

## The Language of *Macbeth*

Jonathan Hope (Strathclyde University, Glasgow) and Michael Witmore (Folger Shakespeare Library, Washington, D.C.)

This is a reformatted version of the following published paper:

Jonathan Hope and Michael Witmore, 2014, 'The Language of *Macbeth*', in Ann Thompson (ed.), *Macbeth: The State of Play* (London: Bloomsbury), pp. 183-208

If you wish to cite this paper, please use the published version.

*Macbeth* has affected its critics perhaps more intensely than any other Shakespeare play, and its language in particular has been the focus of outrage, puzzled unease, and close examination. Modern responses attest to the play's 'continuous sense of menace' and 'horror', a feeling that the play itself, rather than individuated characters, speaks the lines, and a sense that what is communicated goes beyond the grammar or strict logic of the surface linguistic forms.<sup>1</sup> Perhaps the most resonant account of the play's mysterious effect is Thomas De Quincey's celebrated short note 'On The Knocking At The Gate In *Macbeth*' (1823), which begins with his childhood 'perplexity' at the strength of the feelings produced in him by the knocking at the gate after Duncan's murder:

---

<sup>1</sup> For 'menace' and 'horror', see George Walton Williams, "'Time for such a word": verbal echoing in *Macbeth*', in *Shakespeare and Language*, ed. Catherine M.S. Alexander (Cambridge, 2004), 240-50 (originally published in *Shakespeare Survey*, 47 (1994), 153-59), 210 and 211. For the play, rather than characters speaking, see Nicholas Brooke (ed.), *Macbeth* (Oxford, 1990), 12-3. For communication beyond surface meaning, see Stephen Booth, 'Close reading without reading' in *Shakespeare Reread: The Texts in New Contexts*, ed. Russ McDonald (Ithaca, 1984), 42-55; and Russ McDonald, *Shakespeare's Late Style* (Cambridge: 2006), 46.

The outrage belongs to Dr Johnson, who objected to what he saw as indecorous vocabulary, though this did not stop him mining the play for examples for his *Dictionary*. Lisa Hopkins, 'Household words: *Macbeth* and the failure of spectacle', *Shakespeare Survey*, 50 (2004), 101-10, gives an account of Johnson's outrage. Arthur Sherbo ('Dr. Johnson on *Macbeth*: 1745 and 1765', *The Review of English Studies*, New Series, 2.5 (1951), 40-7) and Anne McDermott ('The defining language: Johnson's Dictionary and *Macbeth*', *The Review of English Studies*, New Series, 44.176 (1993), 521-38) detail the use he made of the play in his *Dictionary*.

Readers wishing to extend work on the play's language beyond this chapter should look to the play's most recent editors for Cambridge and Oxford (A.R. Braunmuller (ed.), *Macbeth* (Cambridge, 1997) and Brooke – see above), both have extended sections in their introductions on the language of the play. Frank Kermode's chapter in *Shakespeare's Language* (2000) is excellent, but most important, and useful, of all is Russ McDonald's account of the play (*Late Style*, 43-52, and 33-4) - we would suggest this as the first port of call after this essay. Braunmuller (ed.), *Macbeth* (Cambridge, 1997) and Brooke – see above), both have extended sections in their introductions on the language of the play. Frank Kermode's chapter in *Shakespeare's Language* (2000) is excellent, but most important, and useful, of all is Russ McDonald's account of the play (*Late Style*, 43-52, and 33-4) - we would suggest this as the first port of call after this essay.

it reflected back upon the murder a peculiar awfulness and a depth of solemnity; yet... I never could see *why* it should produce such an effect<sup>2</sup>

(p. 389)

De Quincey concludes that the knocking, heard repeatedly both as *actual* knocking, and in the echoing vocabulary of the scene ('Here's a knocking indeed... Knock, knock, knock... Knock, knock... Knock, knock, knock... Knock, knock' 2.3.1-15) serves to mark the point where 'the human', or everyday, starts to reassert itself over 'the fiendish' which has dominated the play hitherto. This manifestation of normality paradoxically makes us more aware of 'the awful parenthesis' that has suspended everyday life. Similarly, it signals the coming focus on Macbeth and his tormented guilt:

in the murderer... there must be raging some great storm of passion, -- jealousy, ambition, vengeance, hatred, -- which will create a hell within him; and *into this hell we are to look*. (p. 392)

De Quincey's final phrase (emphasis ours) is as threatening as anything in the play itself, but most critics, while agreeing that the play has a particular, uncanny, power, locate the source of that power, not at the mid-point of the action, nor in stage business, but right at the start, and in the language. For them, the effects of dis-ease produced by what one critic has called the 'infected' language of the play<sup>3</sup> can be traced to vocabulary, metrics, grammar, and sound patterning. These features begin with the witches in scene one and spread out via 'the mouths of the Macbeths and their porter',<sup>4</sup> so that eventually characters as diverse as the First Murderer, Duncan, and Banquo speak 'for the play' rather than themselves.<sup>5</sup>

---

<sup>2</sup> Thomas De Quincey, 'On The Knocking At The Gate In Macbeth', *The Collected Writings of Thomas De Quincey*, ed. D. Masson (1896-97), vol. x, 389-94. (We have followed modern editions in correcting 'murderer' to 'murder' in the quotation.)

<sup>3</sup> McDonald, *Late Style*, 44.

