Maxent Grammars for the Metrics of Shakespeare and Milton*

Bruce Hayes Colin Wilson Anne Shisko[1]

To appear in *Language*, December 2012

Abstract

We propose a new approach to metrics based on maxent grammars, which employ weighted constraints and assign a well-formedness value to every metrically distinct line type. We claim two advantages for our approach. First, it offers an explicit account of metricality and metrical complexity, an account that has a principled mathematical basis and integrates information from all aspects of metrical scansion. Second, our approach permits statistical evaluation of proposed constraints. This makes it possible to determine when constraints are vacuous, their work being already done by simpler, independently needed constraints.

We begin by setting up a system built on earlier work that defines the set of possible constraints following principles of stress matching, bracket matching, and contextual salience. Our analyses of two data corpora— Shakespeare's *Sonnets* and Books VIII and IX of Milton's *Paradise Lost*—shows that the basic concepts of this system work well in describing the data. However, one well-known type of constraint, based on the principle of the stress maximum (Halle and Keyser 1966 et seq.), turns out to be vacuous; our testing indicates that the work of stress maximum constraints is better done by other constraints of the grammar.

* For helpful advice we would like to thank Arto Anttila, Sandra Chung, Robert Daland, Nigel Fabb, Paul Kiparsky, Kristin Hanson, Mark Johnson, Stephanie Shih, Rachel Walker, Kristine Yu, Kie Zuraw, and talk audiences at UCLA, Stanford, Stony Brook University, and the annual meeting of the Linguistic Society of America. We also thank three anonymous reviewers for *Language* and the Associate Editor. This research was supported by a grant from the Committee on Research of the UCLA Academic Senate. The article website, containing databases, grammars, tableaux, and our software, is at http://www.linguistics.ucla.edu/people/hayes/ShakespeareAndMilton/.

[1] Author affiliations: Hayes, Shisko: Department of Linguistics, UCLA; Wilson: Department of Cognitive Science, Johns Hopkins University.

1. Introduction

We propose a new approach to analysis in metrics, illustrating our approach with grammars that describe two well-studied bodies of verse, Shakespeare's *Sonnets* and the eighth and ninth books of Milton's *Paradise Lost*. Our work builds on work in the research tradition of generative metrics (Halle and Keyser 1966, 1971 et seq.), as well as our earlier work (Hayes and Wilson 2008) in the use of maxent grammars for the analysis of phonotactics. We suggest that our approach offers the following advantages.

First, maxent grammars provide a fully explicit measure of metrical well-formedness. The measure is gradient, and thus meets a widely-adopted requirement on adequate metrical grammars (Halle and Keyser 1971:142; Kiparsky 1975:580; Youmans 1989; Tarlinskaja 1989:122, 2006). Unlike previous measures, our well-formedness metric rests on sound mathematical foundations, representing essentially the *probability* of a verse line, and unifies information from all parts of the metrical grammar rather than representing just one particular area.

Second, our approach permits controlled comparison of competing analyses of the data. We argue that, when scrutinized under maxent analysis, some of the constraints proposed in the research literature provide no insight because their effects can be reduced to simpler, independently necessary constraints.

As the basis for such model comparison, we first offer a systematization of the constraint system that establishes a space of possible constraints and provides a way of diagnosing when a particular constraint is actually a tacit combination of two simpler constraints. Our system is based on one single principle ((5) below), amplified by various further hypotheses about the contexts in which stress matching is particularly salient. On this basis we address a number of claims that have been made about the metrics of iambic pentameter in the research literature, including the special salience of word-internal stress contrasts, of phrase-final sequences, and of line endings.

On the whole, the earlier hypotheses do well: in our tests the maxent system uses all of them in improving the fit of grammar to data. However, there is one exception: the Stress Maximum Constraint (Halle and Keyser 1966 et seq.), which in various versions has played a prominent role throughout the history of generative metrics, turns out to be vacuous, in the sense that all of the work done by stress maximum constraints can be done independently by simpler or independently-needed constraints. In sum, we believe that our results provide a clearer picture of how English metrics works and offers a more rigorous framework for metrical study.

The article is organized as follows. We give an overview of the issues, then discuss maxent grammars, then present a general theory of English metrics that accommodates most of the constraints in the research literature, ultimately taking the form of a specific constraint set. Next we review our Shakespeare and Milton data corpora and how they were phonologically annotated. We then use maxent and model comparison methods to find optimal grammars for the two corpora. After assessing the sensibleness of these grammars, we discuss their implications for the metrics of our two poets.

2. Metrical well-formedness

Even fairly inexperienced readers of English poetry have some sort of metricality intuitions. For example, it is clear that (1a) below is a very straightforward, simple iambic pentameter (we give its scansion in traditional notation, separating the iambs with slashes). Line (1b) is also an iambic pentameter, but it is felt to be more difficult, a more complex instantiation of the meter. Line (1c) was invented by Halle and Keyser (1971:139) to illustrate the point that one can create lines of the appropriate length that don't sound like iambic pentameters at all; in Halle and Keyser's terms such lines are "unmetrical".

(1) *Metrical, complex, and unmetrical lines*

a. And short / retire- / ment ur- / ges sweet / return. *Milton, Paradise Lost 8.250*
b. So say- / ing, her / rash hand / in e- / vil hour *PL 8.780*
c. Ode to / the West / Wind by / Percy / Bysshe Shel- / ley[2] *(constructed line, HK 1971:139)*

English phonology makes available a vast number of prosodically distinct line types. For this reason, metrical intuitions are unlikely to be based on memorization of types but must result from general principles, which we assume take the form of a metrical grammar. As elsewhere in linguistics, we seek to construct grammars faithful to native speaker intuition by scrutinizing the available data. We also seek to ground our metrical grammars in theoretical principles governing what such grammars can be like. These research goals have long characterized generative research in metrics starting with Halle and Keyser (1966, 1971).

The principles that determine metrical well-formedness for one poet are not necessarily the same as for another: Shakespeare and Milton, for instance, speak different "metrical dialects" in various respects (Tarlinskaja and Teterina 1974; Kiparsky 1975, 1977).[3] Thus, the native intuitions that are our focus of interest are not those of any arbitrary reader of iambic pentameter, but of Shakespeare and Milton themselves. Obviously, this limits the inquiry to what we can learn from scrutiny of their verse, without the privilege of consulting their well-formedness intuitions.

3. A conundrum in the theory of metrics

A salient aspect of the metrics research literature is that when proposals conflict, it is hard to determine which proposal is should be favored. In principle, theoretical questions ought to be settled by appeals to the data: we say an analysis is wrong when it undergenerates (classifies as unmetrical line types that are attested) or overgenerates (predicts to be metrical line types that are systematically missing). The problem is that very few of the constraints proposed in the literature

[2] The eleventh syllable *-ley* is extrametrical, a standard license in iambic pentameter (§5.5.4), and is not the source of unmetricality here.

[3] In addition, the same poet can speak a different dialect when writing in different genres, as in the differences between Shakespeare's poems and his dramatic verse, described in Hanson (2006) and below. In studying fairly uniform verse corpora we will also abstract away from the changes in metrical practice that took place over the course of Shakespeare and Milton's careers (Oras 1960, Bridges 1921, Tarlinskaja 1987).

are exceptionless.[4] In actual practice, metrists have been forced to take the view that a constraint is justified when configurations it forbids are *underrepresented*. In such cases, the analyst claims that exceptions do exist, but are rare enough that we can infer nonetheless that the poet seeks to avoid violating the constraint when composing verse.

We think this approach is sensible—indeed, given the data, there seems to be no alternative. However, once we accept constraints as permitting exceptions, we are in need of a quantitative approach that will allow theoretical claims to be assessed rigorously: how few exceptions should a constraint have in order to be taken seriously? To our knowledge, no answer has yet been given to this question.

The quantitative problem becomes harder when we consider that constraints can interact: if constraint A rules out lines that are also ruled out by an overlapping constraint B, then merely saying that violations of A are rare will not suffice—we need to take B into account in making our evaluation. The problem becomes harder still when, seeking generality, we follow the strategy of setting up whole families of related constraints on the basis of phonetic or structural scales. This is a standard practice in Optimality Theory (Prince and Smolensky 1993, §5.1) and we adopt it below. When the constraints are thus multiplied, it may become necessary to untangle thick knots of overlapping constraints.

In such a situation, it is essentially impossible to make reliable claims about the relative success of different theories unless we have appropriate mathematical tools with which to express and assess them. A primary goal of this article is to argue that maxent grammars are helpful in this respect.[5]

4. Maxent grammars

Maxent grammars are closely related to harmonic grammars (e.g. Legendre et al. 1990, Smolensky and Legendre 2006, Goldwater and Johnson 2003, Boersma and Pater 2008, Potts et al. 2010). They consist of a set of weighted constraints and evaluate candidate forms by calculating the weighted sum of their constraint violations.[6] Within the class of harmonic grammars, maxent grammars are defined by the formula given in (2) below, which assigns probabilities to candidates on the basis of this weighted sum.

In the present application, we are interested in assigning a probability to every metrically distinct way of filling the ten positions of the iambic pentameter. By "metrically distinct" we mean that we ignore the actual lexical items and syntactic form of lines, and consider only the

---

[4] Indeed, there is something of a tradition of "debunking" work in the literature, where scholars offer existing lines by a poet that are predicted by another researcher's analysis to be unmetrical (see, e.g., Magnuson and Ryder 1970, Barnes and Esau 1978, Koelb 1979, Golston 1998, and Tarlinskaja 2006).

[5] The problem we have described is in the theory of metrics, but analogous cases arise in other areas of linguistics. For an application of maxent methods to a comparable problem in phonotactics, where constraints are similarly exceptionful and overlapping in function, see Wilson and Obdeyn (under review).

[6] For an early version of this idea, see Halle and Keyser (1971:176), whose grammar obtains complexity scores by summing the violation counts of two constraints.

factors that are likely to be metrically relevant, namely stress and phonological phrasing.[7] Given the way we describe these elements (see §6 below), there are about 2.15 x 10^{14} distinct ways of filling the slots of a ten-syllable iambic pentameter template. A null grammar would render these equiprobable, but in fact only a small minority would qualify as acceptable iambic pentameters. Our goal is to develop grammars that allocate the bulk of the total probability to these phonological forms.[8]

The probability value assigned to these strings by a maxent grammar is in principle a testable quantity: we predict that if a living clone of William Shakespeare were to compose new sonnets, the phonological forms seen in them would follow a distribution very similar to the one specified in our Shakespeare grammar. In real life, we can submit our work to empirical discipline in several ways. First, it should closely match the statistical distributions in the data corpus. Second, we can evaluate the analysis for consistency across the data, testing whether the patterns attested in some subset of the corpus hold up when tested in a different subset. Lastly, as a kind of sanity check on the analysis we can inspect the probabilities of lines ourselves. When the probability of a line (more precisely, the phonological type of which it is an exemplar) is high, the line should regarded by experienced readers as a fully acceptable, highly unmarked iambic pentameter. When the assigned probability is extremely low (that is, vanishingly close to zero), the line should be considered fully unmetrical. And when the probability is in the middle, the line should be considered marginally metrical, or (to use different terms) highly complex.

