include certain multi-word units/collocations among these markers. But one thing is extracting and quantifying those units, another is to determine which units are distinct and likely to be of individual origin, and which are distinctive of a collective speech community rather than an individual. In Hoey's view, collocations can denote anything from incompetence in the writer to a writer striving after unusual effect (2009, 45), while Burrows (2012, 381) argues that some identifiably multi-word units may be little else than quasi-phrases or gibberish. Brian Vickers' acceptance of N-grams that are not *topical*, i.e. not linked to contextual semantics (Vickers

---

[7] A reviewer of this essay kindly supplied this example: in *A Funerall Elegie* (1612) we find 'Reason's golden meane' where the searchable texts in EEBO with {golden meane near.5 reason(s)} produce only 3/17348 returns before 1652. This is definitely a collocation yet it is not frequent. In *Love's Labour's Lost*, 'O, these barren taskes, too hard to keepe, / Not to see Ladies, study, fast, not sleepe' reveals after a search of EEBO with {tasks near.20 ladies} only 3/34874 returns before 1690. This is also a collocation but again it is not frequent. Both instances are notably examples of rare lexical collocation.

2011b) follows early Firthian anti-structural notions of collocation (Firth 1957, 196), but at the same time is not easily reconciled with recent cognitive theories of syntactical binding (Pulvermüller and Shtyrov 2003, Pulvermüller and Knoblauch 2009, Pulvermüller 2010).[8] Somewhere in between these poles lies the rather low-practical possibility that some collocations perform useful dramaturgical functions in the playtexts, enforced by collective priming, performance tradition and oral-memorial transmission (Pettitt 1988, 2001; Petersen 2010). To impose some sort of qualitative order on the quantification of multi-word units, I propose to follow Seretan (2009, 94; 2011; Seretan and Wehrli 2006), who suggests that structural or *syntactic* information is crucial for accurately detecting true collocations in corpora. This stance favours meaning and syntactical soundness of the units as well as frequency. Burrows (2012, 381) may not have been quite so specific when he advises against 'lumping together' several kinds of multi-word unit, but a remedy against lumping the evidence would certainly be achieved by categorising the kind(s) of multi-word unit or collocation we identify. Simultaneously, more knowledge about the stylistic features that *do* in fact indicate authorship would be made available. One way to enable such categorisation is to consistently apply syntactically-annotated corpora when extracting multi-word units. Another is the systematic description of the units identified in terms of their contextual *and* phraseological qualities.

A number of very good purpose-made resources for extracting collocations/multi-word units already exist.[9] I would highlight Martin Mueller's 'WordHoard' (a deeply tagged information-rich corpus and search

[8] For Vickers an N-gram is a string of N consecutive words whether grammatically meaningful or not, and so in this definition an N-gram is not necessarily committed to memory (retained) and verbally uttered (produced). I am inclined to discourage using this definition of a style marker *if* our aim is to attach modern neurolinguistic theory to practical text-mining techniques. That said, I am not ruling out that non-topical (longer) matching strings may work for yet other reasons; phonoaesthetics and phonology being prime areas for further investigation. See Firth (1968, 18), and Vickers (2011b): 'Unique matches of three [or more, sic.] consecutive words in *The Troublesome Reign* with comparable strings in other plays by Peele' in *The Troublesome Reign of John, King of England*, ed. by C.R. Forker (2011), Appendix 2, 335-336.

[9] The list is growing, and presently includes a range of KWIC (Key Word in Context) applications (also used by EEBO), the TACT program developed by Lancashire; Mueller's MONK tools; WordSmithTools; Seretan and Wherli's FipsCoWeb (2011); Mike Scot's WordCopyFind; Pl@giarism, etc. Using these and similar resources, identification and extraction of collocations can certainly be automated and replicated, and subsequently optimised if subjected to further conceptual checks and 'manual' analysis.

application);[10] 'KEMPE' (a POS-tagged corpus and search resource),[11] and Seretan and Wehrli's 'FipsCoWeb' (a retrieval resource for own uploaded files).[12] Other programs for textual analysis include Pl@giarism, WordCopyFind, WordSmith Tools and Concordance (all downloadable online, although not all free). Lancashire's TACT and Mueller's MONK tools for use in restricted corpus environments are equally effective, I am sure. Beyond these resources, a growing number of text editing applications for handling very large files, such as 'InfoRapid Search & Replace' or 'Dreamweaver', have proved extremely useful in both pre and post-editing of corpus studies.

To test some of the above theoretical considerations, along with the currently available software, and to add to already relevant research, I have chosen to do something very practical and review the collocations identified by Brian Vickers (2012) as evidence for Shakespeare's hand in the *additions* to *The Spanish Tragedy*.

---

[10] WordHoard (Mueller 2004-2013) is downloadable from <http://goo.gl/xYWSBy>. This stand-alone application allows searches in various corpora, incl. Homer, Spenser and Shakespeare. 'WordHoard Shakespeare' is a joint project of the Perseus Project at Tufts University, Northwestern University Library, and Northwestern University Academic Technologies. Texts are sourced from *The Globe Shakespeare*; the one-volume Cambridge Shakespeare, edited by W. G. Clark, J. Glover, and W.A. Wright (1891-1893), while *Internet Shakespeare Editions* of the quartos and folios, ed. by Michael Best, have been consulted 'to create a modern text that observes as closely as possible the morphological and prosodic practices of the earliest editions'. The corpus uses standardised spelling, and all text is fully lemmatised and morphosyntactically tagged. Mueller and others have further developed a set of text-mining tools, MONK or 'Metadata Offer New Knowledge', which may be used with the corpora in the WordHoard. Unfortunately, neither the MONK tools nor the enhanced *SHC Corpus* (50 million words, containing about 500 texts from 1533 to 1625 from the Text Creation Partnership [TCP], including plays and major Shakespearean sources) is freely available to scholars. Mueller's own corpus-based attribution studies have yielded promising results for multi-word units: "plays by the same author are likely to share more dislegomena [i.e. 2-word collocations]. ... If we look more closely at shared dislegomena by same-author play pairs, we discover that on average plays by the same author share five dislegomena, and the median is four. Roughly speaking, plays by the same author are likely to share twice as many dislegomena as plays by different authors. Clearly some author effect is at work, and there is some virtue in adding some precision to this intuitively plausible conclusion'. See Mueller 2014a.

[11] KEMPE: *Korpus of Early Modern Playtexts in English* contains approx. 9 million words of syntactically annotated early modern playtexts and masques. It is fully POS-tagged, and allows both plain-text searches and syntactical/POS/wildcard-based searches through corpus query tools. The resource readily facilitates both collocation and colligation extraction along with concordance data for all search types. The syntactically annotated version of KEMPE was prepared for online publication by Petersen (2004).

[12] FipsCoWeb was developed by Violeta Seretan *et al*. as an online application for extraction of collocations in own uploaded texts (Seretan and Wehrli 2011). It is freely accessible through: <http://goo.gl/f7b6XM>, and its rationale is explained in Seretan and Wehrli 2010.

In his study, Vickers, in collaboration with Marcus Dahl, uses a combination of Pl@giarism and InfoRapid Search & Replace to locate 116 instances of multi-word units in the *additions.* Those units are then matched by units in a number of other early modern canons, predominantly in Shakespearean plays, and as such presented as evidence for Shakespearean authorship. For the purposes of the study, Vickers and Dahl looked at plays produced between 1586-1642 (Vickers 2012, 35), using the database of plays underlying the KEMPE corpus (compiled by Petersen and Dahl 2001-2003 as part of joint PhD research).[13] I observe the same chronological constraints, using, however, the POS-tagged information-rich version of KEMPE (Petersen 2004). While particular emphasis is granted to plays preceding the publication of the *additions* in 1602, the time span I apply matches the remainder of corpus searches in this essay, where an expansive scope (1561-1652)[14] is favoured in order to gauge the units' general prevalence in early modern drama over time and across canons. The dating of playtexts is checked against *Annals of English Drama* in Wagonheim's revised edition (1989; Vickers and Dahl use Wiggins 2011).