<sup>4</sup> David L. Kranz, 'The sounds of supernatural soliciting in *Macbeth*', *Studies in Philology*, 100.3 (2003), 367.

<sup>5</sup> Brooke (ed.) *Macbeth*, 12-3.

L.C. Knights fixes the first element of this common language, writing of the 'sickening see-saw rhythm' of the witches, in an early focus on metrical form subsequently extended by David Kranz and Russ McDonald.<sup>6</sup> It is worth unpicking this phrase in precise technical detail, as Knights' insight really does open up the linguistic peculiarity of the play. What Knights means by 'sickening see-saw' is the trochaic tetrameter the witches use:

When shall we three meet again?

When the hurlyburly's done,  
When the battle's lost and won.

Fair is foul, and foul is fair (1.1.1, 3-4, 11)

Trochaic verse uses feet consisting of two syllables arranged in a **strong**-weak pattern ('**hur-lybur-ly**'), which contrasts with Shakespeare's more normal iambic pattern, where syllables are arranged weak-**strong** ('suc-**cess**', 're-**flect**'). Trochaic verse is more obviously metrical than iambic: we hear the rhythm as insistent because it is less natural in English, and this is what Knights means by the 'sickening' effect:

**When** shall **we** three **meet again**?

**When** the **hurlyburly's done**,  
**When** the **battle's lost and won**.

**Fair** is **foul**, and **foul** is **fair** (1.1.1, 3-4, 11)

---

<sup>6</sup> L.C. Knights, *Explorations: Essays in Criticism Mainly on the Literature of the Seventeenth Century* (1946), 20; Kranz 'Supernatural Soliciting', 351-2; McDonald, *Late Style*, 47-8.

Notice, in contrast, how much more natural, less incantatory, Macbeth’s iambic pentameter echo of the witches sounds:

So *foul* and *fair* a *day* I *have* not *seen*. (1.3.38)

And there is another, perhaps more subtle, metrical disruption in the witches’ language. In normal tetrameter, each line has four feet of two syllables each, so in well-behaved trochaic tetrameter, we expect eight syllables per line (4 x 2 = 8). For the weird music of the witches however, Shakespeare uses what is called a ‘catalectic’ metre, dropping a syllable: in this case the final syllable of the line. So the witches’ lines characteristically have seven rather than eight syllables (see Table 1).

**Table 1: Metrics in *Macbeth* 1.1.4**

<i>foot</i>	1		2		3		4	
<i>number:</i>								
<i>syllable</i>	1	2	3	4	5	6	7	8
<i>number:</i>								
<i>stress:</i>	s	w	s	w	s	w	s	w
<i>line:</i>	<b><i>Fair</i></b>	is	<b><i>foul</i></b>	and	<b><i>foul</i></b>	is	<b><i>fair</i></b>	-

It is tempting to note, as a result of this, that when Macbeth calls the Witches, ‘imperfect speakers’ (1.3.70), his words are true metrically, as well as in all their other senses. One consequence of the catalectic nature of the metre is that the Witches’ lines end on a strong syllable, rather like a ‘normal’ iambic line. The verse-music of the Witches can thus be seen as a partial pre-echo of the dominant metre of the play; an early suggestion that all is not right with the moral world, and that the chaos to come is fated, prefigured in the very metrical forms of the language.

Critics have found this 'see-saw' rhythm in other characters' verse. McDonald (47) detects it in Macbeth's 'This supernatural soliciting / Cannot be ill; cannot be good' (1.3.130-1), and Kranz notes:

While the witches disappear near the end of 1.3, many of the poetic patterns they engendered do not. The selfsame tune, the aural embodiment of their unholy spirit, makes its way into the mouths of several characters (‘Supernatural Soliciting’, 357)

It is not only trochaic metre that spreads from the witches' first scene; as Frank Kermode comments:

Here, perhaps more than in any other of Shakespeare's plays, an idiosyncratic rhythm and a lexical habit establish themselves with a sort of hypnotic firmness. 'Lost and won', say the Sisters at the beginning of the first scene: 'What he hath lost, noble Macbeth hath won', says Duncan at the end of the second, having just before that rhymed 'Macbeth' with 'death'.

(*Shakespeare's Language*, 203-4)

Russ McDonald explores the possible effects on the reader of this dense sonic repetition:

A major result of such extreme verbal compression is to magnify the relations between similar words and sounds, making their identities more audible and more potentially, or at least apparently, significant... echoing is aurally satisfying and intellectually tantalising... echoing sounds register with unusual force because they reverberate in so short a space. Not only are words repeated ('double, double, toil and trouble') but consonants and vowels are doubled and trebled, rhythmic configurations repeated insistently, and phrases and images reiterated, not just immediately but memorably, across several scenes (Late Style, 45-7)

There is more to this than simple repetition and echo, which are, after all, a

normal feature of Renaissance verse. George Walton Williams points out that the echoes of *Macbeth* often have an 'eerie *secondary force*' ('Verbal Echoing', 240). When Macbeth apparently repeats the witches' 'fair' and 'foul', for example, he has yet to meet them. For Williams, his use of their words suggests that

there is a bond between them and him, more significant than mere repetition of diction. He is ready to receive them when they come to him. Macbeth did not hear the Witches, but he knows how they speak and so knows how they think; speaking their words, he speaks their thoughts

(240-1)

Similar echoes occur elsewhere, perhaps most notably in scene 1.5 when Lady Macbeth greets her husband with his titles,

Great Glamis! worthy Cawdor!