The method by which a maxent grammar assigns probabilities to representations is described in detail in Hayes and Wilson (2008). We will not review this in detail here, but offer a brief summary. First, the prosodically relevant aspects of the lines must be coded with explicit phonological representations (see §5.1). Second, the grammar itself consists of a set of constraints, each of which is a function that inputs a scanned line (meter aligned with phonological representation) and outputs a count (perhaps zero) of violations. Each constraint bears a weight (a nonnegative real number), and the probability of a line is computed based on the constraint weights and the violations that the line incurs. This computation works as in (2) (from Della Pietra et al. 1997:1).

(2) *Maxent probability computation*

$$p(\mathrm{L}) = \exp(-\Sigma_i\, \lambda_i \chi_i(\mathrm{L}))/\mathrm{Z}, \text{ where } \mathrm{Z} = \Sigma_j \exp(-\Sigma_i\, \lambda_i \chi_i(\mathrm{L}_j))$$

$p(\mathrm{L})$	predicted probability of line L
$\exp(x)$	e to the power of x

---

[7] Our analyses are oversimplified precisely because they focus on just these traditional elements. A more elaborate coding of the data would have permitted us to examine interesting findings by Kelly (2004) (that onset complexity is metrically relevant in Milton's verse), and by Kelly and Bock (1988) (that nouns pattern in verse differently than verbs). We have also not addressed Keshet's (2005) proposal (motivated by sung-verse data) that word-medial stressless syllables should be counted as less stressed than word-peripheral ones.

[8] Our analysis is an instance of "unconditional maxent", sometimes called density estimation (Duda et al. 2001:9). Conditional maxent — for instance, the analysis of Hopkins's sprung rhythm in Hayes and Moore-Cantwell (2011) — evaluates a set of outputs affiliated with a particular input, much as in Optimality Theory and Harmonic Grammar.

Σ_i summation across all constraints
λ_i weight of the *i*th constraint
$\chi_i(L)$ number of times L violates the *i*th constraint
Σ_j summation over all possible lines

Expressing the same thing in prose: the weight of each constraint is multiplied by the number of times the constraint is violated by the candidate line, and the result across all constraints is summed, yielding a *penalty score*. This score is negated, yielding a value often referred to as the *harmony*. Taking *e* to the power of the harmony yields what we have called (Hayes and Wilson 2008:384) the *maxent value*. Lastly, the maxent value of all possible lines is summed, yielding *Z.* The candidate line's share of Z is its probability. In what follows, we skip most of these steps and report only the penalty scores, which suffice for comparative purposes. In general, the higher the penalty score, the lower the metrical well-formedness of the line is predicted to be.

A consequence of the formula (2) is that the higher the weight a constraint bears, the higher will be the penalty, and thus the lower the probability of lines that violate it. In other words, weight is an intuitive expression of the "strength" of a constraint. Thus maxent grammars are a way of providing an explicit mathematical basis for a widely-held intuition about violable constraints: that violations of them cause a gradient reduction in the overall level of well-formedness; see for example Halle and Keyser (1971:147-164), Lerdahl and Jackendoff (1983), Golston (1998), Keller (2000), and Friedberg (2002, 2006).

The constraints of the grammar represent a theoretical choice (a hypothesis about the theory of metrics) made by the analyst, and much of the discussion below covers our proposed constraint set. The weights are obtained by closest fit to a data corpus (here, our corpora from Shakespeare and Milton), using mathematics developed by Berger et al. (1996), Della Pietra et al. (1997) and others. The fundamental principle used in finding the weights is that of MAXIMUM LIKELIHOOD: the weights are set so as to maximize the predicted probability of all the lines in the corpus. This is a rational criterion, since (as probabilities sum to one) it sets the grammar so as to minimize the predicted probability of data *not* observed, corresponding to the linguist's traditional goal of formulating a maximally restrictive analysis. In calculating the weights proposed here, we used software developed by the second author, an extension of the program described in Hayes and Wilson (2008).

4.1 Excursus: the distinction of absolute metricality

Our choice of a maxent model brings up a traditional debate in metrics. In one view (e.g., Halle and Keyser 1971; Kiparsky 1975, 1977), metrical theory should make a fundamental distinction between metrical and unmetrical lines, which is then supplemented by a complexity metric that makes well-formedness distinctions among the metrical lines. An alternative (Youmans 1989, Golston 1998) is that there is just one single continuum, extending from essentially perfect to fully unmetrical. Since maxent is based on probabilities—and the assignment of zero probabilities is actually incompatible with the formula in (2)—it commits us to the latter view.

We can use the data we gathered for this project to get a clearer view of this issue, focusing on particular constraints. To start, the constraint ALIGN(Line, W), given below as (9h), is a constraint that almost anyone would agree induces a sense of complete unmetricality: it forbids lines that end in the middle of a word, as in the hypothetical line sequence in (3).

(3) A woman's face with nature's own hand paint-
ed hath the master mistress of my passion; *(construct; after Shakespeare, Son. 20)*

This constraint is completely exceptionless in our corpora. In contrast, here is a constraint that almost anyone would agree is no more than a source of metrical complexity: ALIGN(P, Line), given in (10f). This constraint penalizes Phonological Phrase breaks inside the line. It is violated about 7000 times in our corpora, which given the size of these corpora (4434 lines total), means well over once per line. Nevertheless, the constraint plays a useful role in our analysis below: it expresses the (modest) preference of both Shakespeare and Milton for deploying their Phonological Phrase breaks line-peripherally rather than internally. The constraint passes the statistical significance test described below in §7.1.[9]

Between these extremes, essentially *all* values are found: some constraints are violated just one or two times, others a handful of times, others a few dozen, and onward through to hundreds or thousands of violations (see Appendix). Similarly broad ranges are seen in the constraint weights ((43), (48)) and in the scores assigned to lines ((44)). Where should the boundary between "metrical" and "unmetrical" be placed? In the absence of a principled criterion, we consider this to be only a terminological question. We do not see any way to give the concept of "metrical/unmetrical" any definite meaning, other than to set an arbitrary cutoff.

Maxent grammars approximate the notion of "unmetricality" with the notion of extremely low probability. Given high enough weights, a maxent grammar can assign to a representation an extremely low probability; low enough to be a legitimate approximation of unmetricality. When a line is judged to be unmetrical, we see this as the result of the metrical grammar's assigning the line an exceedingly low probability. The one inviolable constraint just mentioned is indeed assigned a very high weight in our system, so that lines violating it receive very low probabilities.

5. A version of the theory of metrics

While the use of maxent grammars for metrics is novel, our actual constraints are mostly traditional ones in generative metrics and are meant to be so. For our purposes it will be important not to adopt a grab bag of constraints from earlier work; we need to have a constraint theory with some internal structure. For this purpose we adopt the following hierarchy of abstraction. At the DESIGN LEVEL, we suggest a general principle that characterizes the task of metrics and how it is to be accomplished. This principle is implemented concretely at the CONSTRAINT LEVEL with families of constraints. Finally, at the GRAMMAR LEVEL, the

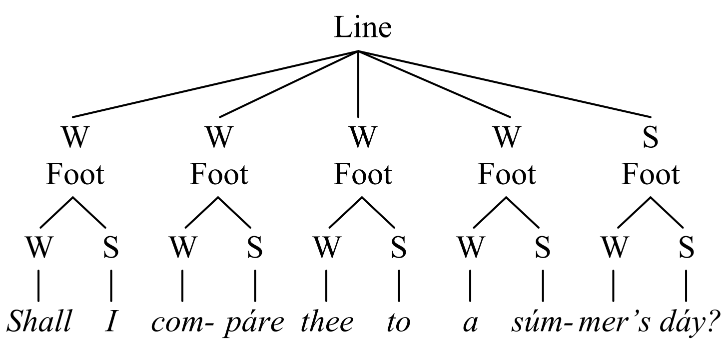
[9] The weight of ALIGN(P, Line) is so low that attempts at directly intuiting its effect may be futile, but here is an example: the very first line of the Sonnets, *From fairest creatures we desire increase,* has (in our codings) a Phonological Phrase break after *creatures*, giving rise to a slight sense of interruption of the line and thus of metrical complexity. This sense is (we judge) lacking in the no-violation lines given in (44a) below.

computational algorithm, referring to poet-specific data, selects from the constraint set and weights the selected constraints, forming an explicit grammar for the poet in question.

5.1 The design level

Following earlier work (Kiparsky 1977, Piera 1980, Prince 1989) we view the meter of a poem as an abstract rhythmic structure. This is sensible, since meters generally have the properties argued to be essential to musical rhythm in Lerdahl and Jackendoff (1983) and similar work. Specifically, the terminal elements of the rhythmic structure are grouped into a hierarchy of abstract constituents having parallel structure, and within each constituent one element is selected as the head; i.e. as the rhythmically Strong one. In the case of the iambic pentameter verse studied here, a simple characterization of the meter is given in (4); here S = metrically Strong and W = Weak. We make the representation more concrete by using it to scan a particular verse line.

(4) *The iambic pentameter template*



Son. 18

The structure given is somewhat simplified: scholars have argued (Piera 1980, Hayes 1988, Youmans 1989, Tarlinskaja 1989:128-129, Duffell 2002) that feet are grouped (either 2+3, or 3+2) into an intermediate Half Line level. Since the effects attributable to the Half Line level in English are subtle and pose computational difficulties,[10] we will dispense with it here.