Using the KEMPE corpus and search tools, which allow both normal searches and POS queries, I have checked the 116 multi-word matches compiled by Vickers,[15] and will venture my commentary below. It will soon become evident that I comment only on those instances that I find unlikely to be *authorial* collocations. Hence the commentary starts with item 2 from Vickers' list.[16] As a guiding principle, I apply a phraseological definition of collocation that stipulates a *syntactically sound expression* that is statistically significant, regardless of its morphological qualities. I.e., there may be distance between the headword and its collocates, but any units that cut across sentence boundaries are not accepted as collocations, *pace* Seretan (2013, 94-99) and Seretan *et al.* (2004, 1871): 'The involved words occur together more often than by chance, thus the collocation is restricted with respect to the collocate substitutability. Both grammatical and lexical collocations are considered (Benson *et al.* 1986), e.g., *abstain from*, *pay attention*, without limitation on the distance between

---

[13] See Petersen 2004.

[14] The earliest playtext in the KEMPE corpus is *Gorboduc* (1561/1562) and the latest is Brome's *A Jovial Crew, or the Merry Beggars*, performed 1641, printed 1652. Dating of plays follow *Annals of English Drama*, ed. by Wagonheim 1990.

[15] John Burrows has indicated that Vickers' phraseological evidence for Shakespeare's *Addition*s to *The Spanish Tragedy* is 'merely a variant of the 'parallel passages' studied by scholars down the centuries' (2012, 357). There is nothing wrong as such with parallel phrases. The *typology* of the phrases simply needs to be examined in greater detail, and it needs to be considered which linguistic agent (author, actor or tradition) is likely to be 'responsible' for the item in question.

[16] All items from Vickers' list (2012) are given in bold type. Where applicable, a hyperlinked KEMPE seearch augments the commentary.

words in text (*apart from the sentence boundaries*)' (Seretan *et al.* 2004, 1871; emphasis mine). In a number of instances I suggest that the multi-word unit in question is likely to derive from oral transmission / dramaturgical tradition rather than individual composition. Section 3 below elaborates upon this aspect.

2) **'…of it. | Besides'** It is questionable whether this 'co-occurrence' is much more than that. It may qualify as an unfixed collocation by Sinclair's definition (1991, 121), but it is neither a grammatical nor a lexical collocation (Benson *et al.*, 1986: ix ff.). This, in turn, makes it difficult to determine which is the headword and which are the collocates,[17] and complicates its relation to language production theories involving both the short-term memory loop and long-term lexical storage. The unit consists of two function words (a preposition and a pronoun) and one conjunction (some would say three function words) belonging to two structurally separate syntagmata. Finally, it traverses a sentence border, which further disqualifies a likely neuronal connection between the constituents (Pulvermüller 2010). It is not a phrase, either. *If* we accept 'besides' as a node, it collocates upward (Sinclair 1991, 121) with the two very frequent function words 'of' and 'it'. The unit is indeed rare, as a search in KEMPE quickly confirms. This *may* of course be an authorship marker. However, it may just as easily be an entirely random combination. In the KEMPE corpus, 'besides' (besides being preceded by commas, full stops and colons) produces the following left-context vectors 'a' (26 times), 'to' (23), 'of' (22), 'and' (19), 'in' (23), 'all' and 'have' (10), …'many', 'me', and 'her' before it collocates with 'it' (3 times in total).

4) **'Run(ne) to'** This grammatical collocation may possibly qualify as a situational dramaturgical formula (see Pettitt 1988, 184-185, and section 3 below). If so, it is probably not a strong indicator of individual, authorial style.

5) **'presently | and bid'** See 2). This example traverses two clauses if not two sentences. A search in KEMPE confirms that this is a rare co-occurrence, yet the matching phrasal repetends listed by Vickers indicate formulaicity, and follow an apparently common Verb + to + Noun + presently structure present in many other playtexts. The example defies categorisation as lexical or grammatical collocation.

---

[17] According to Burgess and Hitch, 'item nodes for familiar stimuli such as letters, digits or words are assumed to have strong pre-experimental connections to the nodes representing their constituent phonemes. When each item in a sequence is presented, an item node is selected by competitive queuing and language storing is accomplished by strengthening *connections* between *simultaneously active nodes* in *adjacent layers*' (2006, 629; emphasis mine).

7) **'Do(e) you hear me'** This clause is probably a dramaturgical formula (see below).

 "Do you hear me sir", by the way, is also found in *Two Angry Women of Abingdon* from 1598 (not in Vickers' list).

20) '**A thing of nothing**' This collocation is also an idiom or commonplace, and is also found in e.g. *A Knack to Know a Knave, Two Angry Women of Abingdon, The Pedlar's Prophecy*, and *The Cobbler's Prophecy*; all 1590s plays but not included in Vickers' list. While this is certainly a lexical collocation, such phenomena should probably be disqualified as authorship markers.

28) **'a sonne. │ For…'** (see 2). This is another syntactically disrupted, unfixed collocation in the *additions*. In Vickers' list, however, it is compared to two fully grammatical (and syntactically sound) collocations from *3Henry VI* and *King Lear*. If we accept that the short-term memory phonological loop deals in unfixed collocations, we need to investigate further the likelihood of such unfixed collocations working across sentence boundaries, which Seretan (2011) dismisses. That said, the (lexical) matches produced are all from Shakespeare's accepted canon.[18]

43) '**See where x comes/stands etc**.'[19] is the best example of an 'oral formula'/ evidence of formulaic dramaturgy in this list.[20] A grammatical clause such as this serves the communicative purposes of authors and actors alike, and therefore should be excepted from any study purporting to quantify evidence of only individual authorship. See further section 3 below.

51) **'No, no, you** …'** A search in KEMPE reveals that the phrase (if not collocation) is a common one across early modern dramatic canons, with pre-1600 examples of **no, no, you (verb)**[21] in e.g. *The Famous Victories, Orlando Furioso, Englishmen for My Money*, including an abundance of other occurrences. The phrase is governed by 'incremental repetition', an oral composition principle perpetuated by repeated performance and subsequently included in printed versions, as defined by Andersen *et al.* (1982).

---

[18] A further KEMPE check for [word="a" %c] [lex="child" | lex="daughter" | lex="son"] [word="for" %c] [pos="PRON" %c] yields further hits in (Rowley's) *When you See me* (1604), Field's *A Woman is a Weathercock* (first performed 1609-1610/publ. 1612) and Brome's *Weeding of the Covent Garden* (performed 1633/publ. 1641), all post-1600.

[19] This unit is strictly singular. More variants exist, including Look where x +VERB.

[20] See e.g. Pettitt's (1988) illustration of 'formulaic dramaturgy' in Marlowe's A and B texts of *Doctor Faustus.*

[21] Search string in KEMPE: [word="no" %c] [word=".?" %c] [word="no" %c] [word=".*" %c] [word="you" %c] [pos="V"%c].

52) '**You are deceiu/v'd'** This is a phrase or a clause, not a collocation. It yields a number of matches in pre-1600 plays besides those listed by Vickers (e.g. in *The Warres of Cyrus, Fair Em, A Knack to Know a Knave, Clyamon and Clamydes, Englishmen for My Money, Soliman and Perseda*, etc.).

59) '**I am not mad'** This is a clause. The repeated instances in *King John* may likewise be examples of internal incremental repetition (as explained above). As such, the phrase probably should be disqualified as a reliable authorial style marker.

61) '**I know thee (to be)'/ 'I know thee'** in *Titus Andronicus* may be another example of internal incremental repetition, and so potentially an oral style marker. Moreover, a KEMPE search yields plentiful examples of this clause structure elsewhere, which suggests that it could be a dramaturgical formula of collective origins. '**I know** [pronoun] **to be'** is found in at least 24 other pre- and post-1600 playtexts.

97) '**Well(,) sir(,) then'** The unfixed collocation 'Well sir' may likely be of oral/formulaic origin, and is very common indeed across the canons. The added **then** does, however, make this unfixed collocation less frequent, and as such the phrase could have authorship determination value. The item regrettably defies easy categorisation as lexical or grammatical.

105) '**Nay(,) then I'** Neither a lexical nor a grammatical collocation, but arguably an unfixed one, this structure is found in a great number of plays, as the hyperlink will testify, including many pre-1600. I would argue that it should be disqualified as an authorial marker.

110) '**Do, do, (do)…'** Some simple verbal incremental repetition is probably at work here. Both *2 Henry IV* and *Troilus and Cressida*'s textual and performative provenance is complex enough for us to expect a presence of oral-formulaic style markers such as those explored by e.g. Maguire (1996) or Petersen (2010). That this should be an authorial style marker is not very likely.

111) '**Not so () │ You…'** (see 2 above).

