Word Length Frequency and Distribution in English: Part II. An Empirical and Mathematical Examination of the Character and Consequences of Isometric Lineation

John Constable and Hideaki Aoyama
Kyoto University, Kyoto, Japan

Abstract

In this paper we build on earlier observations and theory regarding word length frequency and sequential distribution to develop a mathematical characterization of some of the language features distinguishing isometrically lineated text from unlineated text, in other words the features distinguishing isometrical verse from prose. It is shown that the frequency Q_n of n syllables making complete words produces a flat distribution for prose, whereas that for verse exhibits peaks at the line length position and subsequent multiples of that position. Data from several verse authors are presented, including a detailed mathematical analysis of the dynamics underlying Q_n peak creation, and comments are offered on the processes by which authors construct lines. We note that the word length sequence of prose is random, whereas lineation necessitates non-random word length sequencing, and that this has the probable consequence of introducing a degree of randomness into the otherwise highly ordered grammatical sequence. In addition, we observe that this effect can be ameliorated by a reduction in the mean word length of the text (confirming earlier empirical observations that verse tends to use shorter words than would otherwise have been selected), and also by the use of lines varying from the core isometrical set. The frequency of variant lines is shown to be coincident with the frequency of polysyllables, suggesting that the use of variants is motivated by polysyllabic word placement. The restrictive effects of different line lengths, the relationship between metrical restriction and poetic effect, and the general character of metrical rules are also discussed.

Correspondence:
Hideaki Aoyama, Faculty of Integrated Human Studies, Kyoto University, Kyoto, 606-8501, Japan
E-mail:
aoyama@phys.h.Kyoto-u.ac.jp

1 Introduction: Verse and Prose

When Auden wrote that 'The difference between verse and prose is self-evident' (1963, p. 23), he was not suggesting anything controversial, and would presumably have agreed with Ricks that the variation presented 'no interesting puzzle' (1970, p. 261). Yet, as Frye has observed, this 'most far-reaching of literary facts' is a 'distinction which anyone can make in practice' but that 'cannot be made as yet [. . .] in theory' (Frye, 1957, p. 13). There has been much discussion since Frye's remarks, and some progress, Brogan even citing this passage in his *Princeton Encyclopedia of Poetry and Poetics* entry, 'Verse and prose', to illustrate the distance now separating the poetics of the mid-1950s from that of the early 1990s (1993c, p. 1346). However, as the rest of that article demonstrates, a clear theoretical distinction has not yet been articulated in detail, and it is still a commonplace in the literature that 'Distinguishing between prose and verse remains as perplexing an issue as ever' (Steele, 1990, p. 81; see also Marks, 1998, p. 148). This is somewhat surprising, as the elements for a satisfactory theory have long been available, at least since Omond remarked that difference was one of 'mechanical method':

> The units of prose are diverse, irregular in length, rarely conformed to a common pattern. In verse, on the other hand, succession is continuous. Something recurs with regularity. (Omond (1903), quoted in Matthews (1911, p. 31))

Lotz (1972) puts this general point in a still more satisfactory way:

> Verse and prose are opposed to each other as two types of which, one, verse, has definable properties, and the other, prose, is characterized by lack of any such features. (1972, p. 1))

> In most languages there are texts in which the phonetic material within certain syntactic frames, such as sentence, phrase, and word, is numerically regulated. Such a text is called verse, and its distinctive characteristic is meter. [. . .] A non-metric text is called prose. Numerical regulation may refer to a variety of phenomena; therefore, verse and prose are distinguished not as two sharply differentiated classes, but rather as two types of texts. (This, however, should not obscure the fact that verse and prose are polar opposites [. . .].) (1972, pp. 4–5)

> [. . .] the deviation [of verse from prose] can be put in terms of numerical regularity, or meter, and this regularity is the *differentia specifica* of verse. (1972, p. 6)

Nevertheless, to our knowledge, there is neither empirical work demonstrating the existence of these regularities, nor theoretical work examining their mathematical character.

We will attempt to remedy this situation by offering theory and data bearing on the distinction between prose and isometrically lineated text in English. Lineation, as may be inferred from Lotz's principles, is a basic

feature of metre, and there is now general agreement that the arrangement of text in lines is the most commonly occurring rule in the world's metrical systems (Brown, 1991, p. 132; Lotz, 1972, p. 19; Tarlinskaia, 1989, p. 122). Indeed, some scholars have gone further and suggested that there are good reasons for accepting the cultural 'universality of the line' (Dell Hymes, quoted in Leavitt, 1997, p. 134). However, even thorough literary linguistic studies and manuals content themselves with only brief discussions of verse lineation (Leech, 1969, p. 114; Williams, 1986, pp. 182–7; Tarlinskaia, 1993, pp. 3–5; Gasparov, 1996, pp. 1–3; Fabb, 1997, pp. 88–91; Burns Cooper, 1998, pp. 92–3), and attention has been mostly directed onto the admittedly very interesting question of rhythm within lines (Attridge, 1982, 1995). In the narrower field of English literary criticism, the phenomenon has drawn much comment relating to its effects on interpretation (see Culler (1975, pp. 183ff.) for a discussion of several interpretative strategies by which line breaks are 'accorded some kind of value'), but little detailed examination of its nature. When not engaged with subtle interpretational questions, critics locate their analysis at a high level of abstraction, and range over and attempt to conjoin the concepts of isometric lineation, the lineation of free verse, and even poetic effect itself, a very broad range of subjects (Holder (1995, pp. 137ff.) provides a useful review of several prominent authors). This extremely ambitious project has served to show the interest of the issues involved but fails to deepen scholarly understanding of the questions under consideration.

The last of these, poetic effect, has been the cause of such a long-standing and extensive literary critical debate on the differentia of *poetry* and prose (see, e.g. Darwin, 1789, pp. 40–1; Newton Scott, 1904; Lotspeich, 1922; Murry, 1922, pp. 47–70; Alexander, 1933, pp. 84ff.; Read, 1949, pp. ix–xiii; Whately, 1963, pp. 333–4; Furniss and Bath, 1996, pp. 12–13; Leonard, 1996, p. 75) that we should, perhaps, make it clear at the outset that we accept the substance of Wordsworth's remarks that 'much confusion has been introduced [. . .] by this contradistinction', and believe him to be correct when he writes that the 'only strict antithesis to prose is verse' (1974, p. 134; see also Hamer (1930, p. 1) and Turco (1986, p. 5) for helpful remarks on this issue).[1] This is not to slight the importance of the relationships between prose, verse, and poetic effect, but we submit that consideration of the third term and its relations must wait on better understanding of the differences between the first two. Consequently, we will concern ourselves initially with the question of illuminating those features that make regularly lineated text distinct from unlineated, and only then will we turn, briefly, to the vaguer and more problematic question of rich poetic effects and their peculiar association with both prose and verse.

By isometric lineation we understand a requirement that lines be of the same length, and we shall focus on length measured in syllables, although this is in fact only indirectly specified through regulation of the number of beats and offbeats (Attridge, 1982, 1995). Variations in length result from flexibility in the realization of the beat/offbeat pattern, and an

1 Wordsworth subsequently revised his text, arguing that as 'lines and passages of metre so naturally occur in writing prose' the division of prose and verse was not 'in truth, a *strict* antithesis' (1974, p. 135; see also Brooks and Warren (1960, p. 122), who come to similar conclusions). As we will show, this is not a valid conclusion.

English text notionally composed in ten-syllable lines, for example, usually displays a number of nine-, eleven-, and twelve-syllable lines in addition to its core isometrical set. Further, we include within the isometric category repetitions of heterometrical structures; for example, the Spenserian stanza, which is a structure of eight decasyllabic lines and a concluding twelve-syllable line. We will not discuss heterometrical lineation, that is to say non-stanzaic verse where there are lines of various lengths and no repetition (Brogan, 1993a), for the simple reason that our methods are not sufficiently sensitive to detect lineation in small samples. Of the lineation of free verse, we will say almost nothing, because, as will be made clear later, it lies beyond the scope of our investigation.

Most readers and writers of English verse are aware that isometrically lineated text is not merely a visually displayed distribution of words (Scott, 1979, p. 158) but is in some sense a matter of the language. Here and there in the literature there are numerous remarks that touch on this matter. Levin, for example, notes that 'Except possibly in free verse, the typographical groupings [. . .] are not random; some organizing principle must thus be at work behind them' (1971, p. 181).[2] Chatman has observed that 'The poet may select his words in part by considering the numbers of syllables they contain' and thus that 'number becomes a part of the poem's mode of existence' (1970, pp. 318–19). Lotz advances the discussion by noting the importance of the regular occurrence of word boundaries (1972, pp. 7–8), a point also made by Jakobson (1960, p. 361) in relation to Russian verse, and repeated by Fabb (1997, p. 88). Rothman, in a defence of the metrical status of pure syllabic verse, cites the word boundary rule, and very pertinently observes that lineation thus exercises some degree of control over the 'phonemic flow' in the text (1996, p. 207). All these remarks, although vague, have something to recommend them, but it is in Wimsatt and Beardsley's paper 'The concept of meter' that we find the most suggestive sketch of the heart of the matter:

> [. . .] to have verses or lines, you have to have certain broader structural features, notably the endings. Milton's line is not only a visual or typographical fact on the page, but a fact of the language. If you try to cut up his pentameters into tetrameters, for example, you find yourself ending in the middle of words or on weak words like 'on' or 'the'. Much English prose is iambic or nearly iambic, but it is only very irregular verse, because if you try to cut it regularly, you get the same awkward and weak result. (1959, p. 591)

Although we do not wish to defend the value of raising the 'weak words' issue in a discussion of the fundamentals of lineation (it is a supplementary matter in our view), or the doubtful claim that English prose is iambic, the proposition that the relation of polysyllables to line boundaries is crucial in understanding the character of lineated text deserves more attention than it has hitherto received. However, progress beyond the intuitive stage represented by these writers has been hindered

2 Some readers may be surprised to see Levin's article cited in support of this view, as he elsewhere in the same piece states that 'the line is a purely typographical device' (p. 179), and in the colloquium discussion that followed, the conference presentation of the paper (pp. 193–6) was taken to task for suggesting that lineation was not 'not linguistic'. Careful reading of Levin's text, however, reveals that his definition of 'linguistic' is a scrupulously restricted technical use, and that in his terms he can claim that lineation is non-linguistic, that it is conventional, while not calling into question the fact that lines 'comprise patterns or structures of language elements'.

by a lack of precise knowledge concerning the nature of word boundary distributions in English. Work of our own on prose (Part I of this paper, Aoyama and Constable, 1999) supplies this deficiency, and much else concerning lineation becomes clearer in the light of its analysis. Initially, our examination concerns itself with the phenomenon as a feature of completed texts, and then in subsequent sections we will change perspective and look at the question from the author's viewpoint, asking what sort of syntactical and dictional choices are forced or encouraged by lineation. It may be observed in passing that although the thesis presented here is exclusively concerned with the phenomenon of lineation in English we anticipate that many of our remarks may be abstracted and applied to other languages, although it seems probable that adjustments will be necessary to take into account features peculiar to those languages.