The poet's task can be stated in the most general terms as (5):

(5) *Key principle of metrics*

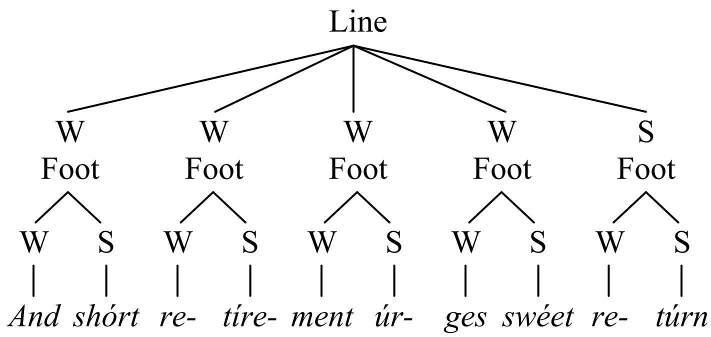
Construct lines whose phonological structure evokes the meter.

English has ample phonological resources that make this evocation possible. It is a stress language, so that poets can manifest the Strong and Weak positions of the meter by matching

[10] The problem is that, as Piera and others have shown, 2+3 and 3+2 structures are commingled in the same poem. In scansion, both variants of the meter have to be tried to find the one that yields the more harmonic outcome. Thus there is circularity in weight-setting; we need to know the right choice of line structures to interpret the data, but that choice itself depends on the setting of the weights. There are ways of attacking this circularity (Tesar and Smolensky 2000), but we postpone this problem to future work; most of English metrics can be done without the half line.

stress to Strong and stressless to Weak. In line (1a), repeated below, this correspondence is perfect:

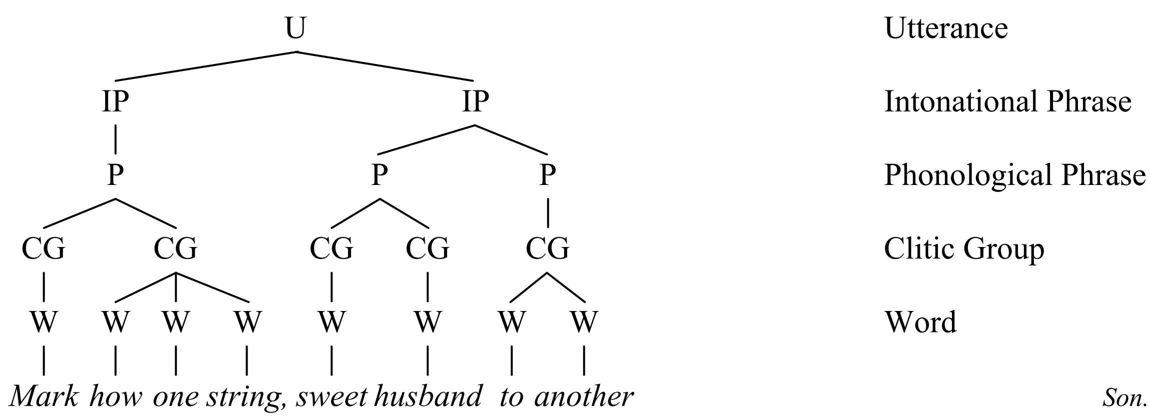
(6) *A stress-perfect line (1a)*



As metrists have often noted, however, only a minority of lines are perfect in this sense (in our data, the figure is about 12%). The characterization of what is allowed as a deviation from this ideal scheme makes the problem more interesting.

English also has the resources needed to manifest the bracketing structure of the meter: the syllables of any line are grouped into words and phrases that can be deployed to match the metrical bracketing. On the phonological side we assume a version of the *Prosodic Hierarchy* theory pioneered by Selkirk (1980) and since elaborated for many languages. We adopt the particular version of the Prosodic Hierarchy, along with the principles of phrasing, proposed in Nespor and Vogel (1986/2007) and further elaborated for English in Hayes (1989). In this version, there are five levels of phrasing, which appear as in (7):

(7) *Example of the Prosodic Hierarchy*



Son. 107

We dispense here with the Utterance for simplicity, since the effects of this highest level of phrasing in metrics are fairly subtle.

As work on various languages has shown (e.g. Kiparsky 1968 (Finnish), Chen 1979 (Chinese), Jakobson 1933, 1952 (Serbo-Croatian), Swiger 1994, Hayes 2008 (Japanese)), poets often use phonological phrasing to manifest grouping in rhythmic structure. They do this by selecting, with greater than chance frequency, lines whose phrasal structure "echoes" the bracketing of the lines, in a specific sense to be defined below.

For English, traditional metrists (e.g. Schipper 1910, §92) have often pointed out the most salient of these phenomena: poets tend to avoid "run-on" lines, which occur when the edges of the Line constituent fail to coincide with a major phrasing break. The existence of "phrasal echoing" for Foot constituents—indeed, the very existence of the foot in English metrics—is more controversial (for skeptics, see e.g. Jespersen 1900; Halle and Keyser 1971:167; Attridge 1982:49). Kiparsky (1977) is notable for the (interestingly indirect) evidence offered in support of foot bracketing, discussed below in §5.5.1. Youmans (1989) buttresses Kiparsky's claim with evidence from word order inversion. Here, we will show that maxent analysis of English metrics indicates modest but very direct "echoing" effects at the foot level.

We turn now from the design level of analysis to the constraint level, where the two basic tasks, manifesting strong-weak relations and manifesting bracketing, are carried out with constraints. Our goal is to create a small but explicit "UM"—universal metrics—consisting of the constraint families posited to be available to poets writing in a stress-based meter such as in English.

5.2 Constraints enforcing bracketing agreement

We begin with a simple case, the constraints that result in the phrasings of the line echoing the bracketing of the meter. These can be expressed using the Generalized Alignment scheme of McCarthy and Prince (1993). For every category of the Prosodic Hierarchy (IP, P, CG and W, as in (7)), we can express an Alignment constraint requiring that it be aligned with some edge of a category—line or foot—in the metrical structure. McCarthy and Prince's Alignment scheme specifies whether it is right edges or left edges that must be aligned. However, in the present instance both meter and phonological phrasing obey the principle of Strict Layering; this means that every constituent abuts only constituents of the same rank, and so right and left alignment amount to the same thing.

Alignment works in two directions: we can either require prosodic categories to have their edges matched with metrical categories, or we can require metrical categories to have their edges matched with prosodic categories. Both kinds of constraints play a role in metrics. For example, ALIGN(Line, IP) says that every Line boundary must be matched by an IP break. This is a rough approximation of the practice of Alexander Pope, whose verse is often noted (e.g. Adams 1997:26-27; Sitter 2007:37) for its avoidance of run-on lines. In contrast, ALIGN(IP, Line) says that every IP break must coincide with a Line break—in effect, this says that lines should not be interrupted with IP breaks. This tendency is less salient but it was also noticed in traditional metrics (e.g. Furnivall and Munro 1910:66-70), where the interrupting breaks were called "central pauses."[11] We will show below that avoidance of such interrupting breaks plays a role in Shakespeare and Milton's verse.

[11] This old term must be used with caution, since phonetic measurement shows that prosodic breaks are often accompanied by pre-boundary lengthening but no actual pausing (see, e.g.. Wightman et al. 1991).

Here are some lines from Milton, listed with the constraints they violate from the ALIGN(Line, X) family. As throughout this article, the reader will find the number of violations for each constraint from the Appendix below; and may also view full tableaux (all lines and constraints) at the website for this article, cited above.

(8)a. *Violates ALIGN(Line, CG), ALIGN(Line, P), and ALIGN(Line, IP)*

But if thou think, trial unsought may find $]_{W}$
$[_{W}$ Us $]_{W}]_{CG}$ both securer then thus warnd thou seemst, *PL 8.370-371*

b. *Violates ALIGN(Line, P) and ALIGN(Line, IP)*

Likeliest she seemd, Pomona when she fled $]_{W}]_{CG}$
$[_{CG}[_{W}$ Vertumnus, $]_{CG}]_{P}]_{IP}$ or to Ceres in her Prime, *PL 8.394-395*

c. *Violates ALIGN(Line, IP)*

Or that, not Mystic, where the Sapient King $]_{W}]_{CG}]_{P}$
$[_{P}[_{CG}[_{W}$ Held dalliance with his fair Egyptian Spouse. *PL 8.442-443*

It can be seen that the Alignment constraints follow a stringency pattern: for example, it is impossible to violate ALIGN(Line, CG) without also violating ALIGN(Line, P) and ALIGN(Line, IP) at the same time. This follows from the principle of Strict Layering: if no CG break is present, then there cannot be a P or IP break, either. Such Strict-Layering-based implicational patterns are discussed by Hayes (1989) and explored further from a maxent perspective below.

Combining prosodic and metrical categories, we obtain the eight Alignment constraints given in (9). Remember that Align(X,Y) is short for 'every X boundary must be aligned with a Y boundary'.

(9) *Constraints on the alignment of metrical categories with prosodic categories*

a. ALIGN(Foot, IP)	e. ALIGN(Line, IP)
b. ALIGN(Foot, P)	f. ALIGN(Line, P)
c. ALIGN(Foot, CG)	g. ALIGN(Line, CG)
d. ALIGN(Foot, W)	h. ALIGN(Line, W)

Constraints that militate against "central pauses" are expressed here as Alignment constraints going in the opposite direction. For example, in the second line of (8b), there is an IP break following *Vertumnus*. Since this break does not coincide with a Line break, this line violates ALIGN (IP, Line). As implied by Strict Layering, it also violates ALIGN (P, Line), ALIGN (CG, Line) and so on. The right edge of *Vertumnus* also falls within the second foot of the line (/ marks foot boundary): *Vertum- / nus*, $]_{CG}]_{P}]_{IP}$ *or / to Ce- / res in / her Prime*. Therefore the line also violates ALIGN (IP, Foot); also ALIGN (P, Foot), etc.

The eight constraints requiring prosodic categories to align with metrical categories are stated together in (10):

(10) *Constraints on the alignment of prosodic categories with metrical categories*

a. ALIGN(IP, Foot)
b. ALIGN(P, Foot)
c. ALIGN(CG, Foot)
d. ALIGN(W, Foot)
e. ALIGN(IP, Line)
f. ALIGN(P, Line)
g. ALIGN(CG, Line)
h. ALIGN(W, Line)

Traditional metrics recognizes a special version of the "central pause" type of constraint: it is especially disfavored to place a large phonological break within the first or last foot, creating the maximally uneven line division of 1 + 9 or 9 + 1 (Bridges 1921:44; Sprott 1953:126). In order to check this hypothesis, we included the following additional alignment constraints in our UM.

