

The final version of: Culpeper, Jonathan and Samuel Oliver (2020) Pragmatic noise in Shakespeare's plays. In: Ewa Jonsson and Tove Larsson (eds.) *Voices past and present - Studies of involved, speech-related and spoken texts*. Amsterdam: John Benjamins. It may contain minor errors and infelicities.

Pragmatic noise in Shakespeare's plays

Jonathan Culpeper and Samuel J. Oliver

Abstract

Pragmatic noise, first coined in Culpeper and Kytö (2010), refers to the semi-natural noises, such as *ah*, *oh*, and *ha*, that have evolved to express a range of pragmatic and discoursal functions. Taking advantage of the regularised spellings and grammatically tagged texts of the *Enhanced Shakespearean Corpus* (Culpeper 2019), this study considers the frequency, distribution and functions of pragmatic noise across Shakespeare's plays and characters. It reveals and discusses, for example, the facts that: whilst particular types of pragmatic noise maintain a steady presence across all the plays, there is variation in token density; female characters have a much greater density of pragmatic noise tokens compared with male; and characters in the middle of the social hierarchy use pragmatic noise particularly often.

Keywords: characterisation, corpus-based methods, pragmatic noise, Shakespeare, social groups

1. Introduction

Compared with the voluminous literary critical literature, linguistic research on Shakespeare's language is somewhat lacking; compared with the voluminous linguistics literature on items that comprise the main or matrix clause, research on interjections is distinctly lacking. This chapter makes a contribution to both areas. Our study focusses on "pragmatic noise". Overlapping to an extent with interjections, pragmatic noise comprises the semi-natural noises – *ah*, *oh*, *ha*, *mhm*, *ugh* – that people make to express angst, anger, pain, surprise, pity, amusement, encouragement, listenership, and so on. Pragmatic noise is intimately connected with spoken interaction, and thus with the work of Merja Kytö, who at multiple points in her career has led research in historical speech-related phenomena, as this volume testifies. Merja is also one of the pioneers of historical corpus linguistics, and this chapter will adopt corpus-based methods to interrogate historical data. Furthermore, the notion at the heart of this chapter, pragmatic noise, was coined by Culpeper and Kytö (2010) in their work on Early Modern English.

In this chapter, we investigate the frequency, distribution and functions of pragmatic noise across Shakespeare's plays, and especially across particular social groups – those constituted by sex and social status. Do particular types of pragmatic noise cluster in particular plays? Does a play being a tragedy, comedy, or history have an influence on pragmatic noise item usage? Do female characters use more and/or different pragmatic noise items compared with male? Is the status of the character reflected in the use of pragmatic noise? These are the key research questions we will address. Our study utilises the resources made available by the AHRC-funded Encyclopedia of Shakespeare's Language project, based at Lancaster University. Not only do we have access to a fully searchable, regularised version of Shakespeare's plays, but we have access to a version that has been annotated for various social categories.

The following section expands a little more on pragmatic noise, and then Section 3 describes the Shakespeare play data, and elaborates on how we extracted pragmatic noise items. Sections 4 and 5 report and then discuss the pragmatic noise in the plays and social groups. Finally, we offer brief concluding remarks.

2. Pragmatic noise

Culpeper and Kytö (2010) devote four chapters of their book on Early Modern English dialogues to pragmatic noise, the first of which introduces the notion (see in particular Section 9.2). Pragmatic noise concerns material that lies outside the main syntax and is pregnant with pragmatic meaning. Unlike many discourse markers or pragmatic markers discussed in the literature, it concerns items that do not have homonyms in other word classes (e.g. *well* can act as a discourse marker in addition to, for example, an adjective), and are almost always monosyllabic and sometimes phonologically unusual (consider MHM) (the use of small caps throughout this chapter signals the inclusion of spelling variants). They are formed of semi-natural or instinctive noises, including not only single types like AH, HA, OH, HUM, but also reduplicative forms such as HA HA.

Pragmatic noise overlaps with two other notions in the linguistics literature. It overlaps to a great extent with what have been called primary interjections (e.g. Ameka 1992), but there are a number of differences. Unlike the label 'interjections', the label 'pragmatic noise' emphasises their pragmatic importance. Interjections are traditionally thought of as performing expressive or emotive functions. More

recently, scholars have recognized that they perform a wider range of functions. This is particularly true of pragmatic noise. For example, HEY and OI typically have directive functions, whilst MHM is typically phatic. Scholars have also noted that the items that constitute pragmatic noise perform discursual functions. Person (2009), studying OH in *Romeo and Juliet*, notes the range of contexts it appears in and the nuances of function it has, including discourse functions, such as marking a change of addressee (2009: 88) or prefacing requests (2009: 97). Jucker (2002), also studying OH, observes a general shift from an exclamatory (i.e. emotive) function to a discourse function. Furthermore, pragmatic noise includes types that are not classified as interjections at all, examples being laughter, pause-fillers, hesitation markers and listenership devices (e.g. MHM).

Pragmatic noise also overlaps with what Biber et al. (1999) refer to as 'inserts', but again there are differences. Pragmatic noise overlaps with central members of inserts, present-day examples including: OH, OI, OOPS, AH, HA, AHA, UH, UM, EH, ERM, MHM, TÛT, WHOA, and WHOOPS (all examples drawn from Biber et al. 1999: 1082–98). However, as Culpeper and Kytö (2010: 203) point out, unlike inserts, pragmatic noise sometimes has unusual phonetic or phonological characteristics, as noted at the beginning of this section. It is this that helps underpin the term "noise" in their label pragmatic noise. Furthermore, Culpeper and Kytö (2010: 199) claim that pragmatic noise items "have less arbitrary meanings compared with most words (they are sound symbolic to a degree)". They are relatively natural noises, evolved as spontaneous reactions to particular cognitive states (see Culpeper & Kytö 2010: Chapter 12, for a detailed elaboration of that evolution). No such claim is made of inserts. In short, pragmatic noise comprises the semi-natural noises that have evolved to express a range of pragmatic and discursual functions.