2 Mathematical Characteristics of Unlineated and Lineated Text

As noted above, an understanding of the character of lineation involves an examination of the distribution of polysyllables in output, and this entails a discussion of both frequency totals and sequential distribution. In this section we will summarize work of our own on prose and demonstrate that verse is significantly different in its sequential distribution.

2.1 Unlineated text: geometric distribution, random sequencing and the flat Q_n distribution

In Part I of this article (Aoyama and Constable, 1999) we have analysed the word-length structure of almost two million words of prose. By using a symbol manipulation program we computed the frequency of *all* sequences of complete words totalling one syllable, all sequences of complete words totalling two syllables, and so on throughout the text up to complete words totalling thirty syllables. Thus, we determined the number of matches to the line definition rule given by Constable, stating that a line of a specified length '*must only be complete words*' (1997, p. 181). The normalized frequency Q_n of such occurrences is defined by the equation

$$Q_n \equiv \frac{L_n}{I} \tag{1}$$

where L_n is the number of matches to a line definition rule of n syllables, and I is the total number of syllables in the data, and was found to be in general flat (the boundary condition was resolved by connecting the end of the data with the beginning), that is to say that Q_n is independent of n (see Fig. 1), in agreement with remarks made earlier by Constable (1997, p. 182).

It was also observed that there is in general no correlation between the syllable counts of adjacent words, that is to say that the number of

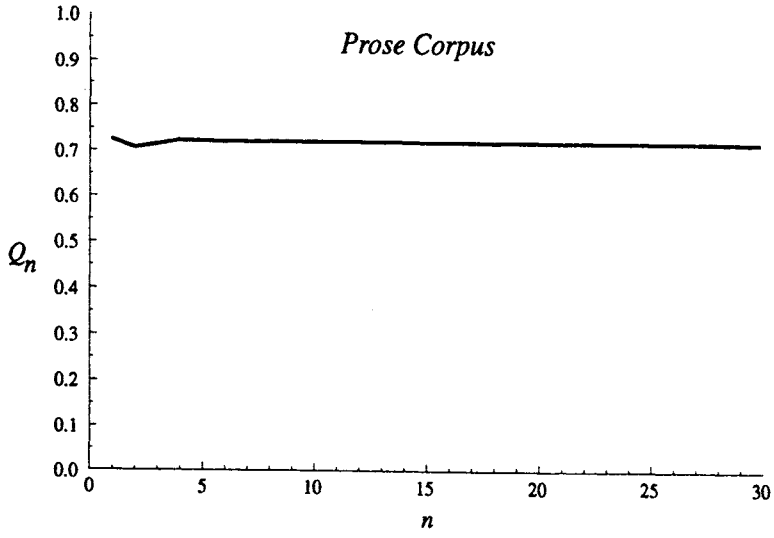

Fig. 1 Plot of Q_n for all 1,977,676 words of the prose corpus. The outstanding characteristic is its flatness, i.e. the independence of Q_n from n.

syllables in a word has no effect on the value of the subsequent word. When taken together with the flatness of the Q_n distribution this random ordering implies a geometric distribution of word length frequency totals, which can be empirically confirmed from the data. This leads us to the conclusion that *prose is randomly segmented*, and it is this finding that will provide the background property against which verse may be distinguished.

2.2 Lineated text: Q_n peaks in isometric verse, empirical studies

As described above, and demonstrated in Part I of this paper (Aoyama and Constable 1999, p. 347), the sequence of word length items in unmetred output, prose, is a random sequence of items from a geometric frequency distribution, and the Q_n frequencies produce a flat distribution. Lineated text differs considerably from unlineated text in this respect. We will show that there are distinct peaks at the position where n is equal to the most probable number of syllables per line, and at subsequent multiples of that line length (in texts where there are a number of variant line lengths, as there often are, these peaks progressively diminish in prominence).

We have analysed five verse texts (details are given in Table 1; the texts were either obtained from public on-line sources or prepared by us), the syllabic data being obtained with a simple marking program, written by one of us (Constable), which uses a custom-built lexicon to determine the syllabic count of each word. The texts chosen are all long, thus ensuring statistical reliability, and cover a broad chronological range. In addition, both rhymed (Fairfax, Chapman, and Dryden) and unrhymed compositions (Milton and Wordsworth) have been examined. It should be noted that all these texts are in five-beat duple rhythm, that is, a pattern where the offbeat is typically a single syllable, and begin their

Table 1 Verse texts analysed: line length distributions

Author	Text	Date	No. of words	Line length distribution							
				7s	8s	9s	10s	11s	12s	13s	14s
Edward Fairfax	*Jerusalem Liberata*	1600	124,856	3	45	1,027	12,653	1,632	112	7	0
George Chapman	*The Odyssey*	1616	68,170	0	3	227	7,217	1,073	67	1	0
John Milton	*Paradise Lost*	1667	79,836	0	1	163	8,315	1,887	178	4	0
John Dryden	*The Aeneid*	1693	106,483	0	1	86	11,544	1,483	515	65	5
William Wordsworth	*The Prelude*	1839 (D text)	57,570	0	1	46	6,087	1,523	181	10	1

7s, seven syllables, etc.

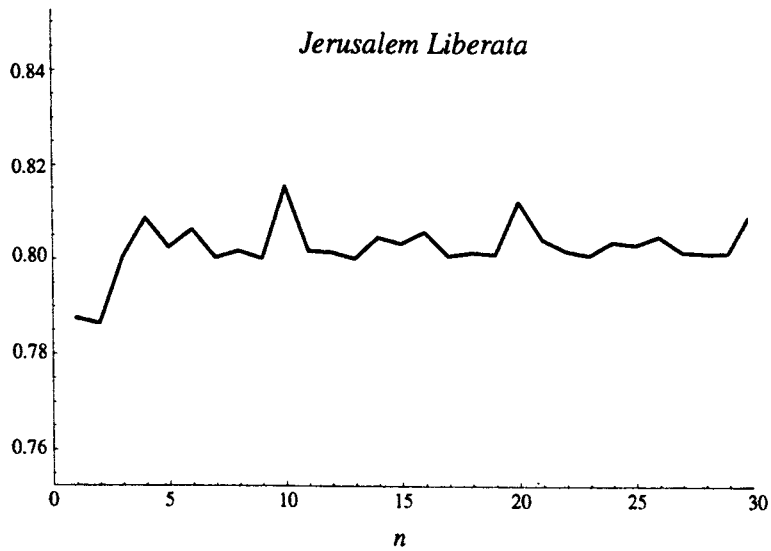

Fig. 2 Edward Fairfax, *Jerusalem Liberata*: detailed Q_n distribution.

alternating patterns with offbeats.[3] This particular structure is, of course, one of the commonest found in English metrical writings, and given our focus in this paper on the fundamentals of lineation, as opposed to the details of the beat patterning within lines, we have deferred consideration of other line forms. However, although these other line forms, for example triple rhythms (where the offbeat is typically two syllables) and patterns beginning their patterns with beats, are expected to be fundamentally similar in the properties of their lineation, the reader should note that some variation, for example in the frequencies with which variant lines are employed, is likely. These points will be touched on below when appropriate.

Figures 2–4 present data for three of our texts. In these figures, we observe that there are significant peaks over the most probable line lengths. The importance of the Q_n peaks should be immediately apparent; in addition to the visual and aural line (Brogan, 1993b, p. 696) we can now speak of the mathematical line. In other words, the view of authors who have supposed that lineation is arbitrary in terms of language, a rare position in its pure form (though see Weirather (1980), quoted in Holder (1995, p. 136)) but prevalent as a general uncertainty as to the ontological status of lines, is shown to be untenable. The claims of Wimsatt and

3 The terminology here is drawn from Attridge (1982, 1995), and is we believe technically superior to the more familiar but potentially misleading equivalent term from traditional metrics, 'iambic pentameter'.

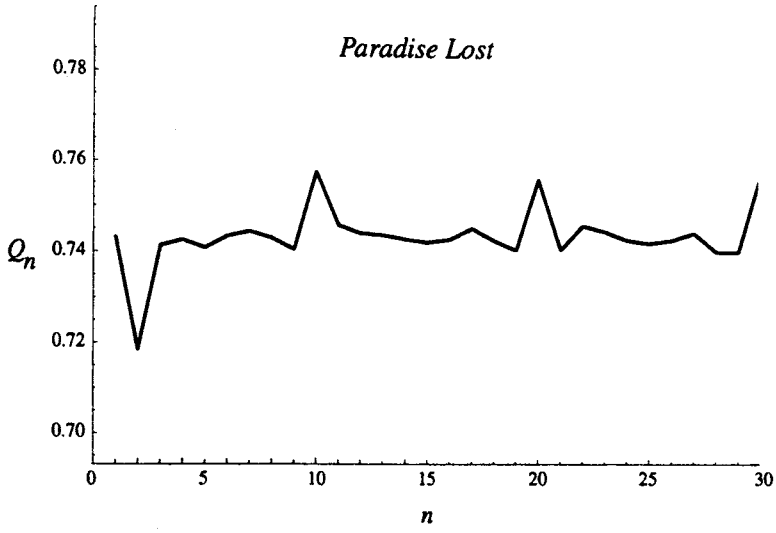

Fig. 3 John Milton, *Paradise Lost*: detailed Q_n distribution.