(11) *Constraints on grossly uneven line division*

a. *]$_{IP}$ - FOOT$_1$
b. *]$_{P}$ - FOOT$_1$
c. *]$_{CG}$ - FOOT$_1$
d. *]$_{W}$ - FOOT$_1$
e. *]$_{IP}$ - FOOT$_5$
f. *]$_{P}$ - FOOT$_5$
g. *]$_{CG}$ - FOOT$_5$
h. *]$_{W}$ - FOOT$_5$

Below are examples of violations of (11a) and (11e).

(12)a. *Violates (11a)*

Not sedulous by Nature to indite
Wars,]$_{IP}$ hi- / therto / the on- / ly Ar- / gument *PL 8.27-28*

b. *Violates (11e)*

Is lust / in ac- / tion: and / till ac- / tion,]$_{IP}$ lust
Is perjur'd, murderous, bloody, full of blame, *Son. 129*

5.3 Stress and metrical strength: simple constraints

We turn next to stress matching. The simplest possible constraints in this domain simply require that S position must be filled with stress and W position with stressless:

(13) *Simple stress-based constraints*

a. *STRESS IN W — Avoid stressed syllables in W position.
b. *STRESSLESS IN S — Avoid stressless syllables in S position.

The "stress-perfect" line of (6) is an example that violates neither of these constraints. Line (1b), repeated below, has one stressed syllable in W (*rash*), violating (13a), and one stressless syllable in S (*her*), violating (13b).

(14)

W	S	W	S	W	S	W	S	W	S
|	|	|	|	|	|	|	|	|	|
So-	*sáy-*	*ing,*	*her*	*ràsh*	*hánd*	*in*	*è-*	*vil*	*hóur*
			|	|					
			(13b)	(13a)					

The question of what counts as "stressed" for purposes of this constraint was investigated by Kiparsky (1977, §2); we adopt Kiparsky's conclusions with the slight modifications specified by Hayes (1983). Constraints equivalent to (13) were proposed in Halle and Keyser (1971:169) and they also form part of the general parametric theory of metrical constraints proposed by Hanson and Kiparsky (1996).

While very simple constraints like (13) can do quite a bit of work in the analyses, other constraints turn out to be needed as well. Jespersen (1900) may have been the first to suggest that metrical principles should evaluate *sequences* of syllables. A sequence of syllables in which the second has more stress than the first would naturally fit a WS sequence in the meter, and similarly a sequence in which the first has more stress than the second would naturally fit into SW.

(15) *Constraints based on the relative stress of consecutive syllables*

a. *RISE FROM S — Do not rise in stress out of an S position.
b. *FALL FROM W — Do not fall in stress out of a W position.

A line is shown below annotated for its violations of (15a,b):

(16)

W	S	W	S	W	S	W	S	W	S	
|	|	|	|	|	|	|	|	|	|	
Féed'st	*thy*	*lìght's*	*fláme*	*with*	*sélf-*	*sub-*	*stàn-*	*tial*	*fúel,*	*Son. 1*
(15b)		(15a)								

A straightforward extension of the Jespersenian constraints in (15) is to posit versions that forbid two syllables with equal stress in SW or WS:

(17) *Constraints based on the relative stress of consecutive syllables—ties included*

a. *NO FALL FROM S — Assess a violation when stress does not fall out of an S position.
b. *NO RISE FROM W — Assess a violation when stress does not rise out of a W position.

The violations of (15a,b) in (16) are also violations of (17a,b). Constraints (17a,b) are moreover violated by the tied sequences of line (4), shown below in (18) (the scansion assumes that *shall*, *I*, *thee, to*, and *a* are all stressless syllables):

(18)

W	S	W	S	W	S	W	S	W	S
|	|	|	|	|	|	|	|	|	|
Shall	*I*	*com-*	*páre*	*thee*	*to*	*a*	*sùm-*	*mer's*	*dáy?*

(17b) (17a) (17b) (17a)

Another plausible extension would be to suppose that differences involving the contrast of stressed vs. fully stressless syllables are treated as more important than those between stressed syllables with different degrees of stress. Hence the two constraints in (19) are also included in the set.

(19) *Regulating stress matching with stressed and stressless syllables*

a. *Stress Mismatch(− +)	Do not align *stressless* + *stressed* against the meter.
b. *Stress Mismatch(+ −)	Do not align *stressed* + *stressless* against the meter.

For example, a phrase with rising stress like *tàll trées* would violate (15a) if placed in SW position, but it would not violate (19a); whereas *thĕ trées* in SW would violate both.

5.4 Stress and metrical strength: modulations

We posit that the fundamental principle of stress matching is "modulated" in various ways, rendering it stricter in certain contexts. The hypothesis behind such modulations is that particular factors can make stress mismatches more salient, hence more disruptive to the meter. In this section we present two modulations from the research literature and embody them in constraints.

5.4.1 Tight phrasal domains

Metrists have postulated that stress matching is stricter within tighter phrasal domains. Magnuson and Ryder (1970) and Kiparsky (1975) propose that when two consecutive syllables are within the same simplex (noncompound) word, there is particularly strong pressure for them to match the meter. The term *lexical stress* (Kiparsky 1977:194) is used to describe this situation: a stress is a lexical stress if it has more stress than a neighboring syllable within the same simplex word. The following two constraints forbid mismatched lexical stress, in both rising and falling configurations.

(20) *Constraints regulating the matching of lexical stress*

a. *Rise from S(lexical)	Do not rise out of S when the two syllables involved are in the same simplex word.
b. *Fall from W(lexical)	Do not fall out of W when the two syllables involved are in the same simplex word.

A comparison from Kiparsky (1975) illustrates (20a). Shakespeare's sonnets include the real line given in (21a). Kiparsky's rewrite in (21b) replaces phrasally-defined stresses with lexical. It represents a configuration that is very rare in Shakespeare (though not outright unattested: Koelb 1979; Tarlinskaja 2006:58).

(21)

	W	S	W	S	W	S	W	S	W	S	
	|	|	|	|	|	|	|	|	|	|	
a.	*Plúck*	*the*	*kèen*	*téeth*	*from*	*the*	*fierce*	*tì-*	*ger's*	*jáws*	*Son. 19*
b.	*Plúck*	*im-*	*mènse*	*téeth*	*from*	*en-*	*ràged*	*tì-*	*ger's*	*jáws*	*(construct)*
		(20a)				(20a)					

5.4.2 Line-final position

In many metrical traditions it is observed that metrical patterns are reflected more strictly by phonological material at the end of the line (for a listing of cases, see Hayes 1983:373). This has been held to be true for English by traditional metrists; see Bridges (1921:40-41) and Sprott (1953:102). Youmans (1983:76) gives an argument based on marked word order that Shakespeare tries to match stress to the meter more strictly in the last foot of the line (see also section 7.3.3), and the statistical data given by Tarlinskaja (1976:279-280) also support this conclusion. In (22) are given three possible formal implementations of this idea; they are closely similar and will be evaluated against the data below.

(22) *Constraints regulating stress matching within the final foot*

a. *NO RISE FROM W(final foot) — Stress must rise in final foot of the line.
b. *FALL FROM W(final foot) — Stress may not fall in final foot of the line.
c. *STRESSLESS IN S(final foot) — The last S position in the line must be filled with stress.

5.5 Tacit constraint conjunctions?

Repeatedly, analysts in the metrics literature have proposed constraints that from the viewpoint of the framework adopted here are *local conjunctions*, in the sense of Smolensky (1995). Here is one example, from Kiparsky (1977). Kiparsky suggests that in Shakespeare's verse, it is illegal to match a rising stress contour to SW position in the meter whenever this configuration occurs at the end of a prosodic phrase. In all such cases, the right edge of the phrase in question will necessarily fall within the middle of an iambic foot, creating a bracketing mismatch. An example of such a violation is given in (23); the text is by Thomas Wyatt, an early Tudor poet whose "metrical dialect" was rather different from what came to prevail later on.

(23) *A line violating both stress and bracket matching*

For good is **thĕ** **lífe**,]$_{\text{IP}}$ end- ing faith- fully — *"The longe love" 14*

↕ ↕ ↕

[W S]$_{\text{foot}}$ [W S]$_{\text{foot}}$ **[**W S**]**$_{\textbf{foot}}$ [W S]$_{\text{foot}}$ [W S]$_{\text{foot}}$

In the present context, Kiparsky's constraint (1977:206) is a local conjunction of a stress-matching constraint and a bracketing-matching constraint, specifically (19a) *STRESS MISMATCH(− +) and (10a) ALIGN(IP, Foot): the former constraint specifies that *stressless + stressed* (as in *thĕ lífe*) should not be mismatched against the meter, and ALIGN(IP, Foot) specifies that an IP edge should not be placed in foot-medial position. The conjoined character of the constraint was noted explicitly by Kiparsky.

We claim that such local conjunctions should be subjected to careful scrutiny, lest one fall into what one might call the "fallacy of expected values." The problem is this: if violations of a constraint *A are rare, occurring in only (say) 1 percent of all lines, and if violations of *B are equally rare (1%), then a constraint that consists of the local conjunction of A and B would be expected a priori to be violated very rarely indeed: no more than .01 × .01 = .0001, or once in ten thousand lines, simply under the hypothesis of statistical independence.[12] We are not claiming that such conjoined constraints are useless, only that an appropriate statistical test needs to be applied in order to assess them. Note that even when a conjoined constraint has *zero* violations in a large corpus, this does not necessarily indicate that it is metrically meaningful: if the two component constraints are violated quite rarely, then the expected value may be very near zero.

Maxent grammars can in some cases detect invalid conjoined constraints. Suppose, for example, that *A is violated 5000 times in a corpus of 50,000 lines, that *B is violated 500 times, and that there are 50 lines (the statistically expected value, if *A and *B are independent) that violate both *A and *B. If we train a maxent grammar whose constraints are *A, *B, and conjoined *A&B, then the weights will be: *A = 4.59, *B = 2.20, and *A&B = 0.[13] In other words, the weighting system can recognize useless constraints and designate them as such by assigning them zero weight.

The question at hand is an empirical one. Kiparsky's constraint is not *a priori* useless; it may indeed be the case that when a mismatched stress is phrase-final, that makes it particularly salient and thus especially in need of being matched to the meter. In other words, if Kiparsky's constraint is correct, then phrase-final position should be added to the set of modulations (§5.4) for stress-matching constraints. In such a system, lines like (23) would receive a triple penalty: the sum of the weights for *STRESS MISMATCH(− +), ALIGN(IP, Foot), and the Kiparskyan conjoined constraint. Assuming suitable weights, this would predict that violations of Kiparsky's constraint should be exceptionally rare, rarer than we would expect given just the two simple constraints on which it is based.