If the goal of attribution studies is to analyse the language of written compositions and assign authorship, and to do so on the basis of a sound theory of language, I agree: we need a strong theoretical rationale for our practical work (Vickers 2011a, 116). If part of our rationale rests on quantifying multi-word units such as those compiled by Vickers, Craig, Burrows, and Hoover, then we need to examine what they contain. As we have already established, collocations can be lexical or grammatical. A majority of the items I have

selected for review above are lexical, while a few are grammatical. Others defy categorisation. With the involvement of lexical neural assemblies and the long-term memory, at least some of the above units will probably have served aptly as mnemonic aids (i.e. situational formulas, or verbal expressions of formulaic dramaturgy, aiding the immediate language production that takes place on stage in the short-term verbal memories of actors. The fact that a collocation is lexical (containing a lexical headword plus collocates) does not then *in itself* vouch for individual, authorial origin. Other units could be construed as simple incremental patterning – a progressive oral composition mechanism, which may involve long-term memory storage of earlier 'lexical increments' too, and which may potentially be introduced through the course of transmitting a play on stage and in print several times (Andersen *et al.* 1982). Yet other items above could be random co-occurrences. Most importantly, most of the collocations, clause structures, and repetitious phrases listed above are *unlikely* to indicate *individual* authorship. Be that by Shakespeare or any other individual candidate. The remainder of Vickers' 116 multi-word units, now filtered for what I call oral or collective formulas, may very likely prove highly effective determinants of individual authorial style. The following section will discuss further the range and contents of those oral / collective formulas.

3. *The Oral Roots of Staged Verbalisation*

While the culture that Shakespeare and his contemporaries inhabited was not oral, the sub-culture or theatrical industry in which they worked was not fully textual either. One might say that they worked under residually oral conditions. At the very least, they belonged to a discourse community heavily reliant on memorial reconstruction. The texts that survive from the early modern stage certainly do so in various formats, versions and qualities quite unlike what we today would call stable text formats. Most of them were written for oral performance. Milman Parry once said that when someone transcribes an oral song into a text, it remains oral. Have we indeed sufficiently understood the provenance of the texts we work with? Print makes for more tightly closed verbal art forms (Ong 1982, 130), but the early modern stage dealt in several 'open' handwritten formats relevant at different stages in a play's lifespan; different printed versions of playtexts were circulating, and often several hands were involved in playwrighting at composition or revision stage. Acting troupes went touring with alternate versions of plays, etc. In other words, we are back where we started with the written/rewritten, performed/reperformed and printed/reprinted sample material.[22] It seems to me appropriate, therefore, to try to link

---

[22] Scholarly resources like *The Lost Plays Database* (curated by Knutson and McInnis 2009-2015), the REED archives (ongoing since 2003), incl. the *Early Modern London Theatres*

the theories of language production introduced earlier with theories from the study of oral transmission of folk material – and in particular Milman Parry and Albert Lord's concept of the 'formula'.[23] The use of formulas in oral composition was probably the most revolutionary concept introduced in Lord and Parry's research on Homeric poetry and traditional folksong. In Parry's definition, a formula is 'a group of words which is regularly employed under the same metrical conditions to express a given essential idea'. Like the functional information they express, 'the standard formulas belong not to one singer, but to tradition' (1971, 80). In Lord's understanding, formulas are 'the phrases and clauses and sentences' of the poet's specialised poetic grammar, which he learns 'by hearing them in other singers' songs', where the process of memorisation is 'unconscious and follows the same principles as the learning of language itself' (2000, 36).

Such formulas are unquestionably linked to the chunks or multi-word units already discussed above. What makes Parry and Lord's perspective different is the centrality of *memory* and *tradition*. By placing the use of formulas firmly in the spoken language as expressions of functional verbal economy, the formulas come remarkably close to the 'permanent sets of associative connections in long-term memory…, which underlie the attainment of automaticity and fluency in language' (Miller 1956 in Ellis 2001, 38-39). 'Chunking', in this context, however, is seen not as a language mechanism, but as the main principle of human cognition: 'A chunk is a unit of memory organisation, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organisation of memory. Chunking appears to be a ubiquitous feature of human memory' (Newell 1990, 3-4).

That formulas occur more frequently in spoken language than in writing has subsequently been confirmed by corpus-based comparisons of written and spoken corpora (e.g. Brazil 1995; Biber *et al.* 1999; Leech 2000). This supposedly means that memory functions differently in oral composition than in writing (which is rightly speaking a form of artificial memory). Parry, Lord, and later Foley, were acutely aware of this aspect, but perhaps we may add some recent insights to theirs.

According to Chafe (1994), modern English utterances 'are constructed as intonation units that usually have a modal length of four words and that are often highly predictable in terms of their lexical concordance (Hopper

---

project (ed. by MacLean 2011), and the *Henslowe-Alleyn Digitisation Project* (Ioppolo 2005-2016) all serve as timely reminders of the unstable composition environment that constitutes the early modern stage(s). The text repository EEBO, in comparison, offers less such 'messy' meta data.

[23] The term 'formula' covers a vast number of technical terms in circulation for 'repetitions', or 'recurrent phrases', 'stock epithets', 'epic clichés', 'stereotyped phrases', 'bound phraseology' to name just a few (Lord 2000, 30).

1998)'.[24] Assuming that early modern English will have had a similar span, we can apply this constraint to actors' oral language production, too, only bearing in mind that on-stage verbalisation is constructed in real time, and so will have imposed even greater working-memory demands on the individual compared with writing. Consequently, we can expect an even greater reliance on fixed collocations in performance, whose lexical content link them with the long-term memory. It is, as Ellis, Simpson-Vlach and Maynard (2008, 376) state, simply 'easier for us [language users] to look something up from long-term memory than to compute it (Bresnan, 1999; Kuiper, 1996)', and even easier, perhaps, if those fixed collocations come in 3-4-gram 'natural' intonation units.

Optimal 'on stage' economy of expression, then, relies on knowing exactly which patterns to *infer/use* in which situation. If actors were *not* familiar with a stock of frequently used word associations, they would presumably struggle to complete their workday both in live, on-stage speech and in learning their parts. For similar reasons, situational formulas are replete in primary oral cultures. Moreover, in folk ritual, ballads and folksong there is a limitation to the form of such formulas (Ong 1982), which in turn allows for easy replication. If we apply this observation to the early modern stage, actors – like traditional oral performers – are likely to have 'found and kept expressions which could be used in a variety of scenes and situations, either as they stood or with slight modifications'. Foley calls this the 'compositional idiom' (1991, 23), suggesting an enhanced definition of the 'formula' as 'an expression regularly used under the same metrical conditions to express an essential idea, but where 'formula types or systems [can] replace one another, allowing for a high degree of economy of expression/ mnemonic economy' (25).

Whether or not Shakespeare and his colleagues wrote to produce literary drama (Erne 2003) or deliberately created texts out of other texts, borrowing and adapting already popular plots, the vast majority of the material they produced was intended for stage performance and was feasibly written with oral rendition or actors' memories in mind (Lancashire 1999, 736). This could mean, then, that authors like Shakespeare were doubly subject to collocation use, in the sense that they are *non-consciously* producing language according to a general 'idiom principle' but also *consciously* writing to accommodate actors' so-called *prehension*[25] on stage (cf. Foley's 'compositional idiom'). We ought thus perhaps to expect playtexts, like other oral genres, to contain an 'additive oral style' or 'high oral residue' (Ong 1982, 36-37):

[24] Chafe and Hopper are quoted in Ellis *et al.* 2008, 376.

[25] According to Firth, 'the collocation of a word or a "piece" is not to be regarded as mere juxtaposition; it is an order or mutual expectancy. The words are *mutually expectant* and mutually *pre*hended' (1957, 196; emphasis mine). Both Lancashire and Vickers refer to the concept of *prehension*, e.g. 'the wide use that natural language makes of such "embedding" produces a kind of *expectation* in the hearer that the rest of the associated "word-material" will soon appear' (Vickers 2012, 136).

The elements of orally-based thought and expression tend to be not so much simple integers as clusters of integers, such as *parallel terms or phrases or clauses*, *antithetical terms or phrases or clauses, epithets*. … In oral performance of folksongs, the performer prefers e.g. not the soldier, but the brave soldier; not the princess, but the beautiful princess; not the oak, but the sturdy oak. (Ong 1982, 38)[26]

Another advantage for the on-stage speaker would be to 'say the same thing, or equivalently the same thing, two or three times' when performing in front of a large audience. If the playgoers miss the 'not only…', they themselves can supply it by inference from the 'but also…' (Ong 1982, 40).