Turning to their presence in written texts in particular, Culpeper and Kytö (2010), examining speech-related late early modern texts, demonstrated how particular sets of pragmatic noise types tend to have particular functions in and were distributed across genres in particular ways. Five types were common to all their speech-related genres: O, OH, ALAS, AH and FIE (Culpeper & Kytö 2010: 268–270). However, there was variation in the density of occurrence across the genres, and also in the rank order of the particular types of pragmatic noise. Play-texts are particularly dense in pragmatic noise, which occurred in their data with a density of 5.5 per thousand words, compared with 1.7 per thousand words in the next most densely populated genre, prose fiction (Culpeper & Kytö 2010: 269). The rank order for particular pragmatic

noise types reported for plays is: O, HA, OH, FIE, AH, HE, ALAS, AY, PSHAW and TUSH (Culpeper & Kytö 2010: 269). This differed for other genres. For example, ALAS occurs in seventh position in play-texts but second position in prose fiction (Culpeper & Kytö 2010: 269). What was not considered, however, are potential differences amongst individual play-texts, and this is one of the goals of this chapter.

Pragmatic noise items have much to do with personal affect, a term which in linguistics has been used to encompass people's feelings, emotions, moods and attitudes, as well as personality (Caffi & Janney 1994: 328). Taavitsainen (1999) examined the role of features of personal affect – features which she termed 'surge features' – in literary characterisation, specifically in *The Canterbury Tales*, and these features include pragmatic noise. She states that personal affect is "a component of participant relations and finds outlets in various forms; thus it gives us a picture of the person's behavioural patterns and mental characteristics" (1999: 219–20). Examples include OH expressing surprise or FIE expressing disgust. What researchers have not done, however, is to consider, as we will do in this chapter, pragmatic noise (or overlapping categories such as surge features) across groups of characters comprising entire social groups.

3. Data and method

3.1 The Enhanced Shakespearean Corpus

The largest single body of Shakespeare's works and the earliest publication of a large group of his works is that constituted by the First Folio (1623). This was the obvious choice for the Encyclopedia of Shakespeare's Language project to have as its core data. Needless to say, scholars have recognised the presence of other hands in plays listed in the First Folio; collaborative works were common at the time. To the First Folio, we added two further plays: *Pericles* (Quarto 1) and *The Two Noble Kinsmen* (Quarto 1), believed to be collaborations with George Wilkins and John Fletcher, respectively. The resultant 38 plays, totalling 1,038,509 words, represent what is generally thought of as Shakespeare's canon.

The *Enhanced Shakespearean Corpus* (ESC) is termed 'enhanced' because of its tagging/annotation. Original spelling texts were kindly supplied by Internet Shakespeare Editions. Pragmatic noise is not immune to spelling variation. Without regularisation, a search on, for example, *alas*, would not retrieve instances spelt *alasse*. Spelling was

regularised with the program *Variant Detector* (VARD), developed by scholars at Lancaster University over more than 15 years, and most significantly by Alistair Baron (see <http://ucrel.lancs.ac.uk/ward/about>). This program regularises spelling by matching variants to “normalised” equivalents using a search and replace script, as well as contextual information, to tackle ambiguities, and an additional lexicon to treat word forms that are specific to or have undergone semantic change since the early modern period. The program does not delete the original spelling, but places it in a specific XML element, thereby making it easily available for inspection. Because the project demanded a high level of accuracy, we did not run the program in fully automatic (whole-text) mode. Instead, the program’s manual (word-by-word) mode can on most occasions suggest regularisation options in order of likelihood, from which the human operator approves a selection. We made no attempt to “correct” the spelling, with very rare exceptions made for obvious printer errors, such as *aud* for *and*.

The ESC is tagged for parts of speech. As we explain in section 3.2, the grammatical category of interjections played an important role in our method for retrieving instances of pragmatic noise. Part of speech tagging was partly achieved through the program CLAWS (the Constituent Likelihood Automatic Word-tagging System: see <http://ucrel.lancs.ac.uk/claws>). In a nutshell, CLAWS works on the basis of (1) a lexicon, including words (or multi-word units) and suffixes and their possible parts of speech, and (2) a matrix containing sequencing probabilities (e.g. the likelihood that the word following an adjective will be a noun), which is applied to each sentence to disambiguate words which could potentially be several parts of speech. However, Early Modern English presents a range of problems, aside from spelling variation. These include vocabulary change over time: some words have disappeared from English over the last 400 years (e.g. *iwis*, meaning ‘certainly’ or ‘assuredly’) and are thus not in the tagger’s lexicon, whilst others still exist but behave differently in grammatical terms (e.g. the word *fee* could equally well be a verb as a noun). In addition, CLAWS overlooks many grammatical features of Early Modern English – for instance, the existence of *thou* and *thee* as forms distinct from *you*, rather than as marginal phenomena as they are today; or the regular use of an inflected second person for all verbs. Our solution to both these problems was to make adjustments to CLAWS, and also to manually check all texts in the ESC.

The corpus itself has also been annotated for speakers’ sex and social status, and other speaker characteristics. This annotation scheme was only applied to characters whose talk makes up at least five per cent

of the total word count of the play in which they appear (this excludes very minor characters – messengers, for example – whose nature can be difficult to determine). Categorising characters as male or female is relatively straightforward, though it was necessary to develop separate categories for characters with an assumed identity (e.g. a female character playing a male character). The social status categorisation scheme drew upon the approach developed by Archer and Culpeper (2003), where further detail can be found. The social hierarchy is as follows: monarchy > nobility > gentry > professional > other middling groups > ordinary commoners > lowest groups. The project was sensitive to the fact that it was working with fictional data. Hence, for example, we added a “supernatural beings” category accounting for more than 40 ghosts, gods, fairies, etc. in the 38 plays. Partly for reasons of space, in this chapter, we will confine ourselves to groups on the social hierarchy.

3.2 A method for pragmatic noise extraction

Pragmatic noise instances had been treated by the tagger as interjections, the relevant label being UH. A search for ‘_UH’ in the ESC produced 92 types with a combined total of 8,179 tokens. Not all of these interjections, of course, counted as pragmatic noise, and the statuses of some were difficult to determine. Consequently, the following were removed from this list:

- 1) Affirmative/negative items: NO, NAY, YES, AYE, YEA.
- 2) Morphologically complex items and/or items with homonyms in other word classes: e.g. GRAMERCY, FAREWELL, WELADAY. ALAS and ALACK are borderline cases, as their second elements have homonyms in other word classes, but their first elements, A or AH, are more clearly pragmatic noise. For this reason, ALAS and ALACK remain in this study.
- 3) Highly restricted items. For example, NONINO and NONNY occur only in song and in phrases following on from HEY.
- 4) Items with fewer than three instances. We instituted this to ensure that we had a sufficient number of occurrences of each item to interpret its function.
- 5) Items lacking clarity. There were two examples here: SESSA, which occurs three times, each with a different original spelling, but no clear function or meaning, and UGH, all three instances of which appear consecutively in a single line.