Maxent weighting can help determine when a conjoined constraint provides added explanatory value. Returning to our *A, *B, *A&B case, we examine what happens when we trim the hypothetical data so that there is just one line that violates both *A and *B—far fewer than the expected value of 50. In this case the *A and *B weights remain essentially the same, but as our calculations indicate, *A&B is also assigned a substantial weight, 3.91. Forms violating *A&B receive the summed penalty of all three constraints, or 10.70, matching their extreme rarity.

---

[12] We actually expect even fewer violations, since the calculation did not assume that the conjunction is necessarily local.

[13] The weights can be found thus: violators of just *A have relative frequency .099, violators of just *B .009, violators of all three constraints .001, and violation-free forms .891. Using these as estimates of probability and applying formula (2), we have $e^{-(Wa)}/Z = .099$, $e^{-(Wb)}/Z = .009$, $e^{-(Wa + Wb + Wa\&b)}/Z = .001$, and $e^{-(0)}/Z = .891$, from which the weights are easily obtained. The numbers will diverge slightly when a Gaussian prior, penalizing high weights, is included in weighting computations (Chen and Rosenfeld 2000). We did use such a prior, but set it so that its effects would be very small; $\sigma^2 = 10^{10}$.

With this in mind, let us survey the cases in which tacitly conjoined constraints have been proposed in the metrics literature.[14]

5.5.1 Conjoined bracketing and stress matching

The constraint described in the previous section is part of a larger family defined by two parameters, which we will now flesh out in order to explore it more carefully. Kiparsky suggests that Milton obeys a constraint forbidding mismatched rising stress of any sort phrase finally (compare constraint (15a)), whereas Shakespeare only forbids rising stress of the form *stressless + stressed* (cf. (19a)). In addition, Kiparsky does not specify which level of prosodic phrasing is intended, so for thoroughness we will render his constraints in multiple versions defined at different phrasal levels. These two dimensions give rise to a six-constraint family, enumerated in (24).

(24) *Constraints conjoining bracketing and stress mismatch*

a.	*RISE FROM S(IP-final)	Do not mismatch rising stress IP-finally.
b.	*RISE FROM S(P-final)	Do not mismatch rising stress P-finally.
c.	*RISE FROM S(CG-final)	Do not mismatch rising stress CG-finally
d.	*RISE FROM S(− +, IP-final)	Do not mismatch *stressless* + *stress* IP-finally.
e.	*RISE FROM S(− +, P-final)	Do not mismatch *stressless* + *stress* P-finally.
f.	*RISE FROM S(− +, CG-final)	Do not mismatch *stressless* + *stress* CG-finally.

Constraints (24a-c) represent local conjunctions of (15a) with three of the bracketing-mismatch constraints in (10) (ALIGN(IP, Foot), ALIGN(P, Foot), ALIGN(CG, Foot)); and (24d-f) represent local conjunction of (19a) with the same three constraints. As before, there are multiple stringency relations within the family.

Here are lines that violate one or more of the above constraints. Note that these lines were selected to illustrate violations and thus are not characteristic lines for either Shakespeare or Milton.

(25) a.	Violation of (24a):	But, like / a sad / slave, $]_{IP}$ stay / and think / of nought[15]	*Son. 57*
b.	Violation of (24b):	Resem- / bling strong / youth $]_{P}$ in / his mid- / dle age,	*Son. 7*
c.	Violation of (24c):	Against / God $]_{CG}$ on- / ly, I a- / gainst God / and thee,[16]	*PL 9.931*
d.	Violation of (24d):	How much / more, $]_{IP}$ if / we pray / him, will / his ear,	*PL 9.1060*

[14] While we have focused on local conjunction as a particularly clear case, this does not exhaust the difficulties at hand. Constraints can overlap in their coverage under many other circumstances, and constraints of significant complexity are likely to have small expected values in corpora of the size studied here. The common theme is that it is necessary to sort out what constraints are doing real work in the grammar.

[15] In (25) and henceforth, for brevity we express only the highest-ranking edge of the Prosodic Hierarchy; due to Strict Layering, all lower edges are always simultaneously present and trigger additional constraint violations.

[16] In (25c) *only I* is scanned as disyllabic by a paraphonological rule of elision; i → j / ___ V, hence [oʊnljaɪ]. For the theory of metrical paraphonology see Kiparsky (1977, §11), and for the specific rules applicable to Milton, see Bridges (1921) and Sprott (1953).

e. Violation of (24e): If thy / soul $]_P$ check / thee that / I come / so near, *Son. 136*
f. Violation of (24f): Hence, thou / suborned / infor- / mer! a / true $]_{CG}$ soul *Son. 125*

Hayes (1989:251-252), discussing the verse of Shelley, proposes a triple tacit conjunction. Shelley allows final phrasal mismatches, violating constraints of (24), and allows mismatched rising lexical stresses, violating (20a), but he avoids violating both at once. This implies the constraint family stated in (26), whose effects in Shakespeare and Milton are examined below.

(26) *Constraints conjoining bracketing, stress mismatch, and lexicality*

a. *RISE FROM S(lexical, IP-final) Do not mismatch rising lexical stress at end of IP.
b. *RISE FROM S(lexical, P-final) Do not mismatch rising lexical stress at end of P-phrase.
c. *RISE FROM S(lexical, CG-final) Do not mismatch rising lexical stress at end of CG.

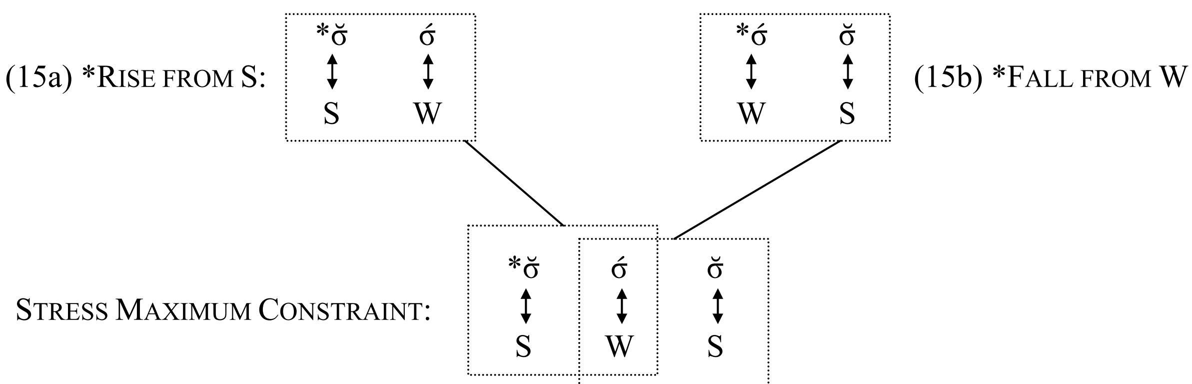
5.5.2 Stress maximum constraints

The first constraint proposed in the generative metrics literature was the Stress Maximum Constraint (Halle and Keyser 1966). In its simplest form, a STRESS MAXIMUM is a stressed syllable that is flanked by two syllables of a weaker degree of stress. A stress maximum constraint forbids stress maxima in W position, as for example in (27).

(27) And dust / shalt eat / all thĕ / dáys ŏf / thy Life. *PL 9.178*

A stress maximum constraint can be regarded as a tacit local conjunction of two Jespersenian constraints of the type given earlier in (15), *RISE FROM S and *FALL FROM W. This is illustrated in (28).

(28) *The Stress Maximum Constraint as local conjunction*



The Stress Maximum Constraint can be fleshed out with a set of variants, just as we have with previous constraints. Halle and Keyser (1971) suggested a less stringent form of the constraint in which a violation is assessed only when the two flanking syllables are stressless, much as the constraints of (15) were restricted in (19); this is annotated in the constraint formulations below with "− + −". In addition, we can impose (as is characteristic in metrics) phrasal conditions on how the stress maximum is defined: the three syllables that define the maximum can be required to be all in the same IP, all in the same P, all in the same CG, or all in the same W. And following Fabb and Halle (2008), we can specify that the stress at the center of the maximum is a lexical stress. We sought to combine these specifications in a sufficiently rich way as to explore much of the logically possible territory, as in (29).

(29) *Versions of the Stress Maximum Constraint*

a. *STRESS MAX IN W
b. *STRESS MAX IN W(IP-bounded)
c. *STRESS MAX IN W(P-bounded)
d. *STRESS MAX IN W(CG-bounded)
e. *STRESS MAX IN W(W-bounded)
f. *STRESS MAX IN W(− + −)
g. *STRESS MAX IN W(− + −, IP-bounded)
h. *STRESS MAX IN W(− + −, P-bounded)
i. *STRESS MAX IN W(− + −, CG-bounded)
j. *STRESS MAX IN W(rising-lexical)
k. *STRESS MAX IN W(falling-lexical)
l. *STRESS MAX IN W(rising-lexical, IP-bounded)
m. *STRESS MAX IN W(falling-lexical, IP-bounded)
n. *STRESS MAX IN W(rising-lexical, P-bounded)
o. *STRESS MAX IN W(falling-lexical, P-bounded)
p. *STRESS MAX IN W(rising-lexical, CG-bounded)
q. *STRESS MAX IN W(falling-lexical, CG-bounded)

For reasons involving our computational implementation of the system, we bifurcated Fabb and Halle's notion of lexical stress: for (29j-q) a lexical stress is falling if the unstressed syllable in the same word follows it and rising if the unstressed syllable in the same word precedes it.

5.5.3 Consecutive S filled by stressless syllables

Sprott (1953) observes that Milton rarely leaves the S positions of consecutive feet unfilled by stress, a pattern noticed for other poets by Bailey (1975:35) and by Duffell (2008). Such an observation is expressible by the constraint given in (30):

(30) *CONSECUTIVE STRESSLESS IN S

This, too, is a form of conjoined constraint, embodying violations of (13b) *STRESSLESS IN S in two consecutive feet. An example violation is given in (31).

(31) *Consecutive S positions filled by stressless syllables*

Undoub- / tedly / he will / relent / and turn *PL 9.1093*

5.5.4 Extrametrical syllables

In our verse corpora, 7.6% of the Shakespeare lines and 3.2% of the Milton lines include an eleventh, so-called "extrametrical" syllable, occurring at the end of the line. We ignore here the constraints that limit these syllables to line-final position[17] and only cover the constraints that penalize their presence or particular configurations.