Greater than both, by the all-hail hereafter!

(1.5.53-4)

Here, Lady Macbeth echoes the words of the witches from scene 1.3: 'All hail, Macbeth! that shalt be King hereafter' (50). She has just read one of those words, 'all-hail', in Macbeth's letter describing his meeting (they '*all-hailed me*' 1.5.6). But Macbeth does not use 'hereafter' in his letter. Like her husband, Lady Macbeth seems to be able to speak the thoughts of the witches. As Williams notes, the word 'hereafter' echoes repeatedly elsewhere, a sign of the play's concerns with time present and future, and the consequences of actions. Perhaps this 'secondary' repetition is the source of the unsettling power critics have so often detected in the play, but found hard to explain. Normally, as Russ McDonald notes, echo and repetition are aesthetically 'satisfying', but their effect in *Macbeth* seems to be discomfiting. This discomfort may come from cognitive dissonance: we recognise the echoes, but are aware that they have no logical explanation in the world of the play.

Stephen Booth offers a further explanation for the unsettling effect of

language in the play. For him, it is characterized by indefiniteness.<sup>7</sup> 'Finality', he argues, is

unattainable throughout *Macbeth*... the beginnings, sources, causes, of almost everything in the play are at best nebulous... it is almost impossible to find the source of any idea in *Macbeth*; every new idea seems already there when it is presented to us (93-4)

One of the ways in which this jarring familiarity is achieved is by the dense verbal and metrical pre-echoes we have discussed. Ideas in the play have no single identifiable source because everyone speaks the same language, and seems already to know what others think. This claim is supported by a suggestive performance tradition: *Macbeth* works extremely well when its roles are explicitly doubled.<sup>8</sup>

The presence of this pervasive linguistic style in *Macbeth* marks, for Russ McDonald, a significant shift in Shakespeare's dramatic technique. McDonald believes that in the mid-1590s Shakespeare 'reached professional maturity' by learning

to make his speakers sound like themselves...one of the triumphs of his craft... and by which he is differentiated from lesser dramatists. Prince Hal, Falstaff, King Henry, Hotspur, Owen Glendower - none of these speakers will be confused with the others... so confident of their individual voices is their creator that he even allows some speakers to parody others

(*Late Style*, 33-4)

Around 1607 however, McDonald identifies a shift in the focus of Shakespeare's language from character differentiation to plot exploration. Whole plays, rather than individual characters, now have 'styles'; and the characters in any one play

---

<sup>7</sup> Stephen Booth, *King Lear, Macbeth, Indefinition and Tragedy* (New Haven, 1983).

<sup>8</sup> A recent Japanese production (Tokyo, 2010) had all characters except the Macbeths played by the three witches, while the 2012 National Theatre of Scotland production had Alan Cumming's Macbeth locked in a mental hospital, obsessively recounting the whole play to himself, taking on the separate roles within his own character.

tend to sound like each other. For McDonald, this constitutes a key feature of Shakespeare's late style.<sup>9</sup>

Literary accounts of the language of *Macbeth* offer, therefore, a clear consensus: the play has a linguistic style that crosses character boundaries, and repetition is a characteristic trope. Can we trace this in more formal linguistic studies? We now propose to investigate the language of *Macbeth* using quantitative techniques as a way of testing literary claims about the play, and to illustrate how linguistic approaches can lead us to surprising findings about literary language.

Given what literary critics have said about the importance of repetition in *Macbeth*, we will begin with word-frequency analysis, as a way of assessing the amount of repetition in the text compared with Shakespeare's other work. Perhaps the simplest measure of vocabulary repetition across texts is the type-token ratio. A text's type-token ratio is calculated by dividing the number of *different words* (types) in that text by the total number of *all words* (tokens). Consider, for example, the text,

Fair is foul and foul is fair

which we analyse in Table 2. The total number of words (tokens) here is seven. But three of those words ('fair', 'is', 'foul') appear twice, so while the number of *tokens* is seven, the number of *types* is four ('fair', 'is', 'foul', 'and'). This gives us a ratio of  $4 \div 7 = 0.57$ .

Compare this with another text,

So foul and fair a day I have not seen

which we analyse in Table 3. This second text has ten tokens, all of which are

---

<sup>9</sup> *Late Style*, 33-4 and 47. McDonald's claim seems impressionistically right to us, but the wide availability of digital tools now makes empirical testing possible. It would not be hard to make linguistic comparisons between the styles of major characters in plays before and after 1607 to establish degrees of differentiation. Certainly our own work, based on plays, rather than characters, tends to suggest that Shakespeare's style as a whole, becomes more consistent later in his career (Jonathan Hope and Michael Witmore, 'Quantification and the language of later Shakespeare', essay for the French Shakespeare Association, forthcoming) - though we would push the beginnings of the later style closer to 1599.



distinct types (each word is a different word). So the ratio is  $10 \div 10 = 1$ .