A total of 4,524 tokens were eliminated, or 55.19% of all tokens tagged as interjections. The remaining 21 types in order of frequency are as follows: O, OH, ALAS, HO, HA, FIE, AH, ALACK, LO, TUT, TUSH, LA, HOLLA, HUM, FOH, SOLA, HEM, HEY, HUSH, PISH and MUM. These amounted to 3,649 tokens in total.

4. Distribution of pragmatic noise across Shakespeare's plays

4.1 Overview of distribution by play

Table 1 displays our frequency results across Shakespeare's plays. Play genre refers to whether the play is a tragedy, comedy or history. For these classifications, which are notoriously controversial, we follow the classification given in the First Folio, with the exception of *Cymbeline*, which we reclassify as a comedy. The plays vary considerably in length. This is an important fact for interpreting the other figures in the table. The middle columns of the table contain the number of different pragmatic noise types and the number of pragmatic noise tokens. Finally, the table displays the normalized frequencies of pragmatic noise (per thousand words), and the rows of the table are ordered according to these results.

Table 1. The number of pragmatic noise (PN) types, the number of tokens and the relative frequencies for each play in the *Enhanced Shakespearean Corpus*, rank-ordered by relative frequency

Play	Play genre	Total words	No. of different PN types	No. of PN tokens	Normalized frequency of PN tokens (per 1,000 words)
Romeo and Juliet	T	29,556	11	218	7.4
Othello	T	32,668	14	240	7.3
Hamlet	T	34,761	15	171	4.9
Troilus and Cressida	T	32,060	14	157	4.9
King Lear	T	29,188	14	137	4.7
Antony and Cleopatra	T	30,277	11	141	4.7
Titus Andronicus	T	24,584	10	113	4.6
Julius Caesar	T	24,037	7	105	4.4
Twelfth Night	C	24,033	11	100	4.2

The final version of: Culpeper, Jonathan and Samuel Oliver (2020) Pragmatic noise in Shakespeare's plays. In: Ewa Jonsson and Tove Larsson (eds.) *Voices past and present - Studies of involved, speech-related and spoken texts*. Amsterdam: John Benjamins. It may contain minor errors and infelicities.

The Tempest	C	20,482	10	81	4.0
Love's Labour's Lost	C	25,867	12	99	3.8
A Midsummer Night's Dream	C	20,126	8	77	3.8
Measure for Measure	C	26,380	8	98	3.7
The Two Noble Kinsmen	C	29,393	10	108	3.7
The Merry Wives of Windsor	C	26,663	11	94	3.5
King John	H	24,768	8	87	3.5
Richard III	H	35,401	10	123	3.5
As You Like It	C	25,954	12	90	3.5
The Taming of the Shrew	C	25,344	11	84	3.3
Cymbeline	C	33,819	11	111	3.3
Two Gentlemen of Verona	C	21,212	9	65	3.1
Much Ado about Nothing	C	25,203	10	77	3.1
Timon of Athens	T	22,510	10	68	3.0
Henry VI, Part 3	H	29,779	8	88	3.0
Henry IV, Part 2	H	31,977	11	94	2.9
Pericles	C	22,073	8	63	2.9
The Winter's Tale	C	31,026	11	85	2.7
Henry VI, Part 2	H	30,763	7	80	2.6
The Merchant of Venice	C	25,065	7	65	2.6
Richard II	H	26,495	9	66	2.5
The Comedy of Errors	C	17,587	7	43	2.4
Henry IV, Part 1	H	29,724	12	71	2.4
Henry VI, Part 1	H	26,083	9	62	2.4
Coriolanus	T	33,722	11	78	2.3
Henry VIII	H	30,022	9	67	2.2
Henry V	H	31,366	8	60	1.9
Macbeth	T	21,118	9	40	1.9
All's Well that Ends Well	C	27,423	6	43	1.6

Table 2 displays the actual pragmatic noise items (types and tokens) for each play, along with their raw frequencies of occurrence in that play.

The plays are listed according to their overall normalized frequency of pragmatic noise tokens; in other words, the order of plays here matches that of Table 1.

Table 2. The pragmatic noise (PN) items in each play in the *Enhanced Shakespearean Corpus* (the plays are rank-ordered by relative frequency of PN tokens per 1,000 words))

Play	Play genre	Pragmatic noise items and their raw frequencies
Romeo and Juliet	T	O (135), OH (16), AH (15), HO (13), ALACK (11), ALAS (8), FIE (6), HA (5), TUT (4), LO (3), TUSH (2)
Othello	T	OH (117), O (34), ALAS (28), HO (26), HA (17), FIE (8), LO (2), PISH (2), AH (1), ALACK (1), FOH (1), HEM (1), HOLLA (1), HUM (1)
Hamlet	T	OH (79), O (39), HO (12), ALAS (9), FIE (7), HA (7), ALACK (3), HOLLA (3), LO (3), AH (2), HEY (2), TUSH (2), FOH (1), HUM (1), LA (1)
Troilus and Cressida	T	O (70), HA (20), OH (17), HO (11), ALAS (9), FIE (7), LO (5), AH (4), FOH (4), HOLLA (4), HUM (2), LA (2), HEM (1), HEY (1)
King Lear	T	O (63), OH (17), ALACK (13), HO (13), HA (12), FIE (7), ALAS (2), HOLLA (2), HUM (2), MUM (2), AH (1), FOH (1), HEY (1), HUSH (1)
Antony and Cleopatra	T	OH (83), HO (15), O (13), AH (10), FIE (5), HA (5), ALACK (4), HUSH (2), LO (2), ALAS (1), LA (1)
Titus Andronicus	T	OH (39), O (30), AH (13), ALAS (10), FIE (7), HA (5), LO (4), TUT (3), HO (1), HOLLA (1)
Julius Caesar	T	O (69), HO (18), ALAS (9), HA (5), LO (2), OH (1), TUT (1)
Twelfth Night	C	O (42), ALAS (14), HO (12), OH (9), FIE (6), HA (6), HEY (6), AH (2), LA (1), LO (1), TUT (1)
The Tempest	C	O (42), HA (9), OH (9), HO (5), LO (5), ALACK (3), ALAS (3), HEY (3), FIE (1), HUSH (1)
Love's Labour's Lost	C	O (75), AH (5), ALACK (5), ALAS (2), HA (2), LO (2), OH (2), SOLA (2), FIE (1), HO (1), HOLLA (1), TUSH (1)
A Midsummer Night's Dream	C	O (55), OH (6), HO (5), ALACK (40), FIE (3), AH (2), HEY (1), LO (1)
Measure for Measure	C	OH (56), ALAS (9), FIE (9), HA (7), HO (7), O (7), ALACK (2), FOH (1)
The Two Noble Kinsmen	C	O (53), ALAS (17), OH (14), HO (5), LO (5), HA (4), HEY (4), FIE (3), ALACK (2), AH (1)
The Merry Wives of Windsor	C	O (25), FIE (14), ALAS (13), HA (13), OH (9), HO (7), LA (7), AH (2), TUT (2), FOH (1), HUM (1)
King John	H	O (46), OH (26), LO (4), AH (3), ALACK (2), ALAS (2), HA (2), HO (2)
Richard III	H	O (67), AH (17), OH (10), ALAS (9), HA (5), LO (5), TUT (5), ALACK (2), FIE (2), HO (1)
As You Like It	C	O (38), HEY (14), OH (12), HO (10), ALAS (7), HOLLA (3), AH (1), ALACK (1), FIE (1), HA (1), HEM (1), LO (1)
The Taming of the Shrew	C	OH (47), FIE (13), HO (50), O (4), TUT (4), TUSH (3), ALAS (2), HA (2), AH (1), LO (1)