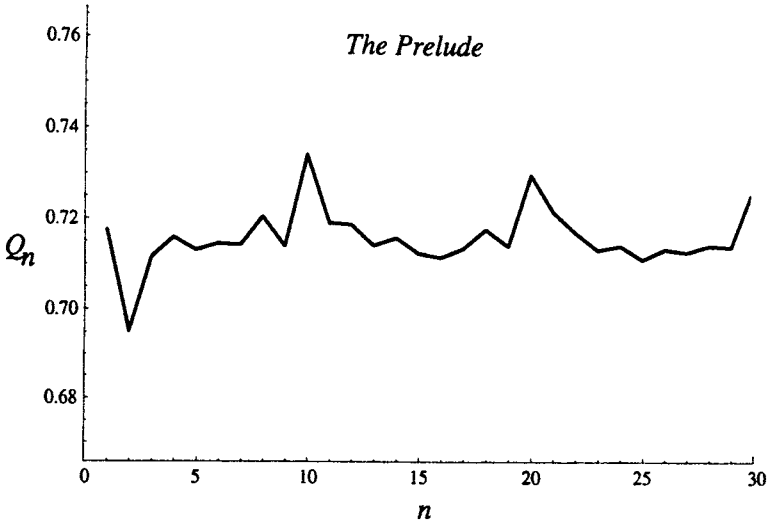

Fig. 4 William Wordsworth, *The Prelude*: detailed Q_n distribution.

Beardsley, and the intuitions of the other authors cited above, are correct. Lineation is a fact of the language.

The concrete demonstration of this point may bring some order to the debate around the issue of the relation between sound and lineation (see Holder, 1995, pp. 129–55). Some critics have wondered, for example, whether Milton's blank verse was 'verse only to the eye' (Johnson, 1967, p. 193; Bradford (1988) summarizes the positions of several other seventeenth- and eighteenth-century authors), and others have argued that pure syllabic verse is not aurally significant, and therefore 'not a metre in English' (Baker, 1996, p. 12). That lines are often perceived aurally, a position taken by many authors, for example by Hopkins (1959), and endorsed very prominently by Jakobson (1960, p. 358), should not be in question. However, our findings show that the salience of lineation for readers may not be of primary relevance to under-

standing the character of lineated text. Consequently, Johnson's scepticism, and Wallace's claim that syllabic verse is a 'kind of free verse' (1996, p. 12) seem to us to be confusing two distinct issues, on the one hand the salience of lineation, and on the other its ontological status. Many scholars have thought they were addressing the latter question, when in fact they were discussing the ease with which readers or listeners might recognize lineation if they were not able to use the typographical appearance of the page to assist them (Gioia (1996) and Rothman (1996) make similar points in their reply to Wallace (1996)). Sonic rhythm, and rhyme, undoubtedly work to mark line endings, as do carefully positioned syntactic boundaries. However, even if a text were composed in isometric syllabic lines without sonic markers, a lineation that might be very difficult indeed to recogize if printed in an undisplayed form, there would still be a real mathematical distinction between the language of such a text and that of a prose text. Whether this suggests that there is an 'essential difference between the language of prose and metrical composition' (Wordsworth, 1974) is largely a matter of terminological definition; if word length sequencing is deemed an 'essential' characteristic of language, then there is an essential difference between prose and verse, if not, then there is none. We will content ourselves with pointing out that there is a computable difference, and leave it to the reader to decide whether this is an essence or not for any particular purpose.

It might be suggested that our procedure constitutes a 'test' for lineation, assisting in, for example, analysis of authors known to have printed verse as prose, such as Melville, Scott Fitzgerald, and Dickens. However, we would caution against this, as our technique is not sufficiently sensitive to detect the presence of short lineated passages within a large body of prose, where any Q_n peak would be so small as to be effectively concealed by random fluctuations. However, the general principle that we have outlined might function as a test in cases where there is doubt as to the metrical status of relatively lengthy texts or substantial canons. For example, we suggest that with suitable adjustments our approach might contribute something to the discussion of Hebrew biblical texts, the line construction principles of which have been much debated (Yoder, 1972; Kugel, 1981, 1990; Geller, 1993). Of course, the language features of Hebrew may be so different that our method is not applicable, and even if it is appropriate it may be that difficulties in establishing reliable texts and syllable counts would stand in the way of productive research (Yoder, 1972, p. 59).

3 Composing in Lines and Q_n Peak Construction

We have now shown that isometrically lineated text is marked by significant empirical properties differentiating it from unlineated text. Although the mathematical dynamics underlying these peaks could be approached in an abstract manner, the analysis can be more firmly motivated by placing it within a consideration of the compositional process. It should

be noted at the outset that the discussion below focuses on duple verse, that is, verse where the offbeat position is typically occupied by a single syllable, and that triple verse, where the offbeat is typically or very much more frequently a pair of syllables, requires a separate and presumably very different discussion.

3.1 Theoretical line violation and alternative placement

If a passage of prose is arbitrarily broken into lines at every tenth syllable, a certain number of these breaks will lie within polysyllabic words. As the sequential distribution of word length items is random, if we know the word length frequency totals we can predict how many of these introduced breaks will actually lie within words. For example, George Eliot's *Middlemarch* contains 317,827 words, and 456,620 syllables, with a word length frequency distribution that is described in Table 2. The number of syllable boundaries internal to words is calculated in column four, and column five presents this information as a proportion of all syllable boundaries in the text.

This calculation can be carried out for the ideal geometric distribution as follows. In part I of this paper (Aoyama and Constable, 1999, p. 345), we defined the following constant frequency distribution:

$$\bar{Q}_n = q. \tag{2}$$

It was shown that this is induced by the geometric distribution for the syllable counts of each word, $\bar{p}_n = q(1 - q)^{n-1}$. The probability of a syllable boundary being internal to a word is given by

$$\sum_{n=1}^{\infty} (n-1)\,\bar{p}_n \Big/ \sum_{n=1}^{\infty} n\bar{p}_n = (1-q). \tag{3}$$

This result leads us to a straightforward interpretation: according to the definition (2), q is the probability of a given syllable boundary being a word boundary. As the above probability (3) is the reverse of this, it is given by (2). In fact, the *Middlemarch* data have an average Q_n of about $\langle Q \rangle = 0.6961$, and using this as the estimate of q, we find that the

Table 2 George Eliot, *Middlemarch*: proportions of syllable boundaries internal to words

Word length	Frequency	Normalized frequency	Number of s/bs internal to words	s/bs internal to words as proportion of all s/bs
1	223,840	0.704	0	0
2	61,588	0.194	61,588	0.135
3	22,256	0.070	44,512	0.097
4	8,118	0.026	24,354	0.053
5	1,806	0.006	7,224	0.016
6	201	0.001	1,005	0.002
7	17	0.000	102	0.0002
Total	317,826	1.000	138,785	0.3032

s/bs, syllable boundaries.

theoretical estimate is $1 - q = 0.3039$, which is in good agreement with the value 0.3032 obtained in Table 2. Returning to our concrete example, let us suppose that the text is broken into 45,662 ten-syllable lines. We know that 0.3 of all syllable boundaries are internal to words, and thus, because words are sequentially distributed at random, we also know that, probabilistically, 0.3 of the line boundaries will fall within words. Thus we can see that approximately 13,700 words will violate line boundaries. That is to say, if this text were to be versified, or if an author were to attempt to construct verse in similar language, then the composer would be faced with the task of moving 13,700 words away from the line boundary positions. This is the core of the act of lineation, and constitutes the informational ordering that is entailed by that act and registered in our procedure as the Q_n peak. It should be noted that longer lines result in fewer theoretical violations, and therefore result in less work for the author, as do texts with lower proportions of polysyllables. It should also be noted at this point that the size of the Q_n peak is, as might be expected, related to the number of theoretical violations, and thus to both mean word length and line length.

We have now seen that lineation is fundamentally a process whereby polysyllables are placed non-randomly with regard to line boundaries. That is to say, as a text is composed an author will from time to time, for communicative and grammatical reasons, be inclined to place a polysyllable in a line boundary violating position, and we can think of lineation as the act of avoiding such violations. There are two techniques, both of which create Q_n peaks: alternative placement and word length reduction. In addition, when these techniques fail to produce satisfactory results authors will vary from the core metrical line length, producing variant lines, thus reducing the Q_n peak size. We will discuss each of these practices in turn.