If the author-playwrights were in fact accomplices in this mechanism, the result may have been near-optimal performance conditions for the oral performers of the written scripts. It means that authors, (scribes) and actors could and would deliberately choose to deploy *similar* functional formulas (such as 'Look where he/she/it comes/goes/is') rather than different expressions for a given (entry) scene. Likewise, audiences would know what phraseology to expect in a number of similar situations. Such a collective stylistic contract is of course speculative. What we *do* know is that multi-word units are central in language production and reception, and so the use of formulas would have come natural to authors and actors (and audience members) alike. We also know that lexical chunks and collocations stored in the mental lexicon enable quick retrieval and both speed and ease communication. To this we can add recent research on priming, which suggests language users' sensitivity to the *frequencies* of occurrence of a wide range of different linguistic elements used within a specific field or domain. In such 'discourse communities', verbal knowledge is shared between speakers who can recognise the chunks used by each other. The shared knowledge enables the interlocutors to process language in a similar way, each 'member' sharing the knowledge of the whole field, industry, or domain. The fact that lexical priming is both a *reciprocal* mechanism and *reinforceable* by a speech community highlights, I believe, the importance of addressing both 'tradition' and 'collectivity' when analysing early modern playtexts. Ellis, applying a usage-based approach (1996, 2002a, 2002b), e.g. provides clear evidence of the influence of each 'usage event' of a given formula and the processing of its component constructions on speakers' language systems. If we grant the early modern stage status of a discourse community, it is likely to have been predicated on a degree of mutual interpretation where collective activities involving the members' interpretation of the vocabulary used take place. Fellow actors and authors' use of similar phraseology will then have primed subsequent usage, perhaps even to the extent that a collective store of phraseology or a

---

[26] Ong of course does not argue that this mechanism is not present in literary composition, only that it is more prevalent in oral composition. See also Ong 1982, 188-212.

'commulect' develops. This is not so far from Firth's notion of 'meaning by collocation', where familiarity with frequently used word associations arouses 'expectancies' in fellow language users (Firth 1957, 195-196).

At the very least, priming would serve some very functional, workaday purposes for the authors and actors in the industry. Similarly, the stakeholders of this community would of course have been highly familiar with cueing systems (see Palfrey and Stern 2007), and so might oddly enough have been more aware of the practical conditions of the human working memory and the brain's language-producing mechanisms than 'normal' language users. Actors and authors, both attuned to systematised cueing, may thus have been exceptionally well-suited to utilising these mechanisms quasi-consciously.

Usage-based theories of language are clearly of great value to understanding both early modern authors' and actors' use of formulas and the degree to which seeing/hearing the beginning of a formula will have primed recognition of its final word(s), collocates or constituents. It would certainly be interesting to look further into how frequency and repetition affect – and ultimately bring about – form in language and expression in the early modern playtexts,[27] and how this knowledge corresponds with general language production (Bybee and Hopper 2001; Ellis 2002a, 2002b, 2008a; Bod, Hay and Jannedy 2003; Hoey 2005, 2009; Robinson and Ellis 2008; Pulvermüller and Knoblauch 2009; Pulvermüller, Cappelle and Shtyrov 2013). Such an undertaking will have to wait, however; and instead the following pages provide some token examples of how we can readily use corpus resources to extract and analyse multi-word units of a potentially 'oral' or 'collective' formulaic nature. By learning more about these formulas, we can hopefully gain a better understanding of the patterns and collocations that characterise the vocabulary of the early modern stage as a whole, and, by implication, an understanding of what individual authors' idiolects are less likely to contain.

The following is a sample of commonly-occurring situational dramatic formulas extracted using the KEMPE corpus and search tools.[28] Using wildcard searches one can locate bigrams, 3-grams or 4-grams, or combinations where words colligate with specific POS in adjacent sentence slots. After this, the researcher is faced with a rather old-fashioned process of manually checking the semantic and syntactic properties of the formulas. Further post-processing includes checks for left or right-context collocates. In the examples below, simple occurrence frequencies are given in brackets. Looking beyond raw frequency

---

[27] See further Petersen 2008, 2010, and Pettitt 2001, 2005).

[28] The searches are carried out in KEMPE, and were chosen on the basis of searches for both lexical and grammatical multi-word units led by e.g. VERB, PRON (pronouns, interrogative), ADV (adverb, time), PREP (prepositions; time/space). The results were then manually checked for fomulaicity.

(and hence statistical significance) across the 9 million words in KEMPE, I treat the formulas as collocations following the phraseological approach (Cowie 1978, 1981; Mel'čuk 1998; Gitsaki 1999); that is, I am as interested in what the units *do* as in how many times they occur:

And so away	(7)
Away, away	(152)
Away, away, away	(9)
Away, away, away, away	(2)[29]
Come away (most frequent right-context collocates: quickly, with, I)	(216)
Come, come / come [noun] (most frequent right-context collocates: 'you', 'my')	(116)
Do you hear	(444)
Do you hear +pronoun (me/my/him/this…)	(67)
Do you hear me	(27)[30]
Far(e)well (,) far(e)well	(42)
Go to (go to) (both a direction and an expletive)	(1448)
Go(e) your ways	(180)
Hear me	(897)
Hear me speak	(104)
Hear me speak	(104)
Hear(e) me (most frequent right-context collocates: 'but', 'sir', 'lady')	897
How now	(1834)
How now + noun (mainly proper nouns)	(752)
Leave me alone	(10)
Let's away /let us away	(139)[31]
Look to (most frequent right-context collocates: your, the, him, my, it)	(570)
Look to + pronoun (it/her, etc.)	(334)[32]
Look where	(81)[33]
Now will I + verb akin to 'leave'	(14)[34]
Now, let me/us	(64)
Now, Sir/sirs/sirrha/sirra	(503)
Run(ne) to	(99)
Saddle my/your horse	(10)
See where (he/she/it/pronoun comes/goes/is/verb)	(286)[35]

[29] Both instances are in *The Puritan Widow*, 1606/1607.

[30] In 7 instances the adjacent right-edge collocate (NOUN colligation) is a proper name or title.

[31] This formula is related to/a variant of 'come let's away to …'.

[32] Expressed as [word="look" %c] [word="to" %c][pos="PRON" %c].

[33] Expressed as [word="look.*" %c] [word="where" %c] [word=".*" %c] [word=".*" %c] [word=".*" %c].

[34] Expressed as [word="now" %c] [word="will" %c] [word="I" %c] [lex="hence" | lex="away.*" | lex="go" | lex="lea.*"].

[35] Expressed as [word="see" %c] [word="where" %c] [word=".*" %c] [word=".*" %c] [word=".*" %c].

We [verb] away (to)[36]
Well met   (166)[37]
What new(e)s   (558)
What new(e)s from   (31)
Why how now +noun   (170)
Why(,) how now   (261)
Within this hour   (71)

Consider the 'Well met' formula as an example. Using the KEMPE search tools, we can locate the most frequent right-context items after 'met' (both collocations and colligations may be quantified here). Most frequent are the function words 'my', 'at', 'in' and 'if'). Of content words we find 'gentlemen', 'maister' and 'sir'; the most frequent adjectival collocates include 'fair' and 'good'. Looking at the collocates teaches us something of the process-oriented, stage-related mechanisms behind the formulas. These combinations are easily assembled/*prehended* by actors' working short-term memory aided by long-term lexical storage, and may thus be present in the texts at the instigation of both authors and actors. Other formulaic units include simple or incremental internal repetition (Andersen *et al.* 1982), e.g.:

Alack (alack) the day   (30)
ay, ay (I, I)   (at least 25)
How, how   (at least 88)
I [Verb] I   (1604)[38]
O, o – / Oh, Oh …   (at least 163)

Moving from vast data sets to single playtexts, we can use the corpus and collocation extraction algorithm of another resource, namely Mueller's WordHoard,[39] to catalogue multi-word units in the multi-textual and co-authored play *Titus Andronicus*.[40] For *Titus*, a total of 2,786 multi-word units pass all the program's filters. Because of the authorship question, *Titus* remains a prime test

---

[36] A case in point: the fact that 'we will away to' only appears is in *The Troublesome Reign of King John* and the *F* and *Q* versions of *The Merchant of Venice* is interesting, but does not sway me that we are talking about an *authorial* '4-gram' match.