**Table 2: Type-token counting in *Macbeth* 1.1.4**

Tokens: (cumulative total)	1	2	3	4	5	6	7
	<b>Fair</b>	<b>is</b>	<b>foul</b>	<b>and</b>	<b>foul</b>	<b>is</b>	<b>fair</b>
Types: (cumulative total)	1	2	3	4	4	4	4

**Table 3: Type-token counting in *Macbeth* 1.3.38**

Tokens: (cumulative total)	1	2	3	4	5	6	7	8	9	10
	<b>So</b>	<b>foul</b>	<b>and</b>	<b>fair</b>	<b>a</b>	<b>day</b>	<b>I</b>	<b>have</b>	<b>not</b>	<b>seen</b>
Types: (cumulative total)	1	2	3	4	5	6	7	8	9	10

Any text which has no repetition will have a type-token ratio of 1. However, most texts contain very many repeated words: words like 'a', 'the', 'and' and so on. Any text which repeats words will have a type-token ratio of less than one, with the figure decreasing as more words are repeated. Lear's

Never, never, never, never, never (5.3.307)

has five tokens, but only one type, giving a ratio of  $1 \div 5 = 0.2$ .

Intuitively therefore, type-token ratios ought to be a good measure of the relative vocabulary richness of texts. If we work out the type-token ratio for each of Shakespeare's plays, we should be able to tell which plays have more repetition, since their type-token ratios will be lower. Brian Brainerd calculated these figures in an early study of type token ratios, and gives results for *Macbeth* and the other major tragedies.<sup>10</sup> On Brainerd's figures, *Macbeth* has a type-token ratio of 0.2, which compares with *King Lear*: 0.17; *Hamlet*: 0.16; and *Othello*: 0.15. Given that a higher ratio implies less repetition, these figures suggest that *Macbeth* is *more* diverse in its vocabulary than the other great tragedies!

There are, however, several possible explanations for this surprising finding, one of which is an unfortunate statistical problem with type-token ratios. While the number of tokens in a text increases uniformly (one more token with every word, all the way through a text irrespective of length), the relative number of types tends to decrease steadily. Every word is a new token, but almost all texts consist mainly of repeated types: and the longer a text goes on, the more repetition there is. This means that type-token ratios are highly correlated with the length of the texts being analysed. Longer texts generally have more repetition, relatively fewer types, and therefore lower type-token ratios than short texts. *Macbeth*, as a notably shorter text than the other tragedies, can be expected to have a higher type-token ratio.<sup>11</sup>

If there are statistical problems with type-token ratios as a measure of vocabulary richness, there are also fundamental literary and stylistic questions

---

<sup>10</sup> Brian Brainerd, 'On the relation between types and tokens in literary text', *Journal of Applied Probability*, 9.3 (1982), 507-18.

<sup>11</sup> A good starting point for exploring measures of vocabulary richness is this post on the Corpus Linguistics blog: <http://corplinguistics.wordpress.com/2011/11/21/vocabulary-richness/>

about the effects of vocabulary richness and repetition. As David Hoover shows, our intuitions about the relative difficulty or quality of an author may not match the facts of their vocabulary use.<sup>12</sup> In a provocative study, Maria Cristina Consiglio compares the type-token ratio of Shakespeare's *King Lear* with that of Nahum Tate's infamous rewriting.<sup>13</sup> Her results confound expectations: Tate's version is more diverse in its vocabulary than Shakespeare's. Similarly, literary critics have frequently assumed that Shakespeare, as an acknowledged genius, must have had a huge vocabulary. Recent linguistic work has shown, however, that Shakespeare's vocabulary was of resolutely average size for his time.<sup>14</sup>

This suggests that we need to move away from crudely mathematical measures of style and effect. As Ward Elliott and Robert Valenza point out:

Shakespeare learned early how to strike deep, not with an outsize inventory of long, inkhorn words, but with a par-for-the-course inventory, mostly of plain words, surpassingly well chosen and put together  
(‘Dwarf’, 47)

Choice and arrangement may be more significant than raw frequency in the production of the kind of stylistic effects critics report for *Macbeth*. It may be that the frequent identification of repetition as crucial in *Macbeth* is because the play repeats certain highly *salient* words, rather than because it has high overall rates of repetition. Frank Kermode's study gives a possible example of this when he states that *Macbeth* ‘is greatly preoccupied with time’ (*Shakespeare's Language*, 202), citing in support the fact that the words ‘time’ and ‘times’ appear 44 times and 3 times respectively. By way of comparison, Kermode offers the much longer *Hamlet*, where the total is 48; *King Lear*: 27; and *Antony and Cleopatra*: 33. He

---

<sup>12</sup> David L. Hoover, ‘Another Perspective on Vocabulary Richness’, *Computers and the Humanities*, 37 (2003), 151-78.

<sup>13</sup> Maria Cristina Consiglio, ‘e-Lears: a corpus approach to Shakespeare and Tate’, in *The State of Stylistics: Pala 26*, ed. Greg Watson (Amsterdam, 2008), 191-206.

<sup>14</sup> See Hugh Craig, ‘Shakespeare's vocabulary: myth and reality’, *Shakespeare Quarterly*, 62.1 (2011), 53-74; and Ward E.Y. Elliott and Robert J. Valenza, ‘Shakespeare's vocabulary: did it dwarf all others?’, in *Stylistics and Shakespeare's Language: Transdisciplinary Approaches*, eds Mireille Ravassat and Jonathan Culpeper (2011), 34-57. Jonathan Hope, ‘Shakespeare and the English Language’, in *English in the World: History, Diversity, Change*, eds Philip Seargeant and Joan Swann (2012), 83-92, gives a summary of this work.

gives Spevack's 'relative frequency' scores, which seek to take play-length into account, and thus allow comparisons between texts of different lengths.<sup>15</sup> These figures are given in Table 4.