3.1.1 *Alternative placement: non-random sequencing within lines*

Rothman has suggested that in pure syllabic metre 'there is a pleasing and challenging strain between numbers of syllables and numbers of words to a line', a remark that is clearly compatible with our theory and observations and, further, that 'at the end of each line word choice becomes highly constrained by syllable count' (1996, p. 207). We take this as suggesting that authors tend to make line-fitting adjustments to their word length towards the end of the line. This point is applicable by extension to other line forms, such as those we are discussing, where line length relation is the outcome of beat and offbeat regulation. We can evaluate this hypothesis by examining the distribution of words of varying numbers of syllables within each line, a question that can best be approached by determining the frequency of spaces following any position in the line. If there is any tendency for words in a region to be longer, then the frequency of spaces in that region will also decline. Figures 5–7 chart data for the three texts listed in Table 1. Figures relating to the core metrical line (dots and continuous line) and the most frequent variant (squares and dotted line) are reported. The horizontal axis n represents syllable

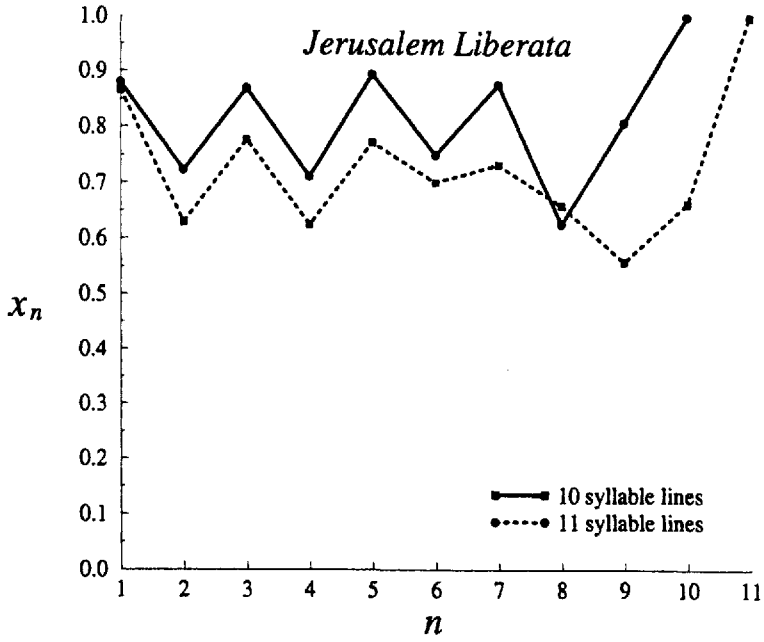

Fig. 5 Edward Fairfax, *Jerusalem Liberata*: distribution of spaces within lines: frequency of spaces following syllable position *n*.

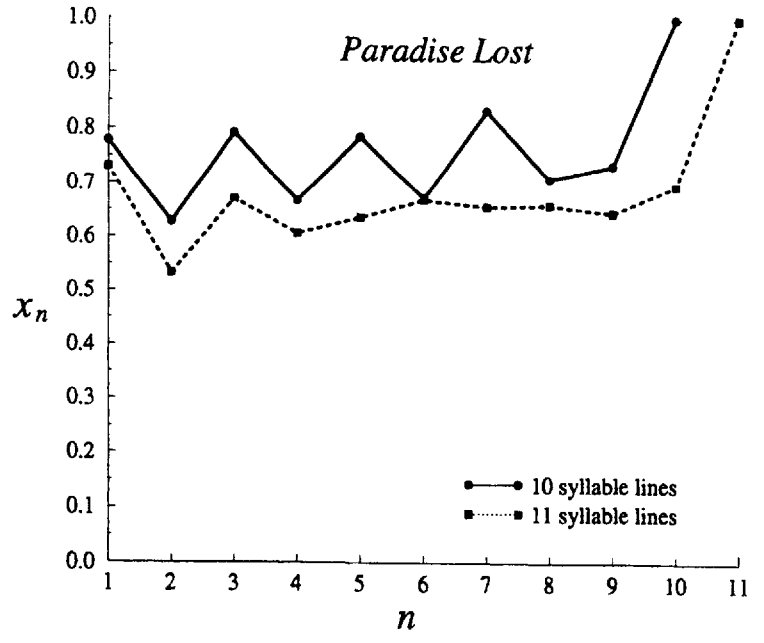

Fig. 6 John Milton, *Paradise Lost*: distribution of spaces within lines; frequency of spaces following syllable position *n*.

positions and the vertical axis x_n represents the normalized frequency of spaces following these positions. There appears to be no tendency for words at the end of the line to be shorter; indeed, if anything, there is a tendency for them to be longer, although the overall distribution is globally even.

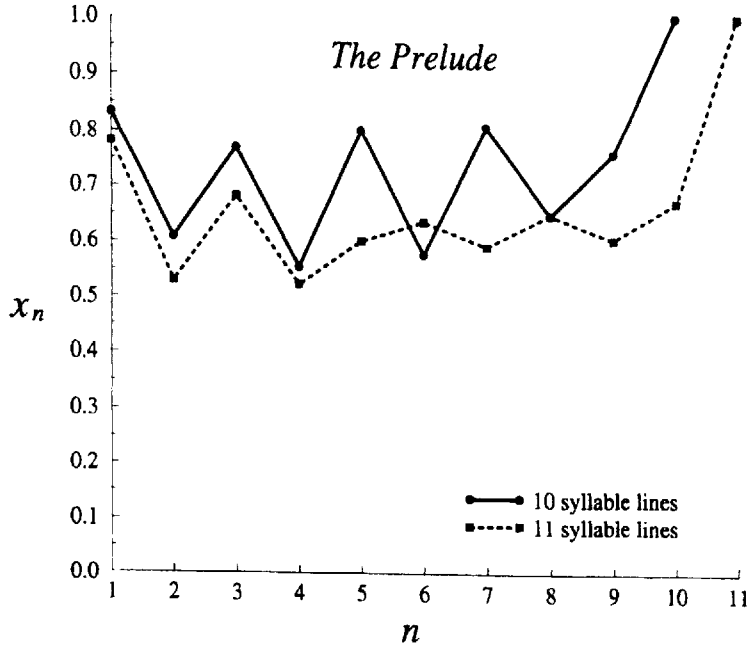

Fig. 7 William Wordsworth, *The Prelude*: distribution of spaces within lines; frequency of spaces following syllable position *n*.

3.1.2 *Word length sequencing within lines: concluding observations and additional remarks*

We have already noted that the overall distribution of word length is relatively even; however, our texts are all by reputable authors, presumably highly skilled in the craft of verse, and it seems possible that less able writers may not achieve such an even distribution of word length throughout the line. Further, Constable (2000) has argued at length that revisions to verse will tend to give evidence of increasingly successful solutions to the problems posed by metrical form, and it would be of great interest to know if revised verse texts tend towards a smoother distribution. In the same paper Constable has already shown that Akenside's revision of *Pleasures of Imagination* uses a much greater variety of structures within lines, and we predict that this will be found to be an indirect indication of this smoother word length distribution.

The space distribution charts (Figs 5–7) also give evidence for two further observations.

- For the most frequent line lengths, the distribution follows an alternating pattern, with enhancements for the odd-numbered positions.
- For the next-to-most-frequent lines, the distribution is stretched toward the end of the line. That is to say that it tends to be flat toward the end of the line, especially after the medial syllable.

The first of these points clearly suggests some degree of non-random ordering of word length, and the second, although peripheral to the main topic under discussion here, is so prominent that it requires explanation.

Alternating pattern. Two facts may be employed to explain the phenomenon of the alternating pattern whereby we see that spaces are much more common after certain positions in the lines (this observation is, of course, valid for only the duple rhythm verse we have examined; triple rhythm may display very different effects, as will duple verse that typically begins its line with beats rather than offbeats). First, disyllables, which make up the bulk of polysyllables in any text, are usually stress initial. In a 17,000-word dictionary compiled by one of us (Constable) from a range of texts in prose and verse, 0.77 of all disyllables, of which there were approximately 7,500, were found to be of this form. Second, in English verse lines stresses in polysyllables are required to coincide almost invariably with one of the beat positions. Consequently, there are more spaces before beat positions than after them. In most of the texts considered the beats occur on the even-numbered syllables, and thus we find a strong alternating pattern of space frequency, with lower frequencies after even-numbered syllables, and higher frequencies before them.

Decay of the alternating pattern in variant lines. A ten-syllable five-beat line is made up of five beats and five offbeats, and can be described thus, where O represents offbeats and B represents beats: OBOBOBOBOB. Each beat is a single syllable, and each offbeat is a single syllable. The O positions are 1, 3, 5, 7, 9, the B positions are 2, 4, 6, 8, 10. (There are some possible variations from this pattern, BOOBOBOBOB for instance, but these, although interesting, are not common enough to detain us here.)

Variant lines that are longer than this pattern will employ one or more double offbeats, that is, offbeats containing two syllables. In the eleven-syllable line this will be a single double offbeat, placed in one of the five possible offbeat positions. When this occurs the beat and offbeat positions subsequent to the double-offbeat are all shifted rightwards by one value. That is, if the double offbeat occurs in the first O position, then the beat positions subsequently become 3, 5, 7, 9, 11, and the offbeat positions 4, 6, 8, 10.

Now, for the purposes of discussion we may assume that the double offbeats in the eleven-syllable lines are evenly distributed over the five positions (in fact, some authors may prefer certain locations to others, but we will ignore this for the time being). Thus we should expect in our sample that 0.2 of all the lines have a double offbeat in the first offbeat position, 0.2 in the second, and so on. Thus, we should expect that 0.2 of the lines are right-shifted subsequent to the first offbeat position, 0.4 of the lines are shifted subsequent to the second offbeat position, 0.6 after the third, 0.8 after the fourth, and all them will be shifted by the line end.

The implications for the alternating pattern are clear. It will be more or less intact for the first few positions, but will decay as the line progresses. In the first positions most of the lines still have beats on the even-numbered positions, but later in the line more of them will be shifted. Indeed, it is reasonable to assume that a shifted alternating pattern will begin to strengthen towards the end of the line, and this appears to be detectable in the charts.

At this point, it becomes reasonable to ask what relationship exists between the x_n and the Q_n distributions. In the following sections we will show that the non-random x_n distributions within lines in fact account for the subsidiary peaks seen in the Q_n distributions.

3.1.3 Q_n *Distribution induced by the* x_n *distribution*

We will first investigate the relation between the space distribution in lines, and the Q_n frequency distribution. Let us assume that the text is made of only one type of core metrical line of length N. If the lines of various pattern types are distributed randomly in this text, the Q_n distribution is expressed as follows, using the x_n distribution:

$$Q_n = \frac{1}{X} \sum_{l=1}^{N} x_l x_{l+n} \tag{4}$$

where X is defined by

$$X = \sum_{n=1}^{N} x_n. \tag{5}$$

In equation (4) any subindex n of x_n should be understood as n (mod N), so that, for example,

$$\begin{aligned} Q_1 &= \frac{1}{X}(x_1x_2 + x_2x_3 + \ldots + x_{N-1}x_N + x_Nx_1) \\ Q_2 &= \frac{1}{X}(x_1x_3 + x_2x_4 + \ldots + x_{N-1}x_1 + x_Nx_2) \\ &\vdots \\ Q_N &= \frac{1}{X}(x_1^2 + x_2^2 + \ldots + x_{N-1}^2 + x_N^2). \end{aligned} \tag{6}$$

Equation (4) is obtained from the following consideration. When a sequence of n-syllables matches the line definition rule, two conditions have to be satisfied: (a) the starting point has to be the beginning of the word (the probability of this happening after the lth syllable in a line is x_l/X); (b) the end of the sequence has to be the end of a word, the probability of which is x_{l+n}. Thus by taking the product of the probabilities of (a) and (b) and summing over l we find equation (4).