[37] The most frequent immediately adjacent right-edge collocates of 'well met', as identified by the sort function in KEMPE, are: 'by' followed by 'at', 'in', and 'my'.

[38] Expressed as [word="I" %c] [pos="V" %c] [word="I" %c].

[39] I use the publically available Shakespeare corpus in WordHoard (texts sourced from *The Globe Shakespeare*). Mueller's larger *Shakespeare His Contemporaries* corpus, where texts are sourced from the TCP and enhanced with dense meta-data search potential, is still a closed environment. I, for one, look forward to the time when a 'user-friendly version of it will be up and running in an open access environment that will cut down on the tedium of some older forms of exploration and enable new forms of exploration that previously were impracticable' (Mueller 2014b).

[40] Mueller's version of *Titus Andronicus* is assumed to be a Folio version, checked against Q variants, as per *The Globe Shakespeare*.

bed for new(er) methodologies of authorship attribution, including, of course, tests for multi-word units. Below follows a sub-sample from amongst the extracted bi-grams and 3-grams, which I would characterise as examples of verbal repetition or simple commonplaces (these can be oral or authorial in kind). Where a unit looks potentially formulaic, I have used the KEMPE search tools, including POS queries, to check for the prevalence of such phrases across the entire corpus. In the same way as above, I have not limited these searches in date since the aim is to find formulaic phrases promulgated by dramatic transmission over time and across canons. Locating a 'formula' in both early and late plays probably testifies to its usefulness; it may also mean that the formula was written for *prehension*. Alternatively, actors' transmission of a text may have caused oral formulas to become embedded in surviving printed editions (Petersen 2008, 2010). Either way, such formulas very probably belong to both authors *and* actors:[41]

An/d if it please…	>100
Away with	>50
Come let us go/e & Come let(')s go/e	>50
Get you (ye) gone (gon)	>100
Go your ways	180
Ha, ha	<100
I'll go fetch	>100
Let me alone	>300
Let me see	>500
Let us/let's go/e	>300
Nay, nay / nay nay	>50
Now will I (+transitive verb)[42]	>100
Well I wote	>100
What's the news …	>100

For the similarly textually challenging case of *Hamlet*,[43] a total of 4,173 multiword units pass all the WordHoard filters. Of these I have selected a number of potentially oral or collective formulas, which have then been checked for commonality across the early modern canons using KEMPE:

[41] Frequencies are expressed as 'greater than' (>100) / 'less than' (<100) approximations, allowing for matches *not* located by the KEMPE search algorithms.

[42] 'Now will I hence' occurs 4 times across the early modern canons contained in KEMPE. The search string applied locates units like 'now will I haste', 'Now will I away ...', 'Now will I leave …', ' Now will I high me...' / '...turn and run', etc. A majority of the immediate right-edge colligates of 'Now will I' are nouns, i.e. objects.

[43] Text from *The Globe Shakespeare*. See notes 10 and 39.

How now	>1800
How now + noun[44]	>750
How now what / How now, what	>250
I have new(e)s to tell you	<50
Look you now	<50
Look(e) where	>50
Now my lord	>100
To a nunnery go[45]	

Using Seretan's FipsCoWeb,[46] I looked for similar instances of formulaic multi-word units in Marlowe's *Doctor Faustus* (*A* and *B* texts), where both texts are assumed to have been exposed to transmission, and possibly also co-authorship (Maguire 1996; Pettitt 1988; Vickers 2012. Vickers asserts that the play is undoubtedly co-authored, 2012, 134). The units identified were once again submitted to checks for cross-canon prevalence using KEMPE. Asterisks mark instances where formulas reoccur internally in the texts and/or across the entire corpus:

Faustus A-text
2* here; come
3 thy;;head;;
2* see;;Pope;;
2 make;;world;
2* mighty;;Lucifer;;
3* great;;Lucifer;;
2* great;;thing;;
2 not;;tell;
2 dead;;time;;
2 raise;;spirit;
2 shoulder;of;mutton;

Faustus B-text
4 thine;;eye;
2 highly;;solemnize;
4* great;;Lucifer;;
2* gentleman;;farewell;;
3 doctor;;have;;
3* see;;Pope;;
2* here;;take;;
2* but;;tell;;
2* sweet;;friend;;
2* now;;tell;;
2 make;;world;;
2 have;;leg;;
2 thy;;life;;

B-text continued:
2 have;;grape;;
2 Pope;;have;;
2 world;;see;;
2 thy;;head;;
2 thy;;body;;
2 beauteous;;paramour;;
2* enter;;angel;;
2 high;;firmament;;
2 lake;of;mud;;
2* mighty;;Lucifer;;
2* plague;;take;;
2 ripe;;grape;;
2 take;;guilder;;
2 world;;admire;;
2* nay;;stay;;
2* then;;wilt;;
2 as;;chary;;
2 in any case;;ride;;

[44] The wild-card search string [word='how' %c] [word='now' %c] [pos='N' %c] [] [word='what' %c] [word='new.*' %c] neatly extracts a 'How now, NN, what news' formula.

[45] There are 5 instances of the unit in F1 *Hamlet* but 8 'to a nunnery goe' in Q1; a text we may assume has been much transmitted (see e.g. Maguire 1996; Petersen 2010). I have elsewhere classified this as a remarkable example of internal incremental repetition and a relatively sure sign that this text holds evidence of transmission-induced oral-memorial style markers (Petersen 2010). Incidentally, 'To a nunnery/nunnerie' also appears in Marlowe's *Jew of Malta* (twice) and in Dekker's *The Welsh Ambassador*.

[46] This user-friendly resource may be accessed at Seretan and Wehrli 2011.

In *The Massacre at Paris* and *Edward II*, also from the Marlowe canon, we are similarly faced with plays with a complicated textual provenance, and where co-authorship of at least one of the texts (*Edward II*) is likely. Again asterisks mark elements that qualify as formulaic in the KEMPE corpus:

The Massacre at Paris		*Edward II*			
11	enter;;king;	3	stay;;while;	2	long;;king;
2	defend;;right;	2	will;;lordship;	2	thy;;sight;
4*	come;;lord;	2	legate;of;Pope;	3*	see;;king;
2	gather;;power;	2	close;;eye;	2	other;;king;
2	good;;morrow;	2	country;;cause;	2	not;;see;
2	country;;good;	2	commit;to;tower;	2	king;;good;
2	noble;;man;	2	noble;;gentleman;	2	thy;;father;
2	be;;good;	2	rend;;hart;	2	thy;;sword;
3*	come;;let;	7	enter;;king;	3	not;;be;
6	enter;;guise;	2	accursed;;head;	3	thy;;life;
2	enter;;messenger;	2	enter;;bishop;	2	thy;;hart;
2	thy;;hand;	2	noble;;birth;	2	thy;;brother;
3	thy;;death;	4	lord;;king;	2	thy;;word;
2	power;of;man;	2	so;;passionate;	2	not;;come;
2	thy;;life;	2	therefore;;trust;	3	thy;;friend;
2*	but;;come;	2*	come;;let;	2	thy;;head;
2	so;;be;	2*	but;;tell;	4	thy;;king;
2	thy;;brother;	2	friend;;do;	2	be;;king;
2*	then;;come;	2	enter;;noble;	2	good;;sir;
2	body;of;king;	2*	here;;come;	2	flying;;fish;
2	guise;;come;	2	honour;of;name;	2	hang;about;neck;
2	make;;king;	2	man;of;birth;	2	subscribe;;name;
2	English;;agent;	2	enter;;baron;	2	thy;;message;
2	French;;king;	2	king;;die;	2;	troublesome;;death;
2	tell;;king;	2*	now;;let;	2	wanton;;humour;
2	reading;of;letter;	2*	come;;prince;	2	kill;;good;
2	rebellious;;king;	2	thy;;face;	2	not;;suffer;
2	humbly;;thank;	2	not;;trust;	2*	away;;base;

2	exit;;king;	2	noble;;man;	2	hand;;traitor;
2	and so;;will;	2	thy;;land;	2	for;;till;
2	do;;deed;				

The very noticeable 'come lord/s', 'come let/s', 'but then', and 'then come' formulas in *Massacre,* like the high number of other repeating units, offer likely evidence that the play has been much transmitted in performance. According to Lord, 'fixed expressions of this sort and of other sorts can be found occasionally in print, indeed can be "looked up" in books of sayings, but in oral cultures they are not occasional. They are incessant' (2000, 34-35). The high degree of formulaicity of *Edward II* likewise suggests that a communicative economy principle is at work in this text. Whether introduced by authors, co-authors, or actors, these units are found amongst the collocations extracted by the FipsCo application (and would be by similar software). It takes manual analysis to exclude/disqualify the collocations as authorial markers.