**Table 4: 'Time' and 'times' in *Macbeth* and other plays**

Play	Relative frequency	Raw Frequency
<i>Macbeth</i>	.267	47
<i>Hamlet</i>	.162	48
<i>King Lear</i>	.107	27
<i>Antony and Cleopatra</i>	.139	33

(Figures quoted from Kermode, *Language*, page 213 - based on Spevack, *Concordance*, vol. 3)

We can see, once play-length is taken into account, that 'time' and 'times' appear more frequently in *Macbeth* than elsewhere, though Kermode is reluctant to make much of these figures, relegating them to a footnote, and failing to explain exactly what 'relative frequency' is (the figure is derived by dividing the frequency of the word(s) being studied by the total number of words in a play and multiplying by one hundred to give a percentage).

Kermode's use of word frequencies here is broadly typical of literary critics' use of statistics: the frequency of a highly salient content word is cited as support for an argument that is actually made using traditional 'quote and discuss' methods. And the initial identification of 'time' and 'times' as words to count is made impressionistically, as a result of subjective reading, rather than emerging from a statistical test. Kermode's citation of 'relative frequency' makes his analysis more sophisticated than most, but our understanding of the relationship between the frequency of content words and 'meaning' is poor. Yes, the words 'time' and 'times' are relatively more frequent in *Macbeth* than in *Hamlet*, but someone reading or seeing a performance of the full text of both plays encounters the words almost the same number of times. Does the saliency

---

<sup>15</sup> Marvin Spevack, *A Complete and Systematic Concordance to the Works of Shakespeare* (Hildesheim: George Olms, 1973).

of the terms increase in *Macbeth* because they are less diluted by other words?

Where Kermode (and literary critics more generally) tend to identify what they see as significant themes first through their own reading, and then use frequency counts to support that claim, there are statistical tests which examine every word in a text, against their observed frequencies in a wider corpus, in order to identify all those words in a text which have unusual frequencies. For example, we can analyse the frequency of every word in *Macbeth* compared to the frequencies of those words in the whole of Shakespeare. The results of such a test (called a log-likelihood test) show us those words that Shakespeare uses more *and less* frequently in *Macbeth* than in his work as a whole.

There are several advantages to such a test over the less systematic method used by Kermode. Perhaps most importantly, log-likelihood identifies words which are unexpectedly reduced in frequency, as well as those which occur more often than normal. Additionally, log-likelihood detects frequency shifts in very common words; notably function words such as pronouns, determiners and prepositions. Again, this is something human readers are relatively poor at doing: we are alert to small increases in rare items such as 'time', but blind even to very large shifts in the frequency of words such as 'and'.

The web-based text analysis tool, WordHoard allows anyone to run log-likelihood tests on Shakespeare.<sup>16</sup> For this chapter, we made a log-likelihood comparison between every word in *Macbeth* and the occurrence of each word in all of Shakespeare. The most significant results are shown in Table 5. The analysis gives a list of thirty-three words which show a difference between their actual frequency in the play, and their expected frequency given Shakespeare's practice elsewhere. The table may look confusing, so we will use the highest scoring word, 'thane' as an example of what each column means.

---

<sup>16</sup> WordHoard can be downloaded from:

<http://wordhoard.northwestern.edu/userman/index.html>

Log-likelihood is explained in more detail in the excellent user documentation found at the WordHoard site:

<http://wordhoard.northwestern.edu/userman/analysis-comparewords.html#loglike>

**Table 5: Log-likelihood comparisons between *Macbeth* and Shakespeare (top 33 results – generated by WordHoard)**

Lemma	Word class	Relative use	Log likelihood	Analysis parts per 10,000	Reference parts per 10,000	Analysis count	Reference count
thane	n	+	156.1****	18	0.35	30	30
hail	v	+	51****	10	0.82	17	71
knock	v	+	49.8****	11	1.17	19	101
cauldron	n	+	41.6****	4.8	0.09	8	8
our	po	+	41.5****	71	36.76	119	3180
she	pn	-	41.3****	21	53.05	35	4590
the	dt	+	41****	422	327.07	703	28298
tyrant	n	+	40.2****	9	0.89	15	77
sleep	n	+	33.3****	9.6	1.38	16	119
weird	j	+	31.2****	3.6	0.07	6	6
i	pn	-	28.6****	283	359.55	472	31108
trouble	n	+	27.3****	5	0.43	9	37
dagger	n	+	26.3***	6	0.61	10	53
wood	n 1	+	25.9***	6.6	0.8	11	69
fear	n	+	24.5***	13.8	3.88	23	336
nature	n	+	23.1***	15.6	5	26	433
we	pn	+	20.5**	98	67.3	164	5823
double	v	+	20.3**	3.6	0.23	6	20
horror	n	+	20.3**	3.6	0.23	6	20
you	pn	-	19.8**	128	171.26	213	14817
which	crq	+	17.5*	48	28.71	80	2484
upon	acp	+	17.5*	37.8	21.07	63	1823
deed	n	+	17*	10.8	3.32	18	287
new	j	+	16.7*	10	3.05	17	264
castle	n	+	15.5*	4	0.55	7	48
air	n	+	15*	8	2.36	14	204

yet	av	+	14.9*	34	19.41	57	1679
bloody	j	+	14.8	9	2.68	15	232
time	n	+	14.8	28.8	15.50	48	1341
strange	j	+	14.5	10	3.36	17	291
worthy	j	+	14	10.8	3.78	18	327
love	v	-	13.7	4	13.15	7	1138
a	dt	-	13.5	144	181.24	240	15681