The Q_n distribution given by equation (4) has the following properties:

(1) Periodicity: $Q_n = Q_{n+N}$.
(2) Reflection symmetry within each period: $Q_n = Q_{N-n}$.
(3) The average value of Q_n is given by the following:

$$\langle Q \rangle \equiv \frac{1}{N} \sum_{n=1}^{N} Q_n = \frac{X}{N} = \langle x \rangle. \tag{7}$$

This is actually a trivial relationship, for $\langle Q \rangle$ is an overall-average probability of a word boundary occurring at a given place.

(4) The Q_N distribution has *the highest peaks* at $n = N \times$ (integer), as long as the text has only one type of core metrical line of length N. This can be proven as follows. First we note the following inequality:

$$0 \leq \frac{1}{2} \sum_{l=1}^{N} (x_l - x_{l+n})^2 = \sum_{l=1}^{N} x_l{}^2 - \sum_{l=1}^{N} x_l x_{l+n} \tag{8}$$

This implies that

$$Q_N \geq Q_n \tag{9}$$

which means that Q_N has the largest value. The equality in the above holds if and only if all x_l are equal to $x_l + n$. Therefore, if the Q_n distribution has peaks at $n = k \times$ (integer) of the same height as the peaks at $n = N \times$ (integer), it means that $x_N - k = 1$, as $x_N = 1$ by definition. In such a case, we find that the text is actually made of two core metrical lines of length k and length $N - k$. Therefore, given that the text has only one type of core metrical line of length N, and Q_n distribution has peaks at $n = N \times$ (integer), which are the highest among all the peaks.

(5) (Corollary of (4)) Unless the entire text consists of monosyllabic words only, the Q_N distribution has *Peaks* at $n = N \times$ (integer). This can be deduced by extending the above discussion, or from the identity, which can be proven by a straightforward calculation,

$$Q_N - \langle Q \rangle = \frac{1}{X} \sum_{n=1}^{N} (x_n - \langle x \rangle)^2 \tag{10}$$

If Q_N is not a peak, from the property (4) we find that $Q_N = \langle Q \rangle$, which is possible only when all $x_n = \langle x \rangle$ from equation (10). As $x_N = 1$, we conclude then that all words are monosyllabic.

In the following, we will use equation (4) for the actual observed x_n distributions to show that the resulting Q_n distributions match the observed distribution.

The reproduction of the observed Q_n *distribution.* Equation (4) can be used to calculate the theoretical Q_n distribution induced by the x_n distribution. Figure 8 plots the actual Q_n distribution (continuous line) against the induced distribution for the core isometrical line length (dotted line) for Fairfax's *Jerusalem Liberata*. (The results for the other texts are similar in character.) This plot shows a remarkable coincidence between the actual Q_n distribution and the induced distribution. More specifically, we note that the peaks at $n = 10$ are perfectly reproduced by the theoretical distribution.

3.1.4 *Non-random sequencing: syntactic disruption*

We have now shown that there is no evidence of a tendency to use shorter words towards the line end, but that there is a significant tendency to order word length sequencing within lines and that this results from beat patterning. This latter ordering leads to significant subsidiary peaks in

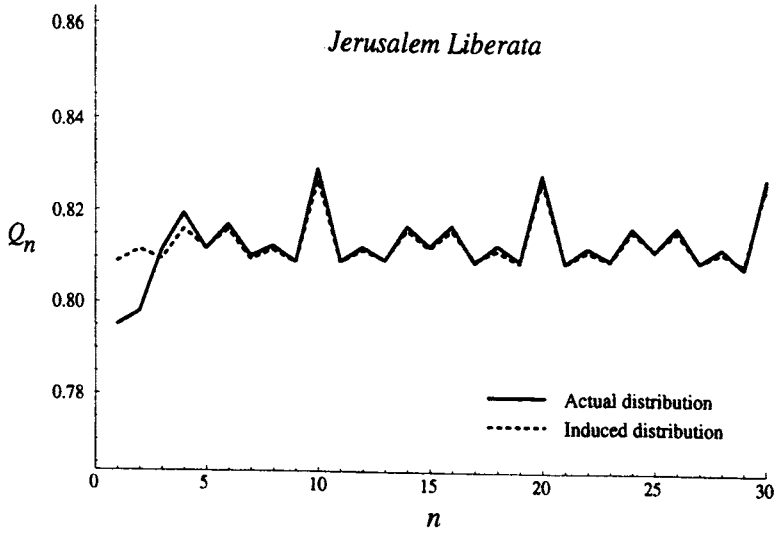

Fig. 8 Edward Fairfax, *Jerusalem Liberata*: the actual and the x_n-induced Q_n distributions.

the Q_n distribution in addition to the major peaks, which are our principal focus of attention. We will now return to these major peaks and consider their importance. Lineation, as has been demonstrated, requires a movement of polysyllables away from the line boundary positions. In other words, the Q_n peak results from a deviation, a non-randomness, in the normal word length sequencing at the line juncture, and any consequent effects within the line are more or less invisible. This is still a real and significant deviation.

Syntax, properly speaking, is the sequential distribution of language features, and, understandably, linguists have concentrated their attention on those features, such as the sequence of semantic category items, that are arrayed in an ordered fashion. That is to say, they have focused on those features with linguistic significance. Nevertheless, the sequential distribution of word length is as real a feature of syntax as any other, although as it is arrayed randomly it is, for most purposes, of little interest. When, however, we come across an exception, as is the case with isometric verse, the word length sequencing of which is non-random, a number of rewarding questions arise. Principally, we are led to wonder to what degree the introduction of order into a normally random aspect of syntax affects other, normally ordered, aspects.

To address this issue we need first to decide whether there is any significant relationship between word length and part of speech. As prose output exhibits no apparent order in the word length sequencing at the global level it is reasonable to assume that there is no strong link. However, as noted in Part I (Aoyama and Constable, 1999, p. 353), there is some deviation from the random sequencing in the fine structure, namely the depression at $n = 2$ in the Q_n frequency charts. From this we can see that a monosyllable is slightly more likely to be followed by a polysyllable than would be predicted from chance, and grammatical syntax seems the likeliest explanation. That is to say that a word class whose mean length is low is more likely than would be expected from

chance to be followed by a class whose length is high. Bearing this in mind it seems probable that detailed examination of the word length characteristics of parts of speech will reveal correlations of interest. However, the effect appears to be weak, and for the time being we will assume that unmetred output presents us with two sequences that are not strongly related: on the one hand, a random sequence of word length values, and, on the other, a highly non-random sequence of parts of speech. Now, if the text is metred and order is introduced into the word length sequence it is clear that there is a *probable* effect on the sequencing of parts of speech, namely, that some degree of randomness will be introduced.

This theoretical consideration is open to empirical investigation, and we will offer a hypothesis in this regard. *If two matched English texts are taken, one in prose, one in verse, the sequential distribution of word class items in verse will be closer to a random sequence than that found in prose.* In other words, given a word at position *n* in a text the frequency totals of word class items in verse will be a better predictor of the class value of preceding and succeeding words than will be the case in prose (obviously, the frequency totals will be a poor predictor in both cases; we merely suggest that it will be significantly less inaccurate in the case of verse).

Thus we may suggest that isometric lineation causes disruption in the sequencing of parts of speech, and, moreover, that this disruption does not merely result in alternatively ordered, and convenient, structures (a 'poetic grammar'), but must constitute a move towards random sequencing. It should be noted that this disruption does not necessarily entail grammatical rule violation or inversion. Neither are necessitated by lineation, although the frequency with which authors resort to them may be increased. The effect to which we are drawing attention is much subtler, and less likely to be salient. Indeed, in carefully written verse it may not be salient at all.

3.2 Theoretical line violation and distortion of the geometric distribution

We have seen that one way of preventing line boundary transgressions in our artificially lineated prose sample is to alter the word order in non-random ways, and that this would probably result in syntactic distortion. Now let us suppose that in fact we are concerned to do as little damage as possible. Assuming that we are already taking as much care as we can in rearranging the syntax, is there anything else that we could do? The answer is yes. We can delete the boundary transgressing polysyllables and replace them with shorter words. That is to say we can introduce more spaces into our text. We can see immediately that this will result in a distortion of the geometric frequency distribution, and that there would be, practically speaking, an increase in monosyllables and a decline in polysyllables in verse. Empirically, this appears to be case. Constable (1997, 1998) has shown that when matched samples of prose and verse are compared the verse has a lower mean word length, and a lower pro-

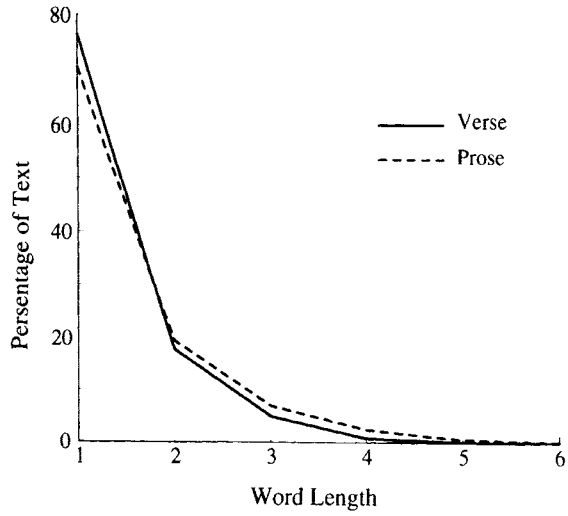

Fig. 9 Word length in George Eliot's *Middlemarch* (prose) and *The Spanish Gypsy* (verse) compared.

portion of polysyllables at every length. Figure 9 charts word length distribution data given in Constable (1998) for George Eliot's *Middlemarch* (dotted line) and her verse narrative *The Spanish Gypsy* (continuous line). The mean word lengths of these two texts are 1.44 and 1.31 syllables per word, respectively.