Not enough has been written about exactly this kind of *functional* multi-word unit nor of these units' relation to the composition practices of the early modern stage. As Pettitt noted some 15 years ago,

We are accustomed to seeking reassurance in the strength and capacity of the memory in oral cultures, which the Elizabethan still partly was, but it is by no means certain that the oral memory is inevitably geared to the *verbatim reproduction* of texts. Nor is it certain, given the *ambiguous status of the playwright* in this particular phase of theatre history, that the *verbatim reproduction of his text*, as opposed to keeping going and keeping the audience satisfied, *was a decisive consideration* with the players. (2001, 414; emphasis mine)

I would like to think that Shakespeare and his colleagues were fully aware of these conditions. Like Ian Lancashire (1999, 736), I would also like to think that Shakespeare, being an actor himself and so accustomed to memorising sound, wrote particularly for the stage (or in a particular way for the stage). But we cannot know this. Judging by the data just compiled there is reason to assume, however, that early modern authors writing for performance at least to some degree made use of collective dramaturgical formulas, and as such were *writing for prehension*.[47] In the case of authors like Shakespeare, who are close to the theatre, the influence on texts of prior and ongoing performance tradition seems inescapable. Perhaps when considering Shakespeare and his contemporaries' composition habits we should assume, then, a 'strong' variant of Parry and Lord's oral-formulaic theory, in which performative transmission effectively partakes of the 'literary' composition function, and composition in pen or quill partakes of the transmission function?

[47] See also Pettitt 1988, 168-169.

With regard to the exact ways in which formulaicity is introduced into the sample material, folkloristics may be no better equipped than orthodox philology at determining authors' writing habits, at detecting theatrical revision, or mapping the routines of ostensible memorial reporters. What a folkloristic approach *can* help diagnose – according to Pettitt – is the

type of change introduced by the performers while they are the bearers of the living verbal tradition, consciously or more likely unconsciously, under the pressure of reconstructing a text from memory in the stress and confusion of live performance before an audience who are under no obligation to remain respectful or attentive. This is what happens—if more slowly and over a longer period—to folktales and ballads. The impact of these processes is not haphazard and leaves quite distinct symptoms. In the case of folk ballads, such symptoms can be convincingly identified, and when the same symptoms occur in the text of a play, it is reasonable to conclude that it has been subjected to similar processes. Common sense, anecdotal evidence, and strict theorizing can, in ballad studies as in Shakespearean philology, suggest what is likely to happen to orally transmitted texts: omissions, garblings, anticipations, and improvisations. But to prove anything we need to compare two texts, one of which is known to be an oral derivative (at one or more removes) of the other. The differences, particularly if repeated in analogous experiments with other text-pairs, are the result, and so symptoms, of transmission. (2005, 216)

As an illustration of the above conditions, we can look at a transmitted broadside ballad, previously used by Pettitt to exemplify the accumulative use of formulas in oral-memorial transmission. Other examples may be consulted in Petersen (2010, 59-61). In this example, we see a broadside ballad composed by a 'ballad hack' and printed in 1828 (within a year of the events described in the song). It is flanked by a derivative version recorded from oral tradition by Cecil Sharp almost three quarters of a century later:[48]

1828 Verses 4, 5, 6, 7, 11, 12	**1911** Verses 1, 2, 3, 4, 5, 6
…	…
If you'll meet me at **the Red Barn**	If you'll meet me at **the Red Barn Floor**
As sure as I have life	As sure as you're alive
I will take you to Ipswich town	I'll take you down to Ipswich Town
And there make you my wife	And make you my dear bride

[48] The broadside version was first printed by James Catnach in 1828 and sold 1,116,000 copies. The derivative was sung by a certain Robert Feast at Ely. The example is quoted by Pettitt in Andersen *et al.* 1982, 77-83, and in 'The Living Text' (Pettitt 2001, appendix 1). Another oral version, collected from Joseph Taylor of Lincolnshire in 1908, retains only verses 1, 2, 7.

I then went home and fetched my gun	He straight went home and fetched his gun
my pickaxe and my spade	His pickaxe and his spade,
I went unto the **Red Barn**	He went unto **the Red Barn Floor**
And I dug her grave	and he dug poor Maria's grave
With her heart so light she thought no harm	This poor girl she thought no harm
To meet me she did go	But to meet him she did go
I murdered her **all in the barn**	She went unto **the Red Barn Floor**
And laid her body down	And he laid her body low
…	…
Her mother's mind being sore disturbed	Her mother dreamed three dreams one night
She dreamed a dream she saw	She ne'er could get no rest
Her **daughter she lay murdered**	She dreamed she saw her **daughter dear**
Beneath the Red Barn floor	**Lay** bleeding at the breast
She sent the father to the Barn	Her father went into the barn
Where he the ground did thrust	And up the boards he took
And **there he found his daughter**	There he saw his **daughter dear**
Lay mingling with the dust	**Lay** mingled in the dust.

Clearly, we are looking at multi-word units/collocations (in bold) in the above parallel texts of original and derivative ballad versions. They do not belong to an author, though, or indeed one singer/performer, but to a collective tradition, and the units are, presumably, transmission-induced.

In collaboration with Pettitt, and elsewhere (Petersen 2008, 2010), I have exemplified the accumulation of similar incremental patterns in the course of transmission of early modern plays. But such an *accumulation mechanism* is very easy to miss or disregard when studying multi-word units *per se* in the early modern canons. In the first quarto of *Hamlet*, for example, the accumulated multi-word units range from a substantial number of simple verbal repetitions (we could call them bi-grams) to large-scale repetition (3 and 4-grams), now verbally homogenised in the Q1 version. None of these units are present in the - presumably earlier - *Folio* text (Bate 2007). A full catalogue of examples can be consulted in Petersen (2010), but a few instances of the 'new' 3-grams and 4-grams in Q1 *Hamlet* may serve as examples here:[49]

[49] Also in Petersen 2010, 96.

F1
1.4.68 Ham. 'It waues me forth againe; **Ile follow it**.'
1.4.79 Ham. 'It wafts me still; goe on, **Ile follow thee**.'
1.4.86 Ham. 'I say away, goe on, **Ile follow thee**.'

1.5.186 Ham. '...**Let us goe in together**.'
1.5.190 Ham. '**Nay, come let's goe together**.'

4.3.12 'Where **the dead body** is bestow'd my
        Lord, / We cannot get from him."
4.3.16 'Now Hamlet, where's Polonius?"
4.3.32 'Where is Polonius?"

5.2.283 '...**Here's** to thy health.'
5.2.288 '**Here's** a Napkin, rub thy browes.'
5.2.289 'The Queene carowses to thy fortune, Hamlet.'

Q1
459 Ham. 'Still I am called, go on, **ile follow thee**.'
465 Ham. 'Go on, **ile follow thee**'
471 Ham. 'Away I say, go on, **ile follow thee**'

643 Ham. "**Nay come lett's go together**.'
647 Ham. '**Nay come lett's go together**.'

1626 '…we can by no means know of him where the
       body is.'
1627 '...where is this dead body?'
1640 '...where is this body?'

2155 'Here Hamlet, the King doth drink a health to thee'
2156 'Here Hamlet, take my napkin, wipe thy face'
2160 'Here Hamlet, thy mother drinkes to thee.'

As Pettitt (1988, 184-185) explains, the replacement of non-formulaic expressions with formulas can sometimes be detected in those instances where we are lucky enough to have variant texts of a single ballad. Evidently, such a mechanism is at work in the multiple texts of *Hamlet*. But similar features can also be found in a number of other, supposedly unproblematic, playtexts in the early modern canons. Both corpus design and data extraction methods should be more attuned to these essentially contextual conditions.