Cymbeline	C	OH (60), O (22), ALACK (7), HO (7), ALAS (6), AH (2), FIE (2), HA (2), HUM (1), HUSH (1), LO (1)
Two Gentlemen of Verona	C	OH (33), O (13), ALAS (9), FIE (5), AH (1), HA (1), HO (1), LO (1), TUT (1)
Much Ado about Nothing	C	O (46), HA (9), ALAS (4), TUSH (4), AH (3), FIE (3), HEY (3), HO (3), HEM (1), MUM (1)
Timon of Athens	T	O (24), OH (13), FIE (60), HO (6), HA (5), ALAS (4), LA (4), HUM (3), AH (2), ALACK (1)
Henry VI, Part 3	H	AH (29), OH (24) O (19), ALAS (9), LO (3) TUT (2) HA (1), TUSH (1)
Henry IV, Part 2	H	O (42), HA (12), OH (12), ALAS (7), AH (5), FIE (5), HO (3), LO (3), ALACK (2), HEM (2), LA (1)
Pericles	C	O (37), OH (9), ALAS (4), HA (4), HO (4), ALACK (2), FIE (2), HUM (1)
The Winter's Tale	C	O (31), OH (27), ALAS (7), FIE (4), HA (3), HEY (3), HO (3), ALACK (2), HOLLA (2), LO (2), LA (1)
Henry VI, Part 2	H	O (26), AH (19), OH (18), ALAS (9), FIE (4), TUT (3), LO (1)
The Merchant of Venice	C	O (33), SOLA (8), FIE (6), HA (6), HO (5), ALAS (4), ALACK (3)
Richard II	H	OH (35), ALACK (6), O (6), AH (5), ALAS (5), HA (3), HO (2), LO (2), TUT (2)
The Comedy of Errors	C	OH (25), FIE (5), ALAS (4), O (4), AH (3), HO (1), LO (1)
Henry IV, Part 1	H	O (42), TUT (6), OH (5), FIE (3), HA (3), HEY (3), HO (3), AH (2), ALAS (1), HEM (1), HUM (1), TUSH (1)
Henry VI, Part 1	H	O (21), OH (18), FIE (4), TUSH (4), AH (3), ALAS (3), HA (3), LO (3), TUT (3)
Coriolanus	T	OH (40), O (12), FIE (7), HO (6), HA (4), ALACK (2), LA (2), TUSH (2), AH (1), ALAS (1), LO (1)
Henry VIII	H	O (29), HA (12), ALAS (10), OH (5), AH (3), FIE (3), LO (3), HO (1), TUSH (1)
Henry V	H	O (47), ALAS (3), HA (3), OH (2), PISH (2), AH (1), FIE (1), TUT (1)
Macbeth	T	O (17), OH (7), ALAS (4), FIE (3), HO (3), HA (2), LO (2), ALACK (1), HUM (1)
All's Well that Ends Well	C	O (31), 6 (OH), ALAS (2), HO (2), AH (1), FOH (1)

4.2 Discussion of distribution by play

A striking feature of Table 1 is the dramatic difference in the normalized frequencies of pragmatic noise tokens. They vary from *Romeo and Juliet* with 7.4 to *All's Well that Ends Well* with 1.6 tokens per thousand words. It is noteworthy that the top eight plays, rank-ordered by normalized frequency, are all tragedies; the histories inhabit the lower echelons, mixed in with some comedies, which take up some of the centre ground. There is no easy way of interpreting this, because on the one hand the notion of a tragedy is not a particularly coherent and consistent one, and on the other hand pragmatic noise covers such a range of functions. Nevertheless, one might speculate that plays, such

as tragedies, that feature high emotions are somewhat more likely to contain more pragmatic noise. We write “somewhat” because much seems to depend on the particular kind of tragedy.

The tragedy *Macbeth* is conspicuous by its positioning almost at the bottom of Table 1. The spread in relative frequencies across the plays is not quite even. In particular, note that *Romeo and Juliet* and *Othello* both score around 7.4 per thousand words, but then there is a drop to around 5 with *Hamlet*. Why are these two plays particularly dense in pragmatic noise? Both plays are love tragedies. *Antony and Cleopatra*, another notable love tragedy, ranks not far behind in sixth place, and *Troilus and Cressida*, in fourth place, has a claim to being a love tragedy. Perhaps this particular type of tragedy, with its extreme emotional ups and downs, results in high density of occurrence of pragmatic noise. Let us probe exactly what lies underneath the high numbers for these plays.