The first column, headed ‘Lemma’ lists the word being analysed.<sup>17</sup> The second column gives its part of speech (in this case ‘n’ for noun). The third column is crucial, as the ‘+’ or ‘-’ sign indicates whether the frequency of the word is raised or lowered in *Macbeth* compared to Shakespeare’s normal use. We can see from the ‘+’ sign that ‘thane’ occurs more frequently in *Macbeth* than we would expect, given Shakespeare’s use of it in his work as a whole. The fourth column gives the log-likelihood score: the higher this is, the greater the shift in frequency over or under expectation. Stars are used to indicate degrees of statistical significance: four indicate a result very unlikely to be due to chance, with the degree of confidence decreasing as the number of stars decreases. ‘Thane’ has by far the highest log-likelihood score in the play, and is given four stars, indicating that the result is very unlikely to be due to chance (or normal variation).<sup>18</sup>

The next columns are very useful in allowing us to judge the extent of any shift in usage. ‘Analysis parts per 10,000’ tells us how many times the word occurs every 10,000 words in the analysis text (which in this case is *Macbeth*). So we can see that ‘thane’ occurs 18 times every 10,000 words in *Macbeth*. We can compare this with the next column, which tells us how many times ‘thane’ occurs in the ‘Reference’ sample (which is the whole of Shakespeare’s work). Here we

---

<sup>17</sup> ‘Words’ are not straightforward entities to define or count. WordHoard distinguishes between ‘spelling’ and ‘lemma’. A ‘spelling’ count operates strictly on spelling: so ‘dagger’ and ‘daggers’ will be treated as two separate words. A ‘lemma’ count works by dictionary entry: so ‘dagger’ and ‘daggers’ would be one word, as would ‘am’ and ‘is’. The choice between using ‘spelling’ and ‘lemma’ lies with the researcher. In this study, we have used ‘lemma’.

<sup>18</sup> We note in passing that Table 5 shows that the lemma ‘time’ is indeed raised in frequency in *Macbeth*, but not to a level which is accorded statistical significance.

see that 'thane' occurs just 0.35 times every 10,000 words in Shakespeare. Another way of assessing this is provided by the next two columns, which give the raw counts for 'thane' in the analysis text (*Macbeth*) and the reference text (the whole of Shakespeare). We can see that 'thane' occurs 30 times in *Macbeth*, and 30 times in the whole of Shakespeare: in other words, every Shakespearean instance of 'thane' comes in *Macbeth*.

Given this, it is hardly surprising that log-likelihood is telling us that the use of 'thane' in *Macbeth* is unusual compared to Shakespeare's normal usage. But it is also fair to say that this result is not very interesting. It is easy to explain why the Scottish word 'thane' appears in Shakespeare's only Scottish play. The presence of the word in this play, and nowhere else, is not an interpretive problem.

There are several other words which have similarly obvious explanations. 'Cauldron' and 'weird' are also found only in *Macbeth*. 'Hail', 'knock', 'tyrant', 'sleep', 'trouble', 'dagger', and 'wood', although used elsewhere, all show a raised frequency over the expected one, and all are clearly linked to particular plot events, or themes. *Macbeth* is a play in which daggers feature; characters are murdered in and worry about their sleep; there is an extended passage about knocking at a door; a wood moves. We hardly need computers and advanced statistics to tell us this.

However, log-likelihood comes into its own as a discovery technique if we shift away from obvious content words. There are several function words in the table whose presence is harder to explain. 'Our' is much more frequent than we would expect, occurring almost twice as often (71 times every 10,000 words against Shakespeare's norm of 37 times). 'She' is reduced in frequency, at 21 per 10,000 words versus 53 (which we might find surprising, given the prominence of Lady Macbeth). Perhaps most puzzling, 'the' is raised in frequency to a highly significant degree. It is easy to skip over these words when 'reading' a log-likelihood table, since the content words tend to catch our eye, but it is the function words that often raise the most interesting problems of explanation.

For us, the most intriguing log-likelihood result for *Macbeth* concerns the determiner 'the'. This is exactly the sort of word literary critics tend not to comment on: indeed, it is exactly the sort of word even early quantitative and



digital analysts excluded from their work on the assumption that it was too common to be interesting. But experience of log-likelihood tests shows that it is often the most frequent words that prove the most interesting, because they are the most difficult to explain: and they are the words that can only be identified as interesting through the use of digital analysis and complex statistics. Most important of all, the effect of such words can only be explained by returning to the text and looking carefully at their contexts of use. Quantitative and digital analysis does not distance us from the text: it sends us back to it with new questions.

First, let's make the increase in 'the' explicit. WordHoard's figures show that 'the' occurs in *Macbeth* 422 times every 10,000 words. In Shakespeare as a whole, 'the' appears only 327 times every 10,000 words. In terms of actual instances, 'the' appears 703 times in *Macbeth*; but if Shakespeare were behaving 'normally', it would only appear 545 times. So we can say that there are about 150 'extra' 'the' forms in *Macbeth*.