Given the restrictiveness of lineation this tendency to use shorter words is hardly surprising, as it would be very difficult indeed to maintain an approximation to normal syntax if the mean word length were not reduced. Indeed, the ability to reach a satisfactory compromise with the requirements of lineation by passing neither too much nor too little of the burden of restriction over from syntax and onto diction may be a hallmark of the competent verse writer.

These observations may be brought to bear on the report made earlier by Constable (1997, p. 194), that authors whose mean word length is low in prose tend to reach a much higher proportion of that prose figure in their verse than authors whose mean length in prose is high. It was speculated that the reason for this was that the latter writers were simply more careful of maintaining an approximation to natural syntax. However, it now seems more probable that the reason for this is that writers with low mean word length have fewer polysyllabic word placement problems and are consequently under less pressure to reduce their word length.

3.2.1 *Variation from the core isometrical line length*

As noted above, the impact of line restriction can be further ameliorated by the use of lines varying from the core isometric length, and in fact very few poems, even in the eighteenth century, are completely isometric with regard to syllable count. This is to be expected, as lines are rarely defined as a rigidly fixed number of syllables in English, lineation being the outcome of a regulation of the distribution of stressed and unstressed syllables to realize beats and offbeats, and, as described above, as offbeat

positions can occasionally be filled in different ways a number of different line lengths result. For example, in duple verse each offbeat is usually a single syllable, but can on occasion be two syllables, or even left blank. In practice, it is rare to find more than two such variants in a line, thus in a work with a core length of ten-syllable lines we find a normal maximum of twelve syllables per line, and a minimum of eight, although it is in fact rather rare to find lines of nine syllables. Neither the motivation underlying the use of variant lines nor their empirical frequencies have been hitherto well understood. In the following discussion we will suggest that the primary motivation is not, as one might think intuitively, the avoidance of sonic monotony, but instead the amelioration of line length restrictions on the placement of polysyllables. We will begin by discussing the line length distributions themselves, and then examine these frequencies in relation to our data and theory. Figure 10 gives distribution plots for three of the texts examined earlier, Dryden's translation of *The Aeneid*, Milton's *Paradise Lost*, and Wordsworth's *The Prelude*, all of which have a core metrical set of ten-syllable lines. In these distributions it is evident that when an author employs lines that vary from the core metrical set they tend to employ longer rather than shorter lines. This is entirely consistent with the suggestion that variant lines are employed as a result of an attempt to minimize the disruption of lineation, and that when an author varies from the core set it is usually to employ longer words, with the consequence that variant lines are usually longer lines. Further strengthening can be given to this approach by considering the mean word length of the various line length groups, where it has been found that longer lines tend to have higher mean word lengths (Constable, 1997, pp. 186–7; see also Figs 5–7, where the charts show that variant lines usually have a much lower probability of a space at any position). However, some doubt may still remain as to whether the word length characteristics of the variant lines are the simple outcome of

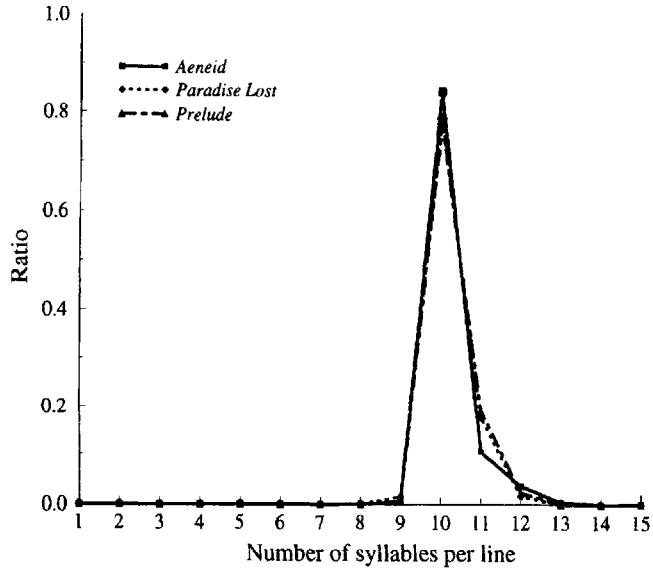

Fig. 10 Line length frequencies for John Dryden, *The Aeneid*, John Milton, *Paradise Lost*, and William Wordsworth, *The Prelude*.

internal rhythmic considerations, for example, the possibility that in order to create an elegant double offbeat authors are more likely to employ a polysyllable. However, consideration of the frequencies of variant lines will suggest that the polysyllabic placement hypothesis is the only plausible hypothesis.

Let us return briefly to the discussion above of the number of theoretical line boundary violations appearing in an arbitrarily segmented text, this time altering our model to incorporate the concept of variant lines. That is to say, let us suppose that we take George Eliot's *Middlemarch* and arbitrarily segment it into ten-syllable lines, *varying up from that length whenever we find our line boundary violating a polysyllable.* Unsurprisingly, as we know that the word length sequence distribution in the overall text is random, the result is a line length profile that coincides almost exactly with the geometric distribution of word lengths, as can be seen if the word length distribution is overlaid on the chart (see Fig. 11). The figures are normalized to facilitate comparison. What we see in such a chart is that the probability of a variant line of a particular length occurring is simply the probability of a polysyllable violating the line boundary position so as to produce a variant of that length. For example, the total number of eleven-syllable lines is the number of disyllables interrupting the line boundary, plus the number of tri-syllables overlapping the line boundary by one syllable, and so on; the total number of twelve-syllable lines is the number of tri-syllables overlapping the line boundary by two syllables, plus the number of tetrasyllables overlapping the line boundary by two syllables, and so on. These frequencies are determined by the geometric distribution.

Now, if we chart the line length frequencies for a representative verse text, comparing it with the word length frequencies of a matched prose sample the correspondence is still remarkably close, although we see small and very important differences between the two distributions. Figure 12 overlays the line length frequencies (continuous line) of

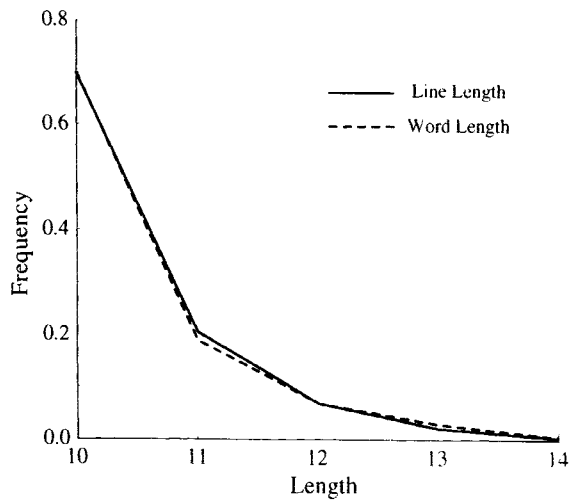

Fig. 11 Line length frequencies in artificially lineated prose, compared with word length frequencies: George Eliot's *Middlemarch*.

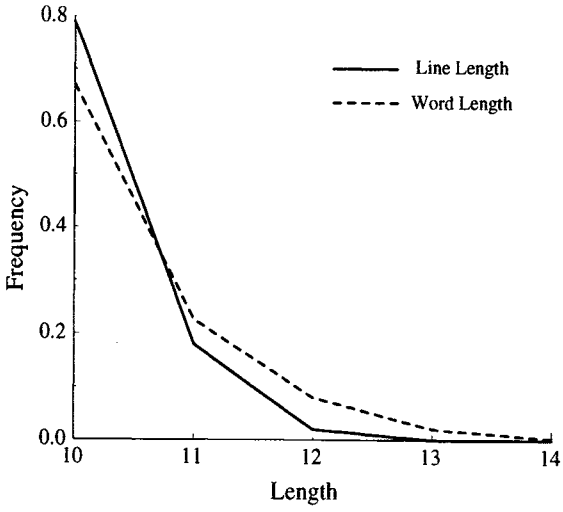

Fig. 12 Line length frequencies in verse, John Milton's *Paradise Lost*, compared with word length frequencies in a matched prose text, *History of Britain*.

Milton's *Paradise Lost* (1667, but begun in the late 1650s) on the word length frequencies (dotted line) of his *History of Britain* (1670, but probably begun in the 1640s and completed in 1655). We may therefore conclude that Milton resorts to variant lines with a frequency that is close to that which we would expect from the number of theoretical line violations arising in his unmetred language. The only plausible explanation for this close coincidence is that his use of variant lines in verse is motivated by polysyllabic word placement.

In the light of these considerations we can see that the more variant lines are employed the less the fact of lineation impinges on the author, and it is in theory possible to use the ratio of actual variants to theoretical violations, discussed above, to judge the degree of flexibility that an author has allowed themselves. For example, in Milton's poem variants are 0.64 of theoretical violations, a figure that we suspect is high. In any case, it seems certain that the variants employed by a composer will hardly ever exceed the number of theoretical violations calculated from a matched prose text (the reader is reminded that these considerations apply to duple verse, and that triple verse may exhibit a different character). As a very rough generalization, we can conclude that as polysyllables are approximately 0.3 of most texts, variant lines are unlikely to approach closely to, and will hardly ever exceed, 0.3 of all lines. It can be further noted that when authors write with a relatively low mean word length, say in the region of 1.2 syllables per word, they will need to employ variant line lengths much less frequently than authors writing with a relatively high mean word length, 1.3 or 1.4 syllables per word; there are simply fewer polysyllabic placement problems confronting such an author. Thus what appear, superficially, to be independent stylistic choices prove to be linked; if an author opts for marked simplicity of language, which generally speaking entails the use of shorter words, their lineation will probably be more regular than if they had chosen to write with higher mean word length.