(32) *Constraints on extrametrical syllables*

a.	*EXTRAMETRICAL	Avoid extrametrical syllables.
b.	*EM WITHOUT FALL	Extrametricals must have less stress than the preceding syllable.
c.	*STRESSED EXTRAMETRICAL	Extrametricals may not bear stress.
d.	*NONLEXICAL EXTRAMETRICAL	Extrametricals must be in same simplex word as the preceding syllable.
e.	*EXTRAMETRICAL(~IP-final)	Extrametricals must be IP-final.
f.	*EXTRAMETRICAL(~P-final)	Extrametricals must be P-final.

Constraint (32a) reflects the general markedness of extrametricals. Constraint (32b) is proposed by Kiparsky (1977:231), and is given a theoretical basis as "beat-splitting" by Prince (1989). Constraint (32c) is violated on occasion in Shakespeare's dramatic verse, but not in the corpora examined here. Constraint (32d) is proposed by Kiparsky (1977:232) as governing Milton's verse but not Shakespeare's. Constraints (32e) and (32f) embody a claim made by Kiparsky (1977:234), that extrametricals cannot occur in run-on lines. Thus they may be regarded for present purposes as tacitly conjoined constraints: (32e) is the local conjunction of *EXTRAMETRICAL and (9e) ALIGN(Line, IP), while (32f) is the local conjunction of *EXTRAMETRICAL and (9f) ALIGN(Line, P). Some sample violations for these constraints are given in (33).

(33) *Sample violations of constraints on extrametrical syllables*

a. *EXTRAMETRICAL	Now is the time that face should form another	*Son. 3*
b. *EM WITHOUT FALL	That cruel Serpent: On me exercise not[18]	*PL 9.927*
c. *STRESSED EXTRAMETRICAL	Quite overcanopied with luscious wóodbìne[19]	
d. *NONLEXICAL EXTRAMETRICAL	Loving offenders thus I will excuse ye:	*Son. 42*
e. *EXTRAMETRICAL(~IP-final)	Of good, how just? of evil, if what is evil $]_P$ Be real, why not known, since easier shunned?	*PL 8.698-699*
f. *EXTRAMETRICAL(~P-final)	For hee who tempts, though in vain, at least asperses $]_{CG}$ The tempted with dishonour foul …	*PL 8.296-297*

In sum, we have located four areas in the research literature where constraints have been proposed that may or may not reflect the "fallacy of expected values". These conjoin bracket matching constraints with stress matching constraints (§5.5.1), two stress matching constraints

[17] In other forms of iambic pentameter, such as Shakespeare's dramatic verse, extrametricals may occur medially before a phrase break.

[18] In Shakespeare and Milton's time, *not* in this syntactic construction was stressless, so there is no violation of (32c).

[19] Shakespeare, *Midsummer Night's Dream* 2.1.251. No examples occur in the corpora.

(§5.5.2), two instances of the constraint against unstressed S (§5.5.3), and the constraint against extrametricals with the constraints against run-on lines (§5.5.4). All of these will be tested below.

5.6 Licensing: the inversion phenomenon

Both traditional metrics and generative work since its beginnings have recognized a "licensing" phenomenon in meter: mismatched falling stress, of types which are otherwise illegal or strongly disfavored, is permitted when the offending stress comes at the beginning of a major phrase.

(34) $[_{\text{IP}}$ Richer / than wealth, $]_{\text{IP}}$ / $[_{\text{IP}}$ prouder / than gar- / ments' cost $]_{\text{IP}}$ *Son. 91*

Often, the phenomenon is called "inversion after a break". Generally, the higher-ranked the phrasal entity, the more able it is to license inversions. It is also suggested by Tarlinskaja (1989:128) that inversions are preferred—all else being equal—when preceded by a line boundary. Both hypotheses must be considered; for instance, Kiparsky (1975) suggests that line boundaries do not actually license inversion; they only *look* like they do because they are normally coincident with major phrase boundaries. We will address this disagreement below.

Licensing of inversion interacts with the lexical status of stresses. Notably, Russian and German verse do not license lexical inversions at all (Bailey 1975:46; Magnuson and Ryder 1970:804; Bjorklund 1978:103-114), but they do license non-lexical inversions after line boundaries or phrasal breaks just as in English.

The licensing phenomenon can be expressed in constraints that require the relevant kind of inversion to be accompanied by its licensing context. We suggest the following inventory of constraints, all of them elaborations on (15) and (20). This does not exhaust the logical possibilities, but permits the testing of a variety of general hypotheses concerning how inversion functions.

(35) *Constraints on inversion*

a.	*FALL FROM W(lexical, ~$[_{\text{IP}}$__)	No lexical inversion unless an IP break precedes.
b.	*FALL FROM W(lexical, ~$[_{\text{P}}$__)	No lexical inversion unless a P break precedes.
c.	*FALL FROM W(lexical, ~$[_{\text{CG}}$__)	No lexical inversion unless a CG break precedes.
d.	*FALL FROM W(CG-level, ~$[_{\text{IP}}$__)	Same as (a), but with Clitic Group-level inversion instead of lexical.
e.	*FALL FROM W(CG-level, ~$[_{\text{P}}$__)	Same as (b), but with Clitic Group-level inversion instead of lexical.
f.	*FALL FROM W(~$[_{\text{IP}}$__)	No inversion of any sort unless an IP break precedes.
g.	*FALL FROM W(~$[_{\text{P}}$__)	No inversion of any sort unless a P break precedes.
h.	*FALL FROM W(~$[_{\text{CG}}$__)	No inversion of any sort unless a CG break precedes.
i.	*FALL FROM W(lexical, ~$[_{\text{LINE}}$__)	No lexical inversion unless line-initial.
j.	*FALL FROM W(~$[_{\text{LINE}}$__)	No inversion of any sort unless line-initial.

Line (34) above incurs one violation of (35i) and (35j), because the inversion on *prouder* is not line-initial; it obeys all of the other constraints of (35) because both of its inversions are arguably IP-initial. Here are lines illustrating violations of other selected constraints above.

(36) a. *Violates (35a, d, f, i , j) (no IP break)*

And peace / proclaims / [$_P$ olives / of end- / less age *Son. 107*

b. *Violates (35d, f, i, j) (no IP break, inversion not lexical)*

But when my glass [$_P$ shows me myself indeed *Son. 62*

c. *Violates every constraint in (35)*

Beyond all past example and [$_W$ future, *PL 9.840*

The rationale for inversion may lie in the "beginnings free, endings strict" principle for metrics alluded to in §5.4.2 above. The idea is that the domains to which the principle is applicable are extended from metrical domains like the line to phonological domains (Hayes 1983, 1989); this has an independent rationale in that the ends of phonological domains are also, it appears, loci of special strictness (§5.5.1).[20]

5.7 Remaining metrical constraints

In what has come so far, we have demonstrated that most of what is proposed in the generative metrics literature[21] falls within a general approach based on the fundamental principle of (5), supplemented by a few additional principles concerning which stress patterns are particularly salient (§5.4, §5.5.1, §5.5.2). One additional family of constraints is not treatable within the proposed approach but deserves attention. Kiparsky (1977:211-213) proposes two constraints that expressed in our notation would appear as in (37).

(37) *Two "puzzle constraints"*

a. *POSTTONIC INVERSION(lexical)	No mismatched falling stress after a weak break when a stressed syllable precedes.
b. *POSTTONIC INVERSION	No mismatched falling lexical stress after a weak break when a stressed syllable precedes.

In (37), "weak break" is taken to be "CG-level or lower". These constraints pertain to the relatively rare cases in Shakespeare and Milton in which an inversion does *not* follow a substantial break. For lexical inversions, these do not occur in Shakespeare at all, it appears, so for Shakespeare the relevant cases are limited to non-lexical inversions. Milton does allow the

---

[20] Another possible rationale for the constraints of (35) is that they represent conjunctions of simpler constraints, specifically, stress-matching constraints together with constraints of the ALIGN(Foot, X) family given in (9a-d). If interpreted thus, the constraints of (35) are not in danger of evaporating in maxent, as in the cases covered previously, because in our grammars (9a-d) turn out to be either feebly weighted or not selected at all; see (48).

[21] Some proposals could not be tested because they are not formalizable with our notation: the "strongest-of-phrase" constraint of Youmans (1983:86), the Bounding Theory proposed for Longfellow's *Hiawatha* by Hayes (1989, 231), and Hammond's (1991, 250) C-command Filter for the verse of James Thompson. As noted earlier, we have also left out constraints based on half-line structure.

occasional lexical inversion not after a break (just a handful in *Paradise Lost*). In both cases, the rare inversion type is largely limited to post-atonic position. An example with violations of both (37a) and (37b) is given below.

(38) More just- / ly, Séat / [$_{CG}$ wórthier /of Gods, / as built *PL 8.100*

This pattern is puzzling: we do not expect to find constraints that ban stress in metrical S position. We explore the validity of these constraints below.

5.8 Phonological constraints

There is no point in using metrical constraints to explain data patterns that follow as a consequence of the phonology of the language in which the verse is written. Thus we have included in our metrical simulations a few phonological constraints, given in (39).

(39) *Phonological constraints*

a.	*STRESSLESS CG	Violated by a Clitic Group with no stressed syllable.
b.	*STRESSLESS P	Violated by a Phonological Phrase with no stressed syllable.
c.	*STRESSLESS IP	Violated by an Intonational Phrase with no stressed syllable.
d.	*EXTENDED LAPSE WITHIN WORD	Violated by three consecutive stressless syllables in the same word.
e.	*WORD-INITIAL LAPSE	Violated by two consecutive stressless syllables at the start of a word.

Of these, (39d) is only rarely violated in English (as in words like *óbstinacy*); (39e) is inviolable (there are no words like *[pətəˈtæmərə]). We have been conservative in including such constraints in our testing; for example, the well-known *CLASH (ban on adjacent stresses) is probably relevant to English stress placement, but would likely obscure the picture for metrics, given that of two clashing stresses at least one must occupy W position.[22]

6. Empirical study: annotated verse corpora

In the previous section, we set forth a general view of English metrics, founded in the basic principle (5). We demonstrated that, with additional assumptions, much of the research literature in metrics can be accommodated within this scheme. These assumptions are that lexical stresses, line-final stress, phrase-final stresses, and stress maxima are salient and thus closely regulated; that extrametrical syllables require end-stopped lines; and that consecutive S positions filled by stressless syllables are metrically disruptive. All of these additional assumptions are vulnerable because the constraints that implement them are "tacitly conjoined" under the theory: structures

[22] We were also conservative in not guaranteeing our phonological constraints a place in the model; in grammar construction (see §7 below) they had to compete on an equal basis with the metrical constraints.

that violate them also violate two or more simpler constraints; these constraints, suitably weighted in a maxent grammar, might turn out to render the conjoined constraints unnecessary.