Two factors seem to be at play. Both *Romeo and Juliet* and *Othello* contain a high number of different pragmatic noise items. *Othello* has 14 types, a number that is only exceeded by *Hamlet* with 15 (and *Hamlet* is in third position), and *Romeo and Juliet* has 11. The other factor, as can be seen from Table 2, is that both plays have an exceptionally high number of tokens of their most frequent pragmatic noise types. In *Romeo and Juliet*, O occurs 135 times, which is almost twice the frequency of its next most frequent occurrence (in *Troilus and Cressida*, where it has 70 occurrences). In *Othello*, OH occurs 117 times, substantially more than its next most frequent occurrence, in *Hamlet*, with 79 instances. It should be noted that, aside from the fact that O has a particular tendency to accompany terms of address as part of a vocative, it overlaps considerably in functional terms with OH (see Culpeper & Kytö 2010: section 11.5). Also, the choice between O and OH was probably susceptible to choices made by compositors, if they needed to save space in a printed line.

Let us examine the use of O and OH in *Romeo and Juliet* and *Othello* a little more closely. In *Romeo and Juliet*, the characters Juliet, Romeo and the Nurse each use approximately three times as many pragmatic noise tokens as any other character. Of those, the Nurse uses them most densely (10.2 times per thousand words, compared with 7.2 for Juliet and 6.3 for Romeo). One of the Nurse's main functions in the play is to act as an emotional mirror for various happenings, particularly in relation to Juliet. In addition, she seems to have been constructed as having a general propensity for pragmatic noise, and O in particular. She uses it for a wide range of functions, including to express the pain she feels (pragmatic noise instances are emboldened):

- (1) My back a' the other side: o my back, my back:
(*Romeo and Juliet* 2.5)

To attract attention, and reinforce a request:

- (2) O holy Friar, O tell me holy Friar, Where 's my Lady 's
Lord?
(*Romeo and Juliet* 3.3)

To swoon (or pretend to) at the thought of Paris:

- (3) O he's a Lovely Gentleman: Romeo's a dishclout to him
(*Romeo and Juliet* 3.5)

However, above all, the Nurse uses it for lamentation:

- (4) O woe, O woeful, woeful, woeful day,
Most lamentable day, most woeful day,
That ever, ever, I did yet behold.
O day, O day, O day, O hateful day,
Never was seen so black a day as this:
O woeful day, O woeful day.
(*Romeo and Juliet* 4.5)

Pragmatic noise was used in the expression of particular rhetorical figures. Here, the relevant figure is *lamentatio*. This figure particularly accounts for ALAS, but also AH and ALACK in plays. As in example (4), *lamentatio* is the conventional reaction in plays to death. (See also Taavitsainen's discussion of ALAS in this volume, which echoes some of our points about emotion and tragedy).

Regarding the particular pragmatic noise items displayed in Table 2, note that quite a few of the rank-ordered pragmatic noise types for play-texts reported in Culpeper and Kytö (2010: 269) – O, HA, OH, FIE, AH, HE, ALAS, AY, PSHAW and TUSH – are in evidence here, with a few exceptions. AY is not included because of our method (see section 3.2). HE has a long history in English, but was a relatively minor form until later in the seventeenth century, when it became regularly used to represent laughter. Similarly, PSHAW only took off in the 1670s. The remaining types, O, HA, OH, FIE, AH, ALAS and TUSH, appear regularly across the plays in Table 2 – a reflection perhaps that plays are a fairly cohesive genre. However, there are some interesting differences in the raw frequency rank orders.

In the plays *The Merry Wives of Windsor*, *The Taming of the Shrew* and *The Comedy of Errors*, FIE makes it to second position. FIE is regularly used to cast shame or pour scorn on something or someone, or to express anger or exasperation:

- (5) **Fie** upon thee! art not ashamed?
(*Much Ado About Nothing* 3.4)

These are indeed plays which are notable for such emotions, *The Taming of the Shrew* being the best exemplar. Another interesting example concerns the positioning of ALAS. It occurs in second position in *Twelfth Night*, *Measure for Measure*, *The Two Noble Kinsmen* and *Henry V*. ALAS, as already mentioned above, is key to the expression of lamentation, but its use extends beyond that. It is used to express regret for something, and sometimes sympathy or empathy with others:

- (6) **Alas**, 'tis a sore life they have I' th' other place – such
burning, frying, boiling, hissing, howling, chatt'ring,
cursing
(*The Two Noble Kinsmen* 4.3)

Measure for Measure and *The Two Noble Kinsmen* are both classified as comedies in the First Folio. However, for over a century, this classification has often been viewed as problematic, and they have been described as tragicomedies (for possibly the earliest comment on problematic plays, see Boas 1910). The threat of death via execution is a key theme in both. *Henry V* is certainly laced with dark moments and death. The odd one out then seems to be *Twelfth Night*. But here too, despite being a comedy, there are dark moments. "Alas the day!" says Antonio, in act two scene one, when reflecting on the death by drowning of his sister. Eight of fourteen instances of ALAS are spoken by Olivia and Feste, four each. Olivia most often uses it to express regret for or sympathy with another's situation (e.g. "Poor Malvolio", 5.1). Feste, being the Fool or Clown, might lead the modern reader to think that he is all levity. In fact, he provides insightful and sometimes somewhat melancholic reflections on himself and the behaviours of others. Example (7) is from the last song he sings:

- (7) But when I came **alas** to wive, with hey ho, etc.
By swaggering could I never thrive, for the rain, etc.
(*Twelfth Night* 5.1)

5. Distribution of pragmatic noise across Shakespeare's social groups of characters

5.1 Overview of distribution by social groups

Table 3 displays the frequencies of pragmatic noise across the sex of the characters in Shakespeare's plays, and Table 4 displays the rank order of pragmatic noise items for each sex, along with their raw frequencies.

Table 3. The distribution of pragmatic noise (PN) in male and female characters in the *Enhanced Shakespearean Corpus*

Sex	No. of characters	No. of words	No. of PNs	Average (mean) no. of PNs per character and standard deviation in round brackets	Normalized frequency of PN tokens (per 1,000 words)
Male	1,235	811,531	2,620	2.1 (7.8)	3.2
Female	166	171,132	937	5.6 (9.2)	5.5

Table 4. The rank-ordered pragmatic noise (PN) items by sex of characters in the *Enhanced Shakespearean Corpus*

Sex	Pragmatic noise types and their raw frequencies
Male	O (999), OH (639), HA (188), HO (181), ALAS (130), FIE (115), AH (105), LO (54), ALACK (45), TUT (36), TUSH (20), HOLLA (16), HUM (13), LA (11), SOLA (10), FOH (9), HEM (5), HUSH (4), PISH (4), MUM (2)
Female	O (371), OH (249), ALAS (109), AH (53), FIE (43), HO (31), ALACK (29), LO (15), HA (12), LA (9), HEY (6), HOLLA (3), HEM (2), TUT (2), HUSH (1), MUM (1), TUSH (1)

Table 5 displays the frequencies of pragmatic noise across the social status of the characters in Shakespeare's plays. Figures 1 and 2 reproduce some of the same information in graphic form. Figure 1 displays the average (mean) number of pragmatic noise types per character of a particular status, and Figure 2 displays the normalized frequency of pragmatic noise tokens (per thousand words). Finally, Table 6 displays the rank order of pragmatic noise items for each social status, along with their raw frequencies.