3.2.2 *Long lines and short lines*

The lines specified by metrical rules come in various lengths, the most common lengths being six syllables, in the short lines of ballad stanzas for example, eight syllables, and ten syllables, and these line lengths are expected to have distinctive Q_n distributions with distinctive levels of disruption. Namely, texts lineated using shorter lines will tend to have larger Q_n peaks than texts lineated using longer lines. The explanation for this is straightforward, and implied by the mathematical account given above, but some brief comments will make its relevance to compositional questions apparent.

It will readily be seen that if lineation requires the composer to vary from the random word-length sequencing of normal output, then a line system that uses shorter lines, with consequently larger Q_n peaks, will tend to have larger consequences on the sequencing of the text. In other words, short lines are more disruptive. Generally speaking, this corresponds to the widespread belief that short lines are less suitable for ambitious content, probably because they are less flexible (Fussell, 1965, p. 450; Tarlinskaia, 1993, pp. 137–8). Our account of the Q_n distribution for long and short lines above, allows us to give a clear quantitative account of this intuition.

If we look back, for a moment, to the calculated theoretical violation it is a simple matter to see that the number of violations is dependent on the line lengths, and that if we examine the relationship between the number of violations, 13,699 for ten-syllable lines and 17,123 for eight-syllable lines, we see that it is given by the relations between the line lengths themselves: 17,123 : 13,699 = 1.25 : 1, and 10 : 8 = 1.25 : 1. Thus we can see that, at least in regard to this measure of the difficulty caused by lineation, an eight-syllable line causes 1.25 times as much work for the writer as a ten-syllable line.

4 Discussion: lineation, poetic effect, and the character of metrical rules

We have now described the mathematical character of lineated text, and given a number of empirical illustrations of the facts of the matter. We have shown that the consequences of lineation for a composer are, first and fundamentally, syntactical, but that to ameliorate this effect an author is likely to use shorter words than they would otherwise have employed (this, it should be noted, is an alternative formulation of an argument given in Constable (1997, 1998)). We have also confirmed that this is true in spite of the use of lines varying from the core isometrical length. Thus we have seen that lineation entails syntactic and encourages dictional disruption.

At this point, we would like to prevent a possible misuse, as we see it, of our position. As the interaction of syntax and lineation has attracted much attention in criticism (Baker, 1967; Leech, 1969, pp. 123–8; Hollander, 1975; Davie, 1976; Ricks, 1984; Bradford, 1988, 1993; Wesling,

1996), it should be recognized that our approach does not address the issue as traditionally framed, a framing that may be represented by Bradford's lucid presentation in his discussion of Milton:

> The reader must decide whether the typographical format of the verse, the 'white space' at the end of each line, is capable of compelling or modifying syntax or whether it is merely a convention of the printed page. (1988, p. 187)

Superficially, our work may seem to assist in answering such a question; we have shown that in one sense lineation does indeed 'compel' or 'modify' syntax. However, whereas our focus has been on the making of text, critics generally concentrate, as Bradford does, on the reading of it. In other words, they are less interested in whether syntax has been distorted during composition than in whether the introduction of line breaks, as visual phenomena, causes, or *should* cause, the reader to experience the syntax in a different way than they would if there were no lineation (Hartman, 1980; Johnson, 1990). When put in this form the question is as relevant to free verse as it is to isometrical verse, and therefore lies beyond the scope of our discussion.

The implications of our findings are numerous, and in concluding we would like to touch on two areas of particular interest, first the question of poetic effect, and second the general character of metrical rules. It should be emphasized that our purpose in these remarks is only to suggest that our findings may assist in opening up these areas for further investigation. Both fields are large and highly problematic, and we cannot hope to do more than sketch connections and possible lines of research.

4.1 Lineation and poetic effect

The importance of disruption of diction for our understanding of poetic effect has been outlined in Constable (1997, pp. 197–8), and described in more detail in Constable (1998), where the argument is put within the framework of theory of relevance (Sperber and Wilson, 1995). In the light of our findings above we can recharacterize that disruption, taking into account syntactic questions, and strengthen the earlier argument. Briefly, it has been suggested that verse form can be regarded as an algorithmic technique for increasing the probability of producing text that simultaneously exhibits intact grammar and ahierarchical implications, and that this can explain the fact that verse form is strongly associated with very rich poetic effects but is neither necessary nor sufficient for their appearance (Buchler, 1974, p. 73).

Communication, it is now widely believed, is not principally a matter of decoding an utterance, but results from a two-stage process of decoding followed by inference construction, the inferences being drawn in accordance with our assumption that a speaker or writer will not require more processing effort of a reader or listener than is merited by the communication (Sperber and Wilson, 1995). A crucial part of this process occurs when an individual selects certain of the implications of an utterance, of which there are very large, even infinite, numbers, and

decides that these were manifestly intended by the composer to be retrieved by a receiver (to distinguish them from other implications these are referred to as implicatures). The means by which the composer manipulates these retrievals are what we know as style (Sperber and Wilson, 1995, pp. 217–24). That is to say, syntactical and dictional choices will tend to structure the implications in certain ways, thus leading the receiver to draw certain conclusions as to which are to be assigned to the category of implicature. Slightly different syntactical and dictional choices will lead to very different conclusions on the part of the reader. Sometimes, often in fact, composers will deliberately arrange for a reader's uncertainty about the strength of an implicature to produce delicate and flexible communicative effects.

The bearing of this on the description of verse given above is straightforward. Verse form forces choice on a number of axes and causes the hierarchy of implications to be ordered randomly to some degree with regard to communicative intent. It will thus, sometimes, be peculiarly difficult to decide which implications are to be assigned to the category of implicature. Such uncertainties do occur in day-to-day circumstances, as for example when we mishear something or misread a word, but, and this is crucial, such cases usually result in extreme incoherence and in grammatical flaws that stimulate the reader either to recover the error or to reject the utterance as irretrievably damaged and undeserving of further interpretative effort. However, the disruptions that occur in high-status verse forms are of a slighter, subtler kind, and are not usually accompanied by grammatical damage, although, as we have shown, lineation alone is sufficient to place grammar under some strain. Readers are thus unable to construct a satisfactory or coherent implicature, but do not abandon the text, and believing, in accordance with the second principle of relevance (according to which every 'act of ostensive communication communicates a presumption of its own optimal relevance' (Sperber and Wilson, 1995, p. 260)), that the author would not put them to unnecessary labour, will conclude that they have not yet expended sufficient effort to produce a clear interpretation. Consequently, they will dig deeper into the hierarchy of implications in search of a still richer resolution. The process is endless, and with every unsuccessful attempt the reader will, instead of abandoning the project, assume still greater but as yet undiscovered rewards.

Here and there in the literature we find writers who have suggested that the disablements of verse might be fundamental to poetic effects, rather than peripheral (Ransom, 1996), but these accounts have not been specific as to the character of the restriction, and their linguistics has not permitted a clear explanation of why restrictions should cause readers so often to find verse, in the poet Ivor Gurney's phrase, 'a collection of words that have inexplicable significance, and give one visions and vistas' (Gurney, 1991, p. 153). We submit that our account of lineation, which is not of course the only retrictive axis in English, goes some way to supply this defect. By encouraging the author to sequence grammatical items in a random way, and to use shorter words than they would otherwise have

used, lineation weakens the hierarchy of implications, and leads to text that appears to offer a plausible but ever-unfulfilled promise of interpretative rewards.

4.2 The general character of metrical rules

The observations in the preceding section put us in a position to offer a remark on the general character of successful metrical rules, a remark that we might have offered earlier but will be seen to be more firmly grounded in this context. As observed by Nowottny (1962, pp. 99–100), there is a considerable difference in reader reaction to patterns determined by 'verse form', on the one hand, and, on the other, reactions to 'verbal schemes', where by verbal schemes is meant 'conspicuous word play [. . .] repetitions and variations involving the central apparatus of meaning'. These latter techniques, Nowottny points out, 'are apparently more open to the objection that they impose formalization on the substance of what is said', whereas verse 'however much it may in fact constrain the poet's freedom to put any words he likes anywhere he likes in the poem, easily allows him to preserve the illusion that he has not wantonly sacrificed the meaning to the successful execution of a move in a word-game'. Nowottny is here describing the well-known phenomenon of reader resistance to particular kinds of manifest patterning, semantic parallelism for example, a valid point and one that requires explanation. What is particularly puzzling is that readers do not put up equal resistance to verse forms, as in fact these forms are as culpably disruptive as verbal schemes. In the light of our approach to the effects of lineation and of the character of rich poetic effects, a solution offers itself. Verbal schemes are the imposition of further extraneous order on features that are already richly ordered in output, with the result that their consequences are extreme and salient. With lineation the case is, as we have demonstrated, very different, order being imposed on features that are randomly arrayed in normal output, and its disruptive consequences being subtle and barely noticeable. We think it possible that the principle we have uncovered here will apply elsewhere in English metrical rules, and may have very wide application to the metrical systems of other languages. Namely, we suggest that the metrical rules that become stable and reliably associated with rich poetic effect in any language will tend to regulate a feature that is randomly arrayed in non-metrical output, and thus will tend to introduce, as a byproduct, a degree of randomness into a feature that is usually ordered. By contrast, those rules that impose further order on previously ordered phenomena will only be regarded as forms of wit, rather than inspiration.

References

Alexander, S. (1933). *Beauty and Other Forms of Value*. London: Macmillan.

Aoyama, H. and Constable, J. (1999). Word length frequency and distribution in English: Part 1. Prose. *Literary and Linguistic Computing*, **14**: 339–58.

Attridge, D. (1982). *The Rhythms of English Poetry*. Harlow: Longman.

Attridge, D. (1995). *Poetic Rhythm: an Introduction.* Cambridge: Cambridge University Press.

Auden, W. H. (1963). Writing. *The Dyer's Hand, and other Essays.* London: Faber.

Baker, D. (ed.) (1996). *Meter in English: a Critical Engagement.* Fayetteville, AR: University of Arkansas Press.

Baker W. E. (1967. *Syntax in English Poetry 1870–1930.* Berkeley and Los Angeles: University of California Press.