To investigate these questions, we took two familiar bodies of verse from Shakespeare and Milton's oeuvres: Shakespeare's *Sonnets*[23] and Books VIII and IX of Milton's *Paradise Lost*. These corpora total 2141 and 2293 lines, respectively. Since hand-assignment of the constraint violations would be unreliable and take much time, we instead coded the prosodic structure of the original lines and assessed the violations by machine. Our prosodic codings were done automatically where this was possible (e.g., by assuming IP breaks as a default at punctuation marks), but most of the coding had to be done by hand.

In coding phrasing, we marked for each syllable the highest-ranking prosodic category of which it is the rightmost syllable; intuitively, this marks the degree of "juncture" following the syllable. For stress, we used 1 to mark stressless syllables and 2-4 to mark degrees of stress on stressed syllables. Coding was done partly on the basis of our intuitions as native English speakers, and partly using rules and principles from the phonological literature.[24] For phrasing, we relied on Nespor and Vogel (1986/2007) and Hayes (1989); for phrasal stress, we made use of familiar generalizations from the stress literature: the Nuclear Stress Rule and Compound Stress Rules (Chomsky and Halle 1968) and occasionally the Rhythm Rule (Liberman and Prince 1977) and Beat Addition (Selkirk 1984, Hayes 1995). The prose document we produced and consulted in making our transcriptions, as well as the transcriptions themselves, are posted at the website for this article. We used original spellings and punctuation. All transcription was completed before we began the analytical phase of our study, in hopes of avoiding theory-induced bias.

Two coders (Hayes and Shisko) transcribed all of the lines without consulting each other's work. They achieved reasonably good agreement, as shown in (40).[25]

---

[23] We omitted Sonnet 145 because it is in iambic tetrameter.

[24] We did a scholarly check for words whose scansions suggested that they might have been stressed differently in Shakespeare and Milton's time: we consulted the *Oxford English Dictionary* and also carried out a "concordance test." Here is an example of the latter: examining all instances of how *triumph* is scanned in Milton, one discovers that this word almost certainly bore final stress for Milton when used as a verb, since in verbal use it consistently occurs in WS, not SW position.

[25] In a small number of lines the two transcribers actually found different scansions. This arose where paraphonological rules (see fn. 16) permitted different syllables to be elided. In such cases, the transcribers consulted one another and agreed on a particular scansion. The initial phonological annotations were kept, suitably adjusted for the elided syllables. In principle, the choice among paraphonological options for scanning a line should be made part of the metrical grammar; our use of hand-coding here represents an idealization made for practical purposes.

(40) *Transcriber agreement for stress*

		Hayes			
		1	*2*	*3*	*4*
Shisko	*1*	24821	202	310	57
	2	427	1064	271	62
	3	203	534	7612	1588
	4	144	85	1154	6156

	#	*pct.*
same:	39653	0.887
1 off:	4176	0.093
2 off:	660	0.015
3 off:	201	0.004
total	44690	

(41) *Transcriber agreement for phrasing*

		Hayes				
		Same word	*Word break*	*CG break*	*P break*	*IP break*
Shisko	*Same word*	9483	27	6	3	9
	Word break	50	16408	342	34	73
	CG break	3	385	6456	651	36
	P break	1	157	894	3797	174
	IP break	1	9	20	45	5626

	#	*pct.*
same:	41770	0.935
1 off:	2568	0.057
2 off:	256	0.006
3 off:	86	0.002
4 off:	10	0.000
total	44690	

We conclude that it is possible to assign prosodic annotations of this kind with reasonably good intersubjective agreement. Since the transcriptions are not identical, however, we model them separately below.[26]

[26] A reviewer expressed skepticism concerning codings for Shakespeare and Milton created by modern native speakers. At least for stress, our codings can be defended by a gradient version of the "concordance method" (see fn. 24 and for its origins Tarlinskaja 1967; Gasparov 1980, 8): we take frequency of appearance in S or W position as a rough diagnostic for stress, thus using the poet's own verse as testimony. Following this procedure for monosyllables, we obtained support for our stress transcriptions: in our combined corpora the ten most frequent of the words we normally coded as stressless (*and*, *the*, *to*, *of*, *in*, *I*, *my*, *that*, *with*, *thy*) appeared in W position 76.2% of the time. The ten most frequent words we coded as stressed (*love*, *now*, *self*, *time*, *death*, *sweet*, *heart*, *fair*, *eyes*, *man*) appeared in S position 76.0% of the time. Full data for words in the corpus are posted on the article website.

The coded data were reprocessed slightly prior to modeling. We reduced stress information to one binary distinction (stressed vs. stressless) and one relative distinction (rising/falling/level) defined between consecutive syllables. This procedure discards information (e.g., numerical 434 and 424 were identically recoded) but in a sensible way, since the literature suggests that it is the relative patterning of stress that is relevant to metrics and phonology.

7. Exploring metrical grammars

We can now connect theory to data. Section 5 laid out 87 candidate metrical constraints, intended to embody the fundamental design-level principle of (5) as well as particular ideas proposed in the literature that elaborate this principle. This set of 87 constraints constitutes our universal metrics (UM) for this form of meter, and the constraints of our metrical grammars will be drawn exclusively from it in the manner of Optimality Theory (Prince and Smolensky 1993:2). We also have the data to be explained: when Shakespeare and Milton wrote the verse in our corpora, what metrical principles guided them when they selected the lines they wrote and not other lines?

The basis for our answer is a fundamental hypothesis made by Halle and Keyser (1971:157): "the more complex the line in terms of [the correct analysis], the less frequently it occurs." That is, in constructing lines of verse Shakespeare and Milton intuitively selected lines biased toward less complexity—in maxent terms, toward higher probability. We use this principle in reverse when we seek the grammar that maximizes the predicted probability of the observed data, i.e. the maximum likelihood criterion given in §4.

The *form* of our answer is dictated by the character of maxent grammars. It will consist of (a) a particular selection of constraints from the original set of 87; (b) a particular weighting of these constraints, expressing their relative importance in describing the data.

We anticipate that only a subset of the 87 constraints will be needed to account for the data; this is almost inevitable given the amount of redundancy they involve. In addition, we anticipate that the best grammar for Shakespeare will not be the same as the best grammar for Milton. Rather, we anticipate that inspection of the constraints and weights of the two grammars will reveal metrical differences between the two along the lines noted in the literature.

7.1 The Likelihood Ratio Test

A simple way to use corpus data to select a grammar from the constraints of UM is simply to use the maxent principles to assign a weight to every constraint in UM. The constraints assigned a zero weight (in practice, a fairly large number) can then be regarded as excluded from the grammar. This naïve approach, however, typically includes in the grammar constraints that have very weak effects, or that, though exceptionless, apply to a very small number of lines. A more legitimate criterion is to find grammars all of whose constraints pass a statistical significance test.

We use the likelihood ratio test (e.g. Wasserman 2004:164), comparing grammars that are in a subset relation. We ask whether given a base grammar with constraints C_1, … C_n, a modified grammar containing the additional constraint C_{n+1} produces a statistically significant improvement in the accuracy with which the data are described. To answer this question we compute the test statistic under (42):

(42) *Significance testing with the Likelihood Ratio Test: test statistic*

$$-2 \times \log\left(\frac{\textit{Probability of corpus under simpler grammar}}{\textit{Probability of corpus under full grammar}}\right)$$

The probability distribution of this test statistic can be approximated by a chi-square distribution with one degree of freedom, from which the probability value of the observed difference can be obtained. When constraints fail to reach a reasonable significance value, it is appropriate to exclude them from the grammar even if they bear a positive weight.

In order to apply the likelihood ratio test, it is necessary to choose a significance level. Here, we will use very high one, namely .15. Our reasoning is this: perhaps the important part of our findings is to demonstrate that certain constraint types should *not* be included in the grammar. If they fail to be incorporated into the grammar by our learning system even when a high significance criterion would favor including them, we have more confidence that they do not belong.

The .15 criterion does mean we should be cautious in assessing the constraints that actually do get admitted to the grammar. For the great majority of these, the likelihood ratio test in fact yielded a high significance value (typically, $p < .001$); for the few that emerged otherwise (see (43) and (48) below) we note this, and invite the reader to view these constraints with a suitable degree of skepticism.

7.2 Two methods of grammar search

The likelihood ratio test can only compare grammars in a subset relation. Hence, to explore the hypothesis space of possible grammars, we need to use binary comparisons to find the optimum. The hypothesis space is enormous (2^{87} grammars), so in fact there is no procedure, as far as we know, that is guaranteed to find the best grammar. However, we can come reasonably close by trying two heuristic search procedures and comparing their results.

In a "top down" search, we first use maxent weighting to assign a weight to all 87 constraints of our UM, starting with either the Shakespeare or the Milton corpus as the training set. [27] Some constraints receive zero weights, meaning (as noted above) that they play no explanatory role. We take the surviving constraints and carry out the likelihood ratio test on each one, comparing with the smaller grammar that remains when the constraint is removed. The set of constraints that survive this test are included in the grammar, which is then reweighted.

In a "bottom up" search, we start with a null grammar and gradually add constraints from the UM. At each stage the constraint is selected that has the highest "gain," following Della

[27] All maxent calculations were carried out with software created by Wilson, an evolved version of the phonotactic learner described in Hayes and Wilson (2008).

Pietra et al. (1997). Gain is a statistic that forecasts how much improvement a constraint will make in the predicted probability of the training set if it added to the grammar. At the point where the gain function indicates that no new constraint would pass the likelihood ratio test, no more constraints are added. Since the addition of later constraints can make earlier constraints redundant, the last steps are to trim back the set of constraints using the same method just described for the top-down grammar, then to reweight the completed grammar.

For both top-down and bottom-up grammars, we did one final check: we added back in all of the excluded constraints and did one more likelihood-ratio test, this time with *n* degrees of freedom where *n* constraints were added back in. All of our grammars failed this test, so we retained them in the form derived as given above.

Lastly, we note that constraint selection is slightly stochastic, for reasons described in Hayes and Wilson (2008, §4.2.1); hence different program runs sometimes yield slightly different grammars. We report representative particular grammars from multiple learning runs. However, all runs yielded similar results; in particular, if we treat unselected constraints as having a weight of zero, we find that the median difference in weights between the grammars we report and the mean of a set of five additional learning trials on the same data was 1.6% for Milton and 0.7% for Shakespeare.