Table 5. The distribution of pragmatic noise (PN) across social status of characters in the *Enhanced Shakespearean Corpus*

Status	No. of characters of that status	No. of words	No. of PNs	Average (mean) no. of PNs per character of that status and standard deviation in round brackets	Normalized frequency of PN tokens (per 1,000 words)
Monarchy	78	164,427	511	6.6 (9.1)	3.1
Nobility	379	404,284	1,547	4.1 (9.6)	3.8
Gentry	263	199,731	786	3.0 (9.2)	3.9
Professional	102	42,626	184	1.8 (4.8)	4.3
Other middling	71	34,557	83	1.2 (2.7)	2.4
Ordinary commoners	90	35,497	144	1.6 (4.0)	4.1
Lowest groups	324	50,946	168	0.5 (2.8)	3.3

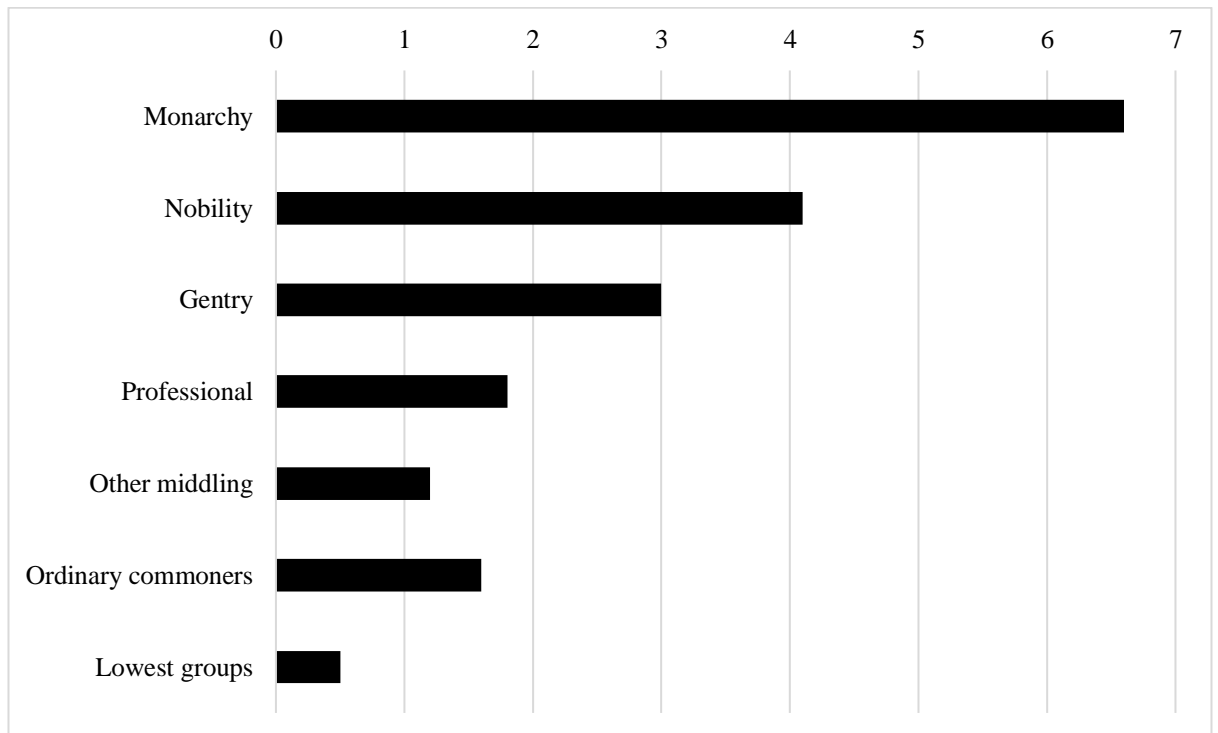


Figure 1. Average (mean) number of pragmatic noise (PN) tokens across social status of characters in the *Enhanced Shakespearean Corpus*