Bradford, R. (1988). 'Verse only to the eye': line endings in *Paradise Lost. Essays in Criticism,* **33**: 187–204.

Bradford, R. (1993). *A Linguistic History of English Poetry.* London: Routledge.

Brogan, T. V. F. (1993a). Heterometric. In Preminger, A. and Brogan, T. V. F. (eds), *The New Princeton Encyclopedia of Poetry and Poetics.* Princeton, NJ: Princeton University Press, p. 525.

Brogan, T. V. F. (1993b). Line. In Preminger, A. and Brogan, T. V. F. (eds), *The New Princeton Encyclopedia of Poetry and Poetics.* Princeton, NJ: Princeton University Press, pp. 694–7.

Brogan, T. V. F. (1993c). Verse and prose. In Preminger, A. and Brogan, T. V. F. (eds), *The New Princeton Encyclopedia of Poetry and Poetics.* Princeton, NJ: Princeton University Press, pp. 1346–51.

Brooks, C. and Warren, R. P. (1960). *Understanding Poetry.* New York: Holt.

Brown, D. E. (1991). *Human Universals.* New York: McGraw–Hill.

Buchler, J. (1974). *The Main of Light: on the Concept of Poetry.* Oxford: Oxford University Press.

Burns Cooper, G. (1998). *Mysterious Music: Rhythm and Free Verse.* Stanford, CA: Stanford University Press.

Chatman, S. (1970). The components of English meter. In Freeman, D.C. (ed.) *Linguistics and Literary Style.* New York: Holt, pp. 309–35.

Constable, J. (1997). Verse form: a pilot study in the epidemiology of representations. *Human Nature,* **8**: 171–203.

Constable, J. (1998). The character and future of rich poetic effects. In Sakurai, S. (ed.), *The View from Kyoto: Essays from the Kyoto Poetry Seminar.* Kyoto: Rinsen, pp. 89–108.

Constable, J. (2000). The composition and recomposition of the pleasures of imagination. In Dix, R. (ed.), *Essays on Mark Akenside.* (In press).

Culler, J. (1975). *Structuralist Poetics: Structuralism, Linguistics and the Study of Literature.* London: Routledge.

Darwin, E. (1789). *The Botanic Garden. Part II containing The Loves of the Plants, a Poem with Philosophical Notes.* London: J. Johnson.

Davie, D. (1976). *Articulate Energy: an Inquiry into the Syntax of English Poetry.* London: Routledge & Kegan Paul.

Fabb, N. (1997). *Linguistics and Literature: Language in the Verbal Arts of the World.* Blackwell: Oxford.

Frye, N. (1957). *The Anatomy of Criticism.* Princeton, NJ: Princeton University Press.

Furniss, T. and Bath, M. (1996). *Reading Poetry: an Introduction.* Englewood Cliffs, NJ: Prentice Hall; London: Harvester Wheatsheaf.

Fussell, P. (1965). Line. In Preminger, A. (ed.) *Princeton Encyclopedia of Poetry and Poetics*. Princeton, NJ: Princeton University Press, p. 450.

Gasparov, M. L. (1996). *A History of European Versification*, Smith, G. S. and Tarlinskaja, M. (trans.), Smith, G. S. and Holford-Strevens, L. (eds). Oxford: Clarendon Press.

Geller, S. A. (1993). Hebrew prosody and poetics: biblical. In Preminger, A. and Brogan, T. V. F. (eds), *The New Princeton Encyclopedia of Poetry and Poetics*. Princeton, NJ: Princeton University Press, pp. 509–11.

Gioia, D. (1996). Meter-making arguments. In Baker, D. (ed.), *Meter in English: a Critical Engagement*. Fayetteville, AR: University of Arkansas Press, pp. 75–96.

Gurney, I. (1991). In Thornton, R. K. (ed.), *Collected Letters*. Manchester: Carcanet.

Hamer, E. (1930). *The Metres of English Poetry*. London: Methuen.

Hartman, C. O. (1980). *Free Verse: an Essay on Prosody*. Princeton, NJ: Princeton University Press.

Holder, A. (1995). *Rethinking Meter: a New Approach to the Verse Line*. Lewisburg, PA: Bucknell University Press; London: Associated University Presses.

Hollander, J. (1975). 'Sense variously drawn out': on English enjambment. In *Vision and Resonance: Two Sense of Poetic Form*. New Haven, CT: Yale University Press, pp. 91–116.

Hopkins, G. M. (1959). Poetry and verse. In House, H. and Storey, G. (eds), *The Journals and Papers of Gerard Manley Hopkins*. London: Oxford University Press, pp. 289–90.

Jakobson, R. (1960). Closing statement: linguistics and poetics. In Sebeok, T. (ed.), *Style in Language*. Cambridge, MA: Technology Press MIT; New York: John Wiley.

Johnson, M. L. (1990). Applying the breaks: a review of the poetic line. *Language and Style*, 23: 469–88.

Johnson, S. (1967). In Birkbeck Hill, G. (ed.), *Lives of the English Poets*. New York: Octagon Books.

Kugel, J. L. (1981). *The Idea of Biblical Poetry: Parallelism and its History*. Baltimore, MD: Johns Hopkins University Press.

Kugel, J. L. (1990). Poets and prophets: an overview. In Kugel, J. L. (ed.), *Poetry and Prophecy: the Beginnings of a Literary Tradition*. Uthaca, NY: Cornell University Press.

Leavitt, J. ed. (1997). *Poetry and Prophecy: The Anthropology of Inspiration*. Ann Arbor, MI: The University of Michigan Press.

Leech, G. N. (1969). *A Linguistic Guide to English Poetry*. Harlow: Longman.

Leonard, J. (1996). *The Poetry Handbook: a Guide to Reading Poetry for Pleasure and Practical Criticism*. Oxford: Oxford University Press.

Levin, S. (1971). The conventions of poetry. In Chatman, S. (ed.), *Literary Style: a Symposium*. Oxford: Oxford University Press, pp. 177–96.

Lotspeich, C. M. (1992). 'Poetry, Prose, and Rhythm'. *Publications of the Modern Language Association*, 37: 293–310.

Lotz, J. (1972). Elements of versification. In Wimsatt, W. K. (ed.), *Versification: Major Language Types*. New York: New York University Press, pp. 1–21.

Marks, E. R. (1998). *Taming the Chaos: English Poetic Diction Theory since the Renaissance*. Detroit, MI: Wayne State University Press.

Matthews, B. (1911). *A Study of Versification.* Boston, MA: Houghton Mifflin.

Murry, J. M. (1922). *The Problem of Style.* London: Oxford University Press.

Newton Scott, F. (1904). The most fundamental differentia of poetry and prose. *Publications of the Modern Language Association*, **19**: 250–69.

Nowottny, W. (1962). *The Language Poets Use.* London: Athlone Press.

Ransom, J. C. (1996). Wanted an ontological critic. In Gwynn, R. S. (ed.), *The Advocates of Poetry: a Reader of American Poet-Critics of the Modernist Era.* Fayetteville, AR: University of Arkansas Press, pp. 3–42.

Read, H. (1949). *English Prose Style*, revised edn. London: G. Bell.

Ricks, C. (1970). Distinctions. *Essays in Criticism*, **20**: 259–64.

Ricks, C. (1984). William Wordsworth 1: 'A pure organic pleasure from the lines'. In *The Force of Poetry.* Oxford: Clarendon Press, pp. 89–116.

Rothman, D. J. (1996). Meter and the fortunes of the numerical imagination. In Baker, D. (ed.), *Meter in English: a Critical Engagement.* Fayetteville, AR: University of Arkansas Press, pp. 197–220.

Scott, C. T. (1979). Typography, poems, and the poetic line. In Jazayery, M., Polomé, E. C., and Winter, W. (eds), *Linguistic and Literary Studies in Honor of Archibald A. Hill: IV: Linguistics and Literature/Sociolinguistics and Applied Linguistics.* The Hague: Mouton, pp. 153–60.

Sperber, D. and Wilson, D. (1952). *Relevance: Communication and Cognition*, 2nd edn. Blackwell: Oxford.

Steele, T. (1990). *Missing Measures: Modern Poetry and the Revolt Against Meter.* Fayetteville, AR: University Arkansas Press.

Tarlinskaia, M. (1989). General and particular aspects of meter: literatures, epochs, poets. In Kiparsky, P. and Youmans, G. (eds), *Phonetics and Phonology, Vol. I: Rhythm and Meter.* San Diego, CA: Academic Press, pp. 121–54.

Tarlinskaia, M. (1993). *Strict Stress Meter in English Poetry Compared with German and Russian.* Calgary, Alta: University of Calgary Press.

Turco, L. (1986). *The New Book of Forms: a Handbook of Poetics.* Hanover, NH: University Press of New England.

Wallace, R. (1996). Meter in English. In Baker, D. (ed.), *Meter in English: a Critical Engagement.* Fayetteville, AR: University of Arkansas Press, pp. 3–42.

Weirather, R. (1980). The language of prosody. *Language and Style*, **13**: 120–45.

Wesling, D. (1996). *The Scissors of Metre: Grammetrics and Reading.* Ann Arbor, MI: University of Michigan Press.

Whately, R. (1963). In Ehninger, D. (ed.), *Elements of Rhetoric.* Carbondale, IL: Southern Illinois University Press.

Williams, M. (1986). *Patterns of Poetry: an Encyclopedia of Forms.* Baton Rouge, LA: Louisiana State University Press.

Wimsatt, W. K. and Beardsley, M. C. (1959). The concept of meter: an exercise in abstraction. *Publications of the Modern Language Association*, **74**: 585–98.

Wordsworth, W. (1974). Preface to Lyrical Ballads. In Owen, W. J. B. and Smyser, J. W. (eds), *The Prose Works of William Wordsworth.* Oxford: Clarendon Press.

Yoder, P. (1972). Biblical Hebrew. In Wimsatt, W. K. (ed.), *Versification: Major Language Types. Modern Language Association.* New York: New York University Press, pp. 52–65.