How can we account for this? There are several issues that we should think about here. One possibility is that the play is using the definite article ('the') in situations where Shakespeare would, in other plays, have used the indefinite article ('a'). When we look at the log-likelihood result for 'a' at the bottom of Table 5, we find something very suggestive. A minus sign in the 'Relative use' column tells us that the frequency of 'a' in the play is indeed lower than we would expect given Shakespeare's normal behaviour. However, the log-likelihood score for 'a' is just 13.5 (as opposed to 41 for 'the'), and there are no stars, suggesting that the effect could be due to chance. So we need to be cautious about this result: it is not as strong as that for 'the', but it is interesting. Shakespeare normally uses 'a' about 181 times every 10,000 words. In *Macbeth*, this drops to 144 times every 10,000 words. The observed frequency (240 'a' forms in the play) is about 60 down on the expected (300 forms). So there is a shift from 'a' to 'the' in the play, though not enough to explain the whole increase in 'the' forms.

Checking the results for the other determiners produces some similar results: 'some' and 'this' are also down on expectation, though again, not at the level of significance (these results are not given in Table 5). 'That' and 'those', on

the other hand, are raised slightly. So there seems a reasonable statistical case for saying that there is a tendency in *Macbeth*, compared to the canon as a whole, for determiners to be definite ('the', 'that', 'those') rather than indefinite ('a', 'some'). We will return to this.

Another possible explanation is that *Macbeth* has more noun phrases than other plays, and that this allows more opportunities for Shakespeare to use determiners. WordHoard uses a 'deeply tagged' corpus of Shakespeare: that is, all of the words have been tagged for grammatical function. So we can search on 'word class'. This isn't going to answer our question exactly, since WordHoard marks 'nouns' rather than 'noun phrases', but comparisons between the number of nouns in Shakespeare's plays will get us close enough to what we want to know. When we look at the log-likelihood table for word class in *Macbeth*, we find that nouns are raised slightly, though with a very low log-likelihood score of 1.0. There are 2,991 nouns in the play, 90 more than the expected result (2905) given Shakespeare's practice elsewhere. So it is possible that what we are seeing in the very strong result for 'the' is a combination of a shift in determiner choice from 'a' to 'the', and a slight rise in noun frequency, with those 'extra' nouns tending to take definite determiners.<sup>19</sup>

So, this is a finding, but is it an interesting one? Does it tell us anything about the way the play works? Common sense suggests that an increased use of definite articles should make the play more definite than Shakespeare's other work. This is puzzling, though, since the subjective experience of reading the play, as reported by generations of literary scholars, is not one of definiteness, but rather of *indefiniteness*; looming, overwhelming, but ill-defined, fear and dread. This mismatch between subjective impression and objective finding is interesting, precisely because it is counter-intuitive and difficult to explain. Generally in literary studies, things that are difficult to explain produce more

---

<sup>19</sup> While investigating this we came across a slightly surprising side-result: proper nouns (the names of people and places, not normally able to take any kind of determiner, so not relevant to our central question here) are strongly lowered in frequency in the play. This is surprising, since our subjective impression is that the play is filled with names (Macbeth, Glamis, Cawdor), which contribute to an overpowering, almost claustrophobic, sense of location. But the figures do not back this up. Perhaps the subjective effect is produced by the repetition of a few names at highly salient moments. This is another warning of the dangers of assuming too simple a relationship between frequency and literary effect.

insightful discussion than things that are easy to explain.

Is it possible that an increase in the use of 'the' is producing an effect of indefiniteness? Once we start examining the ways 'the' is used in the play, we notice some fascinating patterns. Just before the murder, the Macbeths are startled by noises in the night, which Lady Macbeth explains as follows:

It was the owl that shriek'd, the fatal bellman,  
Which gives the stern'st good-night. (2.2.4-5)

A more expected determiner here would have been 'an': 'It was *an* owl that shriek'd'. This is because Lady Macbeth is explaining an unexpected event, assigning to it a previously unknown cause. The expectation in English is that *new* information, like this, is introduced using indefinite articles, while *given* information (something we already know about) is marked with definite articles. So why does Lady Macbeth use 'the' here? The effect is to present the owl, not as an actual, specific owl, but as a generalised, mythical or proverbial owl; and this shift is clear in her epithet 'the fatal bellman', which also, much more expectedly, uses the definite article. So Lady Macbeth's choice of determiner shifts the owl from the immediate, specific 'now' of the play, into a less determinate mythological space and time. The owl becomes an idea, rather than a thing.

There are further instances elsewhere of 'the' being used to introduce ideas which, strictly speaking are new to the discourse of the play:

To beguile the time,  
Look like the time; bear welcome in your eye,  
Your hand, your tongue: look like th'innocent flower,  
But be the serpent under't. (1.5.62-5)

Here, 'th'innocent flower' and 'the serpent' are treated as given information: the language acts as though we already know about these things. This assumption of knowledge produces the murky, claustrophobic feeling critics have often detected in *Macbeth*: it gives the feeling that everything has been decided already. Similarly in Macbeth's speech:

If it were done, when 'tis done, then 'twere well  
It were done quickly: if th'assassination  
Could trammel up the consequence, and catch  
With his surcease success; that but this blow  
Might be the be-all and the end-all - here,  
But here, upon this bank and shoal of time,  
We'd jump the life to come. (1.7.1-7)

Although these ideas are new (indeed, several of the nouns introduced by 'the' have their first recorded appearance in English in this passage), they are presented as given information. There is thus a tension in the language: unfamiliar, abstract concepts and formal terms are presented as if they were familiar, everyday things. The language behaves as if the things it talks about were certain, but it deals with abstracts, concepts, metaphors.