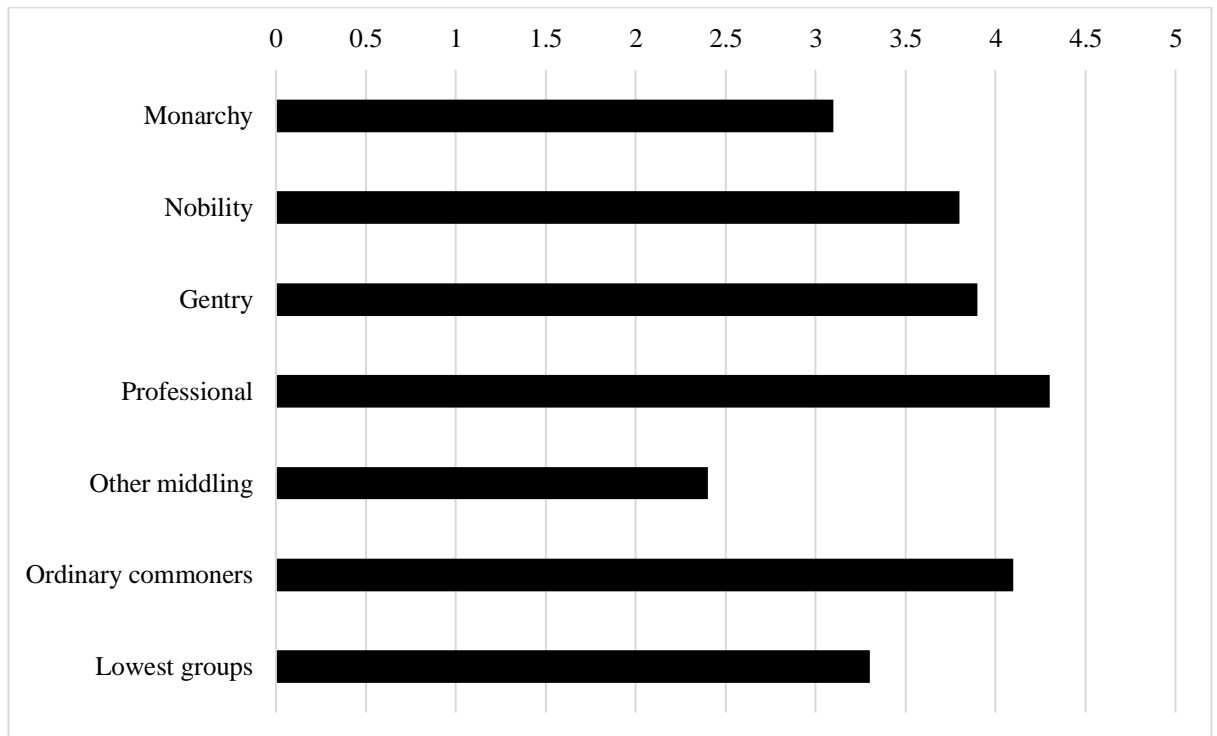


Figure 2. Normalized frequencies of pragmatic noise (PN) tokens (per 1,000 words) across social status of characters in the *Enhanced Shakespearean Corpus*

Table 6. Rank-ordered pragmatic noise (PN) items by social status of characters in the *Enhanced Shakespearean Corpus*

Status	Pragmatic noise types and their raw frequencies
Monarchy	O (183), OH (159), AH (43), HA (34), ALAS (26), HO (23), FIE (20), ALACK (14), LO (6), HEY (2), HUM (1), HUSH (1), LA (1)
Nobility	O (673), OH (360), ALAS (102), AH (73), HA (72), HO (70), FIE (57), ALACK (40), LO (31), TUT (18), TUSH (13), HEY (9), HUM (8), LA (8), HOLLA (5), HEM (4), HUSH (2), FOH (1), PISH (1)
Gentry	O (241), OH (240), HA (57), HO (56), ALAS (48), FIE (42), AH (25), LO (16), TUT (14), ALACK (12), TUSH (8), LA (7), FOH (6), HOLLA (5), HEY (4), HEM (2), HUSH (1), MUM (1), PISH (1)
Professional	O (68), OH (25), FIE (24), HO (18), ALAS (14), HA (14), HEY (9), LO (3), AH (2), ALACK (2), HUM (2), FOH (1), HOLLA (1), LA (1)
Other middling	o (29), OH (15), ALAS (13), FIE (6), HO (6), AH (5), LA (3), HA (2), ALACK (1), HEY (1), HOLLA (1), TUT (1)
Ordinary commoners	O (57), OH (20), HO (17), ALAS (13), SOLA (8), HA (7), LO (5), AH (4), HOLLA (4), FIE (3), HEY (3), HUM (2), PISH (1)

<i>Lowest groups</i>	O (56), OH (44), ALAS (18), HO (14), AH (7), HA (7), LO (6), TUT (5), ALACK (4), HEY (2), HOLLA (2), FIE (1), HEM (1), PISH (1)
----------------------	---

5.2 Discussion of distribution by social groups

Male characters speak many more words than female characters in the plays. Consequently, weighted or normalized frequencies are crucial in discussing frequencies of pragmatic noise. In Table 3, the average number of pragmatic noise tokens per female character is 5.6, whereas it is 2.1 for male. This tendency for pragmatic noise to be more frequently uttered by female characters is also reflected in the overall frequencies relative to the number of words spoken. Female characters speak 5.5 pragmatic noise tokens per thousand words, compared with 3.2 for male. We should of course remember that female characters here are constructed by a male author (or male authors, for plays which are thought to involve collaborations). These representations in Shakespeare's plays may partly reflect a wider male stereotype of women as being more emotional. But they could also partly reflect the frequent role ascribed to female characters in the plays – Desdemona in *Othello* being a good example – namely, to demonstrate the emotional consequences of what the male characters are doing.

Certainly, at the time women were associated with two specific areas of emotion, grief and fear, as these quotations from texts of that period make clear: “Yet he doth not with womanly weeping bewaile his departure” (Rudolph Gwalther & John Bridges, 1572, *An hundred, threescore and fiftene homelyes or sermons* [...]); “but somewhat moved with her too womanly tymerousnes and fear” (Thomas Bentley, 1582, *The sixt lampe of virginities conteining a mirrour for maidens and matrons* [...]). The rank order of the raw frequencies of pragmatic noise tokens displayed in Table 4 provides evidence for the first of these areas of emotion. ALAS appears in third position for female characters, but in fifth for male. Furthermore, other pragmatic noise items point in the same direction. Both ALACK, in seventh position for female characters and ninth position for male, and AH, in fourth position for female characters and seventh position for male, overlap to an extent with the core grief and sorrow emotional functions of ALAS, though they encompass a somewhat wider array of states relating to emotional distress. Conversely, it is worth noting that pragmatic noise types with some kind of directive function play a larger role in male character talk compared with female, reflecting the key role male characters are given in the direction of events, and also the patriarchal society of the time. TUSH, for example, is in 11th position with 20 occurrences, whereas it is

at the end of the rank-ordered list for female characters with a single instance. That single instance is spoken by Queen Margaret in *Henry VI, Part 3*:

- (8) *Suffolk*: Sweet Madam, give me hearing in a cause.
Margaret: **Tush**, women have been captivate ere now.
(*Henry VI, Part 3, 5.3*)

This is significant because Margaret is characterized, not least by other characters, as an unnatural woman on account of her masculine behaviours (e.g. leading an army). HOLLA and HUSH also have a similar pattern to TUSH.