4. *Epilogue*

Few of the collaborative plays analysed in recent authorship attribution studies conform to any easy definition of authorial composition (Maguire 1996, Vickers 2002, Erne 2003, Petersen 2010). Collaborative practises involve a number of stakeholders – author, actors, and theatrical scribes, for a start. If the play is successful, it almost always involves textual change (spoken, and sometimes recorded) over time, and it involves tradition, as both players and playwrights could resort to dramatic formulas and verbal borrowing from other plays on similar topics. It may not always be the case that these features are recorded in the playtexts under scrutiny, but they potentially might be, like some of the examples from Vickers' study of the *Additions* to *The Spanish Tragedy* (2012). This means that otherwise sound stylistic attribution testing may be 'blurred' by oral or collective features. I have argued that it is necessary to further consider this cross-cutting verbal fomulaicity if stylistic investigation into early modern authorship is to be successful.

       The approaches of recent years somehow fall short of establishing entirely certain stylistic evidence for the author vs. the actors – and it has been difficult to determine how to proceed and which methodologies to apply. In a discourse

community like the early modern stage, which is very much characterised by tradition (because of its medieval, popular roots and because of the pressures on dramatic authors), it is very likely that a stock of verbal and dramatic formulas are operative in written composition as well as re- and de-composition in performance. As I hope to have illustrated, the multi-word formula appears to be as useful to an author under pressure as to an actor under pressure. The *prehension* of word material thus potentially becomes an operating principle rather than something secondary to stylistic investigation. There are, as I see it, three types of potential contributors to the shape of a surviving early modern playtext: the author (always), the actors (sporadically – in the *surviving* texts), and tradition (almost always); tradition being the entire vocabulary of all the plays written before the given text under scrutiny, within the same cultural system, and which will directly or indirectly influence authors' (scribes'), and actors' choice of words.

The language written for and spoken on the early modern stages evidently deals as much in *probable* language as it does in *possible* language. Among the many multi-word units (lexical and grammatical) in use in the plays, some are bound to stem from individual authors' idiolects. Quite possibly it will be the *rare* (i.e. possible) *lexical* collocations that we should turn to as appropriate authorial style markers (cf. note 7 and Vickers 2011a). The grammatical collocations of the type 'look where he/she/it comes/goes/is', by contrast, are highly probable choices and as such much more likely to qualify as collective or oral in origin. Other unfixed, non-syntactical units should perhaps be discarded altogether as evidence for authorship until such strings can be linked more closely to the known neurolinguistic processes in the human brain. All other things being equal, more controls and more description is needed of the multi-word unit as style marker. It may yet hold the key to the success of future attribution studies. Meanwhile, I remain convinced that attribution scholars, whose only available sample material remains the deceptively neat digital editions of the once 'live' playtexts of the early modern stage, 'need to better understand how they came about' and more fully 'engage with the manifold questions surrounding their production and transmission' (Craig 2014, 15). *The play*, in other words, *is still the thing, wherein ...*

Works Cited

Printed Sources

Andersen F.G., Otto Holzapfel and Thomas Pettitt (1982), *The Ballad as Narrative*, Odense, Odense University Press.
Bate Jonathan (2007), 'The Case for the Folio', <http://goo.gl/ei3fE5>, accessed 4 March 2016.
Benson Morton, Evelyn Benson and R.F. Ilson (1986), *The BBI Combinatory Dictionary of English*, Amsterdam, John Benjamins.

Biber Douglas, Susan Conrad, Edward Finegan, Geoffrey Leech and Johansson Stig (1999), *Longman Grammar of Spoken and Written English*, Harlow, Pearson Education.

Bod Rens, Jennifer Hay and Stefanie Jannedy (2003), *Probabilistic Linguistics*, Boston, MIT Press.

Brazil David (1995), *A Grammar of Speech*, Oxford, Oxford University Press.

Bresnan Joan (1999), *Linguistic Theory at the Turn of the Century*, Plenary presentation. Paper presented at the 12th World Congress of Applied Linguistics, Tokyo, Japan, August 1999.

Burgess Neil and G.J. Hitch (2006), 'A Revised Model of Short-Term Memory and Long-term Learning of Verbal Sequences', *Journal of Memory and Language* 55, 4, 627-652.

Burrows J.F. (2012), 'A Second Opinion on "Shakespeare and Authorship Studies in the Twenty-First Century" ', *Shakespeare Quarterly* 63, 3, 355-392.

Bybee J.L. and Paul Hopper (2001), *Frequency and the Emergence of Linguistic Structure*, Amsterdam, John Benjamins.

Chafe W.L. (1994), *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*, Chicago, University of Chicago Press.

Cowie A.P. (1978), 'The Place of Illustrative Material and Collocations in the Design of a Learner's Dictionary', in P. Stevens, ed., *In Honour of A.S. Hornby*, Oxford, Oxford University Press, 127-139.

Cowie A.P. (1981), 'The Treatment of Collocations and Idioms in Learner¹s Dictionaries', *Applied Lingustics* 2, 223-235.

Craig Hugh (2014), 'Mingled Yarn', in T. Bishop, A. Huang and B.D. Hirsch, eds, *Digital Shakespeares*, *The Shakespearean International Yearbook* 14, 3-37.

Craig Hugh and A.F. Kinney (2009), *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge, Cambridge University Press.

Ellis N.C. (1996), 'Sequencing in SLA: Phonological Memory, Chunking, and Points of Order', *Studies in Second Language Acquisition* 1, 18, 91-126.

Ellis N.C. (1997), 'Vocabulary Acquisition: Word Structure, Collocation, Grammar, and Meaning', in M. McCarthy and N. Schmidt, eds, *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge, Cambridge University Press, 122-139.

Ellis N.C. (2001) 'Memory for Language', in P. Robinson, ed., *Cognition and Second Language Instruction*, Cambridge, Cambridge University Press, 33-68.

Ellis N.C. (2002a), 'Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition', *Studies in Second Language Acquisition* 24, 143-188.

Ellis N.C. (2002b), 'Reflections on Frequency Effects in Language Acquisition: A Response to Commentaries', *Studies in Second Language Acquisition* 24, 297-339.

Ellis N.C. (2008a), 'Phraseology: the periphery and the heart of language', in F. Meunier and S. Grainger, eds, *Phraseology in Language Learning and Teaching*, Amsterdam, Benjamins, 1-13.

Ellis N.C. (2008b), 'Usage-based and form-focused SLA: The implicit and explicit learning of constructions', in A. Tyler, K. Yiyoung and M. Takada, eds, *Language in the Context of Use: Cognitive and Discourse Approaches to Language*, Amsterdam, Mouton de Gruyter, 93-120.

Ellis N.C., Rita Simpson-Vlach and Carson Maynard (2008), 'Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, Psycholinguistics and TESOL', *TESOL Quarterly* 42, 3, 375-396.

Erne Lukas (2003), *Shakespeare as Literary Dramatist*, Cambridge, Cambridge University Press.

Firth J.R. (1957), *Papers in Linguistics 1934 – 1951*, Oxford, Oxford University Press.

Firth J.R. (1968), 'Linguistic analysis as a study of meaning', in F.R. Palmer, ed., *Selected Papers of J.R. Firth*. Harlow, Longman, 12-26.

Foley J.M. (1991), *Immanent Art. From Structure to Meaning in Traditional Oral Epic*, Bloomington and Indianapolis, Indiana University Press.

Forker C.R., ed. (2011), *The Troublesome Reign of John King of England*, Manchester, Manchester University Press.

Gitsaki Christina (1999), 'Teaching English collocations to ESL students', *NUCB Journal of Language Culture and Communication* 1, 3, 27-34.

Halliday M.A.K. and R.H. Robins, eds (1966), *In Memory of J.R. Firth,* London, Longman.

Hoey Michael (1991), *Patterns of Lexis in Text*, Oxford, Oxford University Press.

Hoey Michael (2005), *Lexical Priming: A New Theory of Words and Language*, London, Routledge.

Hoey Michael (2009), 'Corpus-driven Approaches to Grammar: The Search for Common Ground', in U. Römer and R. Schulze, eds, *Exploring the Lexis Grammar Interface*, Amsterdam, John Benjamins, 33-47.

Hoover D.L. (2003), 'Frequent Collocations and Authorial Style', *Literary and Linguistic Computing* 18, 3, 261-286.

Hopper P.J. (1998), 'Emergent Grammar', in M. Tomasello, ed., *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, Mahwah, Erlbaum, 155-176.

Ioppolo Grace, ed. (2005-2016), *The Henslowe Alleyn Digitisation Project*, <http://goo.gl/xKuF3T>, accessed 4 March 2016.

Knutson R.L., David McInnis and Matthew Steggle, eds (2009-2016), *The Lost Plays Database*, <http://goo.gl/95YDn6>, accessed 4 March 2016.