Elsewhere, we find 'the', and other definite determiners operating to form nouns out of adjectives:

And you whose places are the nearest (1.4.36)

to concretise abstract ideas:

Great Glamis! worthy Cawdor!  
Greater than both, by the all-hail hereafter!  
Thy letters have transported me beyond  
This ignorant present, and I feel now  
The future in the instant. (1.5.53-7)

and to give definiteness to reference that is general:

Let not light see my black and deep desires;  
The eye wink at the hand; yet let that be,  
Which the eye fears, when it is done, to see. (1.4.51-3)

Why does Macbeth not say 'my hand' and 'my eye' here? We would suggest that this is the beginning of the dissociation between desire, act and guilt which wracks him through the play; and which is manifest in his ascribing independent agency to the eye and the hand, achieved linguistically by substituting 'the' for the expected possessive pronoun. We see this process elsewhere in Shakespeare's plays: for example, in Hamlet's descriptions of how the will fails to inspire the body to put its intentions into act. But Hamlet's more intellectual analysis of the distributed nature of human agency is given grammatical form in *Macbeth* as body parts take on an identity of their own; unclaimed, and so perhaps un-owned, by the person to which they belong.

We can link these linguistic observations to Stephen Booth's critical claim about indefiniteness in *Macbeth*. As we have seen, Booth claims that 'it is almost impossible to find the source of any idea in *Macbeth*; every new idea seems already there when it is presented to us' (*Indefiniteness*, 94). Booth made this claim without knowledge of the statistical rise in 'the' forms in the play, but the mathematics back him up. *Macbeth* is a play which presents the new as old, and the indefinite and uncertain as certain. Paradoxically, the effect of this apparent certainty is uncertainty and dislocation in the reader: the formal structures of the language, the function words, are telling us one thing (things are familiar, things are fixed, things are certain); but the content words are telling us something else ('into this hell we are to look'). A determiner tells us *how* to think about the specifics being mentioned. It doesn't simply specify the number of a noun, but an assumed perspective on the world in which that noun appears: exhaustive, partial, hypothetical, matter of fact. That perspective is, in *Macbeth*, a jarring one: the determiner calls attention to the fact that the listener does not know the noun being discussed, and indeed, could not know such a thing. It is as if, in some of these uses, listeners are encountering a certain kind of category mistake, like hearing someone claiming to know just how many hairs are in Macbeth's beard. We reject not the proffered count, but the presumption that such a thing can be counted in the first place.

We can sense this type of ongoing category mistake in the following passage, which alienates the listener by assuming he or she ought to be in the

know – or that the characters know something terribly specific that cannot be stated:

Glamis thou art, and Cawdor; and shalt be  
What thou art promis'd. - Yet do I fear thy nature:  
It is too full o'th' milk of human kindness,  
To catch the nearest way. Thou wouldst be great;  
Art not without ambition, but without  
The illness should attend it: what thou wouldst highly,  
That wouldst thou holily; wouldst not play false,  
And yet wouldst wrongly win; thou'dst have, great Glamis,  
That which cries, 'Thus thou must do,' if thou have it;  
And that which rather thou dost fear to do,  
Than wishest should be undone. Hie thee hither,  
That I may pour my spirits in thine ear,  
And chastise with the valour of my tongue  
All that impedes thee from the golden round,  
Which fate and metaphysical aid doth seem  
To have thee crown'd withal. (1.5.14-29)

'th' milk of human kindness'; 'the nearest way'; 'The illness should attend it': these all assume, or imply, a knowledge and familiarity that is impossible. The sickening lack of specificity is reinforced by the run of pronouns and pronominal phrases whose referent is assumed, but difficult to identify: 'what thou wouldst'; 'That'; 'That which'; 'it'; 'that which'. And by a set of ellipted verbs and pronouns:

what thou wouldst [*have*] highly,  
That wouldst thou [*have*] holily; [*thou*] wouldst not play false,  
And yet [*thou*] wouldst wrongly win [*it*]

As much as repetition of content words and sounds, this forms the characteristic, queasy music of *Macbeth*: high density of pronoun replacement and ellipsis – both features that assume shared knowledge on the part of speaker and hearer.

When the audience can supply a referent for pronouns, and can fill the ellipsis with a suitable word, the effect is to build a strong bond between speaker and hearer: the hearer guesses the speaker's words without actually hearing them, as the Macbeths do the witches'. But if the referents of pronouns shift without warning, as they do in Macbeth's

If it were done, when 'tis done, then 'twere well

It were done quickly

(1.7.1-2)<sup>20</sup>

and if the shared knowledge on which ellipsis depends is absent, then the effect is dislocation, and, ultimately, horror.

---

<sup>20</sup> Jonathan Hope, *Shakespeare's Grammar* (2003), 13-6, discusses this passage in grammatical and stylistic detail.