Turning to pragmatic noise across characters of different social status, Table 5 and Figure 1 tell a story made complex by the fact that there are widely varying numbers of characters at each status level who speak widely varying numbers of words. Focusing on the average number of pragmatic noise occurrences per character, we see a relationship which almost exactly follows social status level such that the higher the status the more items used. The rank order is: monarchy > nobility > gentry > professional > ordinary commoners > other middling > lowest groups. The most likely reason for this is that the characters of higher social status have larger parts (as is apparent from the number of words spoken by the number of characters in Table 5), and thus have more opportunities to use pragmatic noise. More illuminating, perhaps, are the frequencies of pragmatic noise tokens relative to the total number of words spoken by the different character social groups, as displayed in Table 5 and Figure 2. Here, the rank order is professional > ordinary commoners > gentry > nobility > lowest groups > monarchy > other middling groups. The final group, other middling groups, is poorly represented – not much can be concluded from it. Broadly speaking, and as can be clearly seen in Figure 2, these results suggest that the middle groups in the social hierarchy are rather more densely populated by pragmatic noise than the groups on the extremities. It is difficult to explain the reasons for this without significant further study, including a consideration of dispersion, particularly as the differences between the groups were quite small. However, we might observe that it is the characters of middling groups that are often engaged in colloquial interactions, and acting as foils for the main characters (witness the Nurse, as discussed in the previous section).

Some clues to the functional characteristics of particular social status groups can be found in the rank-ordered lists in Table 6. Perhaps

the most striking feature is the position of FIE. It is dominant in the middle of the social hierarchy, especially the professional group, but also other middling and to some extent gentry. It is present in reasonable quantities in the nobility group and also monarchy, but barely exists at the bottom of the hierarchy in the ordinary commoners and lowest groups. The two plays in which it occurs most densely are *The Merry Wives of Windsor* and *The Taming of the Shrew*. For example, in the former, Mistress Page, who is of gentry status, pours scorn on Master Ford:

- (9) **Fy, fy**, M Ford, are you not asham'd?
(*The Merry Wives of Windsor* 3.3).

In the latter, Grumio, a professional groom, habitually scorns others:

- (10) **Fi**, **fi** on all tired Iades, on all mad Masters [...]
(*The Taming of the Shrew* 4.1)

Note that in both these cases we see the reduplicative form (we take a reduplicative form to involve repeated adjacent constituents which together form a single conventionalized unit). Whilst we have not systematically investigated this, it does seem to be the case that the groups in the middle of the hierarchy engage in reduplicative forms or repeated forms (as in the Nurse's speech; see example (4)) more than the groups at the extremities of the social hierarchy. This, obviously, could be a factor in their relatively high frequencies.

6. Conclusion

One overall finding of our work is that there seems to be more variation in frequency of pragmatic noise tokens than in pragmatic noise types. Love tragedies seem to attract particular densities of pragmatic noise, probably as a consequence of their extreme emotional ups and downs. In contrast, the pragmatic noise items O, HA, OH, FIE, AH, ALAS, and TUSH maintain a fairly steady presence across all of the plays. We noted how some pragmatic noise items are a part of rhetorical figures. ALAS, for example, is part of *lamentatio*. These figures are played out in particular contexts. Thus, death would trigger *lamentatio*, and hence the utterance of ALAS (and to some extent AH and ALACK). Plays that had these contexts would see an increase in density of pragmatic noise.

We found that female characters have a greater density of pragmatic noise tokens, perhaps partly as a consequence of their frequent function in the plays as an emotional mirror, but also perhaps as a reflection of the male stereotype of women as more emotional. More specifically, there is evidence that female characters are constructed to perform the emotions of grief and sorrow, through ALAS, ALACK and AH, whereas male characters are constructed as exercising power through the direction of events with items such as TUSH, HOLLA and HUSH. The distribution of pragmatic noise across social status groups proves complex. Whilst overall the higher a character is in terms of social status the more pragmatic noise they use, that could simply reflect the fact that higher status characters get to speak so much more than lower; in other words, they have more opportunities to use pragmatic noise. However, if we factor in the amount that each character speaks overall (i.e. we consider normalized frequencies), we are more likely to hear groups in the middle of the social hierarchy using pragmatic noise more often than at the extremes. This may be because those characters also tend to function as emotional foils for the main characters. These middle characters gravitate in particular to the form FIE. We also noted that they seem to tend towards reduplicative or repeated forms, which obviously would help account for their relatively high densities of use.

We have of course only scraped the surface of what is possible. Future studies would benefit from more probing of specific forms in their contexts. In addition, a more sophisticated statistical model, and also one that combines a number of variables and examines dispersion, would help shed light. Nevertheless, we hope that we have shown that studying pragmatic noise and how it varies across genre and groups of characters is a fascinating enterprise worthy of study.

Funding

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

References

- Ameka, F. 1992. Interjections: The universal yet neglected part of speech. *Journal of Pragmatics* 18(2-3): 101-118.
- Archer, D. & J. Culpeper. 2003. Sociopragmatic annotation: New directions and possibilities in historical corpus linguistics. In *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* [Łódź Studies in Language 8], A. Wilson, P. Rayson & T. McEnery (eds), 37-58. Frankfurt am Main, Berlin, Bern, etc.: Peter Lang.
- Bentley, T. 1582. *The sixt lampe of virginities conteyning a mirrour for maidens and matrons* [...]. At the three cranes in the vintree, by Thomas Dawson and Henry Denham, for the assignes of William Seres. London. (STC 1894).
- Biber, D., S. Johansson, G. N. Leech, S. Conrad & E. Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Boas, F. S. 1910. *Shakspere and his predecessors*. London: John Murray.
- Caffi, C. & R. W. Janney. 1994. Toward a pragmatics of emotive communication. *Journal of Pragmatics* 22(3-4): 325-373.
- Culpeper, Jonathan. 2019. Lead compiler of the *Enhanced Shakespearean Corpus*, part of the Encyclopedia of Shakespeare's Language Project (AHRC grant reference AH/N002415/1). Lancaster University, U.K.
- Culpeper, J. & M. Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.
- Gwalther, R. & J. Bridges. 1572. *An hundred, threescore and fiftene homelyes or sermons* [...]. By Henrie Denham, dwelling in Pater noster rowe, at the sign of the starre. London. (STC 25013).
- Jucker, A. H. 2002. Discourse markers in Early Modern English. In *Alternative histories of English*, R. Watts & P. Trudgill (eds), 210-230. London and New York: Routledge.
- Person, R. F. 2009. "Oh" in Shakespeare: A conversation analytic approach. *Journal of Historical Pragmatics* 10(1): 84-107.
- Taavitsainen, I. 1999. Personality and styles of affect in *The Canterbury tales*. In *Chaucer in perspective: Middle English essays in honour of Norman Blake*, Geoffrey Lester (ed.), 218-234. Sheffield: Sheffield Academic Press.