Kuiper Koenraad (1996), *Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters*, Mahwah, Erlbaum.

Lancashire Ian (1996), 'Phrasal Repetends in Literary Stylistics: Shakespeare's *Hamlet* III.1.', in S. Hockey and N. Ide, eds, *Research in Humanities Computing 4: Selected Papers from the ALLC/ACH Conference*, Christ Church, Oxford, 34-68.

Lancashire Ian (1997), 'Empirically Determining Shakespeare's Idiolect', *Shakespeare Studies* 25, 171-185.

Lancashire Ian (1999), 'Probing Shakespeare's Idiolect in "*Troilus and Cressida*", 1.3.1-29', *University of Toronto Quarterly* 68, 3, 728-767.

Lancashire Ian (2004), 'Cognitive Stylistics and the Literary Imagination', in S. Schreibman, R. Siemens and J. Unsworth, eds, *A Companion to Digital Humanities*, London, Blackwell, 397-414.

Leech Geoffrey (2000), 'Grammars of Spoken English: New Outcomes or Corpus-oriented Research', *Language Learning* 50, 675-724.

Lord A.B. (2000 [1960]), *The Singer of Tales*, Cambridge-London, Harvard University Press.

MacLean Sally-Beth, dir. and gen. ed. (2011), *REED/Early Modern London Theatres*, <http://goo.gl/B0wKDm>, accessed 4 March 2016.

Maguire L.E. (1996), *Shakespearean Suspect Texts*, Cambridge, Cambridge University Press.

Mel'čuk I.A. (1998), 'Collocations and Lexical Functions', in A.P. Cowie, ed., *Phraseology: Theory, Analysis and Applications*, Oxford, Clarendon Press, 23-53.

Miller G.A. (1956), 'The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information', *Psychological Review* 63, 2, 81-97.

Mueller Martin (2004-2013), General Editor, WordHoard, <http://goo.gl/xYWSBy>, accessed 4 March 2016.

Mueller Martin (2014a), 'Repeated n-grams in Shakespeare His Contemporaries (SHC)', in Martin Mueller, *Scalable Reading. Dedicated to DATA: digitally assisted text analysis*, Sep 3 2014, <http://goo.gl/xYWSBy>, accessed 4 March 2016.

Mueller Martin (2014b), 'Shakespeare His Contemporaries: collaborative curation and exploration of Early Modern drama in a digital environment', in *Digital Humanities Quarterly*, 8.3 <https://goo.gl/EHptyu>, accessed 4 March 2016.

Newell Allen (1990), *Unified Theories of Cognition*, Cambridge, Harvard University Press.

Ong W.J. (1982), *Orality and Literacy: The Technologizing of the Word*, London-New York, Methuen.

Palfrey Simon and Tiffany Stern (2007), *Shakespeare in Parts*, Oxford, Oxford University Press.

Parry Milman (1971), *The Making of Homeric Verse: The Collected Papers of Milman Parry*, ed. by A. Parry, Oxford, Clarendon Press.

Pawley Andrew and Frances Syder (1983), 'Two Puzzles for Linguistic Theory', in J. Richards and R. Schmidt, eds, *Language and Communication*, London, Longman, 191-226.

Petersen L.B. (2004), *KEMPE: Korpus of Early Modern Playtext in English*, <http://goo.gl/F48ldA>, accessed 4 March 2016.

Petersen L.B. (2008), 'De-composition in Popular Elizabethan Playtexts', *Oral Tradition* 23, 1, 118-147.

Petersen L.B. (2010), *Shakespeare's Errant Texts*, Cambridge, Cambridge University Press.

Pettitt Thomas (1988), 'Formulaic Dramaturgy in Doctor Faustus', in R. Gill, K. Friedenreich and C.B. Kuriyama, eds, *A Poet and a Filthy Play-maker*, New York, AMS, 167-191.

Pettitt Thomas (2001), 'The Living Text: The Play, the Players, and Folk Tradition', *Leeds Studies in English* n.s., 32, 413-429.

Pettitt Thomas (2005), 'Marlowe's Texts and Oral Transmission: Towards the Zielform', *Comparative Drama* 39, 2, 213-242.

Pulvermüller Friedemann (2010), 'Brain Embodiment of Syntax and Grammar: Discrete Combinatorial Mechanisms Spelt out in Neuronal Circuits', *Brain and Language* 112, 3, 167-179.

Pulvermüller Friedemann and Yuri Shtyrov (2003), 'Automatic Processing of Grammar in the Human Brain as Revealed by the Mismatch Negativity', *Neuroimage* 20, 1020-1025.

Pulvermüller Friedemann and Andreas Knoblauch (2009), 'Discrete Combinatorial Circuits Emerging in Neural Networks: a Mechanism for Rules of Grammar in the Human Brain?', *Neural Networks* 22, 1, 161-172.

Pulvermüller Friedemann, Bert Cappelle and Yuri Shtyrov (2013), 'Brain Basis of Meaning, Words, Constructions, and Grammar', *Oxford Handbooks Online*, <http://goo.gl/6VlcMK>, accessed 4 March 2016.

Robinson Peter and N.C. Ellis (2008), *Handbook of Cognitive Linguistics and Second Language Acquisition*, London, Routledge.

Seretan Violeta (2004), 'Using the Web as a Corpus for the Syntactic-based Collocation Identification', in M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa, R. Silva, eds, *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC 2004), 1871-1874.

Seretan Violeta (2009), 'An Integrated Environment for Extracting and Translating Collocations', in M. Mahlberg, V. González-Díaz and C. Smith, eds, *Proceedings of the Corpus Linguistics Conference CL 2009*, Liverpool, University of *Liverpool*, <http://goo.gl/sbG2gy>, accessed 4 March 2016.

Seretan Violeta (2011), *Syntax-Based Collocation Extraction*, vol. XXXIV, *Text, Speech and Language Technology Series*, Dordrecht, Springer.

Seretan Violeta (2013), 'A Multilingual Integrated Framework for Processing Lexical Collocations', in A. Przepiórkowski, M. Piasecki, K. Jassem and P. Fuglewicz, eds, *Computational Linguistics*, vol. 458, *Studies in Computational Intelligence*, Berlin, Springer, 87-108.

Seretan Violeta, Luka Nerima and Eric Wehrli (2004), 'Using the Web as a Corpus for the Syntactic-based Collocation Identification', in M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa and R. Silva, eds, *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004, Lisbon, may 26-28)*, Paris, ELRA (European Language Resources Association), 1871-1874.

Seretan Violeta and Eric Wehrli (2006), 'Accurate Collocation Extraction Using a Multilingual Parser', in *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, Association for Computational Linguistics, 953-960.

Seretan Violeta and Eric Wehrli (2010), 'Tools for Syntactic Concordancing', in M. Ganzha and M. Paprzycki, eds, *Proceedings of the International Multiconference on Computer Science and Information Technology*, IEEE, 493-500.

Seretan Violeta and Eric Wehrli (2011), *FipsCoWeb* <http://goo.gl/f7b6XM>, accessed 4 March 2016.

Sinclair John (1966), 'Beginning the Study of Lexis', in C.E. Bazell, J.C. Catford, M.A.K. Halliday, R.G. Robins and J. Sinclair, eds, *In Memory of J.R. Firth*, London, Longman, 410-430.

Sinclair John (1991), *Corpus Concordance Collocation*, Oxford, Oxford University Press.

Sinclair John (1996), 'The Search for Units of Meaning', *Textus* 9, 1, 75-106.

Vickers Brian (2002), *Shakespeare Co-Author*, Oxford, Oxford University Press.

Vickers Brian (2011a), 'Shakespeare and Authorship Studies in the Twenty-First century', *Shakespeare Quarterly* 62, 106-142.

Vickers Brian (2011b), 'Unique Matches of Three Consecutive Words in *The Troublesome Reign* with Comparable Strings in Other Plays by Peele', in C.R. Forker, ed., *The Troublesome Reign of John, King of England*, Manchester, Manchester University Press, Appendix 2, 335-336.

Vickers Brian (2012), 'Identifying Shakespeare's Additions to *The Spanish Tragedy* (1602): A New(er) Approach', *Shakespeare* 8, 13-43.

Wagonheim S.S., ed. (1989 [1964]), *Annals of English Drama 975-1700*, 3rd ed., London-New York, Routledge.

Wiggins Martin (2011), *British Drama, 1533-1642: A Catalogue*, Oxford, Oxford University Press.