In the end, our procedures obtained eight grammars, the result of three binary combinations: choice of verse corpus (Shakespeare or Milton), prosodic transcriber (BH or AS), and mode of constraint selection (top-down or bottom-up). As it turned out, for any given poet, the four grammars learned (two coders, two search procedures) were very similar. To avoid cluttering the exposition below, we chose for the main presentation the grammars whose weights agreed most closely with the mean weights of the grammars as a whole; these were (by a narrow margin) the grammars learned bottom-up for coder BH. The remaining grammars are listed and discussed in §9.

7.3 Representative grammars

The bottom-up grammars for the BH-coded data had 26 constraints for Shakespeare and 30 for Milton, with substantial overlap between the poets. All of the selected constraints are listed in (43); this table also provides the violation counts for these constraints, their weights, and the *p*-values for the likelihood ratio tests. The expression *ns* means that the constraint was not selected for this poet by the gain criterion.

(43) *Grammars for Shakespeare and Milton (BH-coded data, bottom-up search)*

		Shakespeare			*Milton*		
		Viols	*Wght*	*p*	*Viols*	*Wght*	*p*
(9)	d. ALIGN(Foot, W)	8,380	*ns*		9,146	0.15	<.001
	e. ALIGN(Line, IP)	302	2.52	<.001	1,279	0.36	<.001
	g. ALIGN(Line, CG)	7	3.09	<.001	50	2.53	<.001
	h. ALIGN(Line, W)	0	11.40	0.002	0	14.12	<.001
(10)	a. ALIGN(IP, Foot)	195	0.48	<.001	678	*ns*	

	b. Align(P, Foot)	1,146	0.41	<.001	1,469	0.28	<.001
	e. Align(IP, Line)	826	0.85	<.001	2,008	*ns*	
	f. Align(P, Line)	3,175	0.19	<.001	3,943	0.29	<.001
	g. Align(CG, Line)	6,624	*ns*		7,209	0.13	<.001
(11)	a. *]$_{IP}$ - Foot$_1$	18	*ns*		18	1.19	<.001
	b. *]$_P$ - Foot$_1$	50	0.81	<.001	77	0.80	<.001
	c. *]$_{CG}$ - Foot$_1$	211	0.67	<.001	320	0.18	0.010
	e. *]$_{IP}$ - Foot$_5$	4	2.58	<.001	8	2.98	<.001
	f. *]$_P$ - Foot$_5$	243	*ns*		170	0.30	<.001
(13)	a. *Stress in W	1,233	2.94	<.001	1,232	2.95	<.001
(15)	a. *Rise from S	428	*ns*		303	0.42	<.001
(17)	a. *No Fall from S	2,605	0.28	<.001	2,694	0.21	<.001
	b. *No Rise from W		*ns*		2,887	0.11	0.015
(20)	a. *Rise from S(lexical)	3	2.59	<.001	13	*ns*	
(22)	a. *No Rise from W(final foot)	215	1.68	<.001	133	2.28	<.001
(24)	a. *Rise from S(IP-final)	8	*ns*		5	2.01	<.001
	d. *Rise from S(− +, IP-final)	1	1.93	0.005	2	*ns*	
(26)	c. *Rise from S(lexical, CG-final)	1	*ns*		0	12.86	<.001
(30)	*Consecutive Stressless in S	158	1.34	<.001	169	1.27	<.001
(32)	a. *Extrametrical	162	2.61	<.001	73	3.56	<.001
	b. *EM without Fall	0	15.04	<.001	1	3.29	<.001
	c. *Stressed Extrametrical	0	10.17	0.055	0	*ns*	
	d. *Nonlexical Extrametrical	14	2.43	<.001	4	3.04	<.001
(35)	c. *Fall from W(lexical, ~[$_{CG}$__)	1	1.44	0.064	5	*ns*	
	f. *Fall from W(~[$_{IP}$__)	191	0.84	<.001	324	*ns*	
	h. *Fall from W(~[$_{CG}$__)	74	*ns*		55	0.52	<.001
	j. *Fall from W(~[$_{Line}$__)	237	*ns*		239	0.37	<.001
(37)	b. *Posttonic Inversion	11	1.70	<.001	9	2.02	<.001
(39)	a. *Stressless CG	99	3.39	<.001	109	3.22	<.001
	b. *Stressless P	29	*ns*		9	1.44	<.001
	d. *Extended Lapse within Word	0	14.73	<.001	4	2.34	<.001
	e. *Word-Initial Lapse	0	20.81	<.001	0	20.27	<.001

We suggest that these grammars fulfill to a fair degree the original goal set in section 4: they allocate substantial probability to the line types actually written by Shakespeare and Milton. We calculate that the average log probability of a line according to the Shakespeare grammar in (43) is −18.58. This is far higher than would be assigned by a 'null' model in which all possible lines are given equal probability (−30.47). For comparison, the highest possible score, −7.33, would be achieved by a grammar that stipulated that all and only the observed lines are legal. In the case of Milton, the corresponding numbers are −19.16, −30.47 (again) and −6.85.

In the following sections, we further assess the models by considering the scores they assign to particular line types.

7.3.1 Sample outputs

The grammars assign penalty scores (§4) to all lines in the corpus. These express the degree to which the lines violate the constraints and may be interpreted as direct predictions about the metrical well-formedness of these lines.[28] The entire set is posted at the website for this article; in (44) are given representative lines predicted by the grammars to be perfect (or closest to it), some lines of medium complexity, and the lines with maximum complexity in the entire corpus. Material in brackets represents the following context, relevant when there is an Alignment violation.

(44) a. *Perfect or near-perfect lines*

Line	*Penalty score*	
And barren rage of death's eternal cold?	0	*Son. 13*
But not to tell of good or evil luck,	0	*Son. 14*
The dear repose for limbs with travel tir'd;	0	*Son. 27*
The sad account of fore-bemoanèd moan,	0	*Son. 30*
But fondly overcome with female charm.	0.99	*PL 8.999*
And short retirement urges sweet return.	1.12	*PL 8.25*
Though others envie what they cannot give;	1.18	*PL 8.98*
Not yet in horrid Shade or dismal Den,	1.28	*PL 8.185*

b. *Medium complexity lines*

Nor taste, nor smell, desire to be invited	7.49	*Son. 141*
No; let me be obsequious in thy heart,	7.51	*Son. 125*
By adding one thing to my purpose nothing.	7.55	*Son. 20*
The rose looks fair, but fairer we it deem	7.57	*Son. 54*
Food for so foule a Monster, in thy power	7.50	*PL 10.986*
From the Suns Axle; they with labour push'd	7.52	*PL 10.670*
But of the Fruit of this fair Tree amidst	7.59	*PL 9.661*
With Men as Angels without Feminine,	7.51	*PL 10.893*

c. *Maximum complexity lines*

Makes black night beauteous, and her old face new.	15.14	*Son. 27*
Find no determination; then you were [Yourself again]	15.29	*Son. 15*
Suffering my friend for my sake to approve her.	16.50	*Son. 42*
Thou dost love her, because thou know'st I love her;	23.82	*Son. 42*
Tine the slant Lightning, whose thwart flame driv'n down	16.21	*PL 9.1075*
Nor I on my part single, in mee all [Posterity stands curst]	16.28	*PL 9.817*

---

[28] The penalty scores in (44) may be converted to probabilities by applying formula (2), in which the value of log(Z) turned out be 14.49 (S), 14.53 (M). For example, the probability of the first lines of (44a), (44b), and (44c) are respectively, 5.09×10^{-7}, 2.84×10^{-10}, and 1.37×10^{-13}.

There didst not; there let him still Victor sway,	17.89	*PL 9.376*
Like a black mist low creeping, he held on	23.56	*PL 8.18*

For reasons given below (section 12.3), we have not attempted to collect metricality judgments from modern readers, but we judge that these numbers are not too far out of line.

7.3.2 Scores assigned to bad lines

Theorists often put forth examples of unattested line types to illustrate what their theories exclude. Sometimes these are invented examples, sometimes actual lines by poets whose verse composition practice was unusual. We assembled a list of such lines from the literature, coded them phonologically, and computed the penalty scores assigned to them by the Shakespeare grammar of (43).

(45) *Scores assigned to lines judged ill-formed by theorists*

a. If it be betray'd, slander doth approve	Magnuson and Ryder (1970:797)	13.01
b. Introduced grandfather to amuse friends	ibid. 801	23.28
c. Ode to the West Wind by Percy Bysshe Shelley	Halle and Keyser (1971:167) = (1c)	21.20
d. How many bards gild the lapses of time	Keats (Halle and Keyser 1971:171)	14.02
e. Fly away! fly away! you dangerous thing!	Magnuson and Ryder (1971:204)	15.16
f. Inbetween, before, beneath and beyond	ibid. 205	35.51
g. A little conceit? What a dangerous thing!	ibid. 212	12.13
h. For when came poison from so sweet flowers?	Kiparsky (1975:578)	12.69
i. Quite overcanopied with luscious green vines.	ibid. 588	37.54
j. As the pallet whereon it must expire	ibid. 589	9.92
k. With malign weakness benumbs feeling parts	ibid. 591	12.21
l. Pluck immense teeth from enraged tigers' jaws	ibid. 592 = (21b)	16.01
m. To restore to old age what youth hath lost	ibid. 592	11.58
n. And to banish old age where youth hath lost	ibid. 592	10.43
o. And restoring old age where youth hath lost	ibid. 592	13.02
p. As gazelles leap a never-resting brook	ibid. 596	6.40
q. To refuse Virtue in thy nakedness	ibid. 596	8.36
r. For good is the life, ending faithfully	Wyatt (Kiparsky 1977:190, 207)	9.79
s. Fore-advised that from you great Rome shall suck	Kiparsky (1977:199)	15.39
t. Alabaster will not outlast this rhyme	ibid. 200	17.40
u. Unlock this casket, if after three nights	ibid. 200	11.99
v. Put up your bright swords, for under this oath	ibid. 200	16.05

On the whole these results seem reasonable. The main weakness of the grammar appears to be the insufficient penalty assigned to the mismatched iambic words *gazelles* and *refuse* in (45p, q), reflecting an insufficiently high weight (2.59) for (20a), RISE FROM S(lexical). This failing might be remedied with sensible post hoc recoding of the training data. We originally included function words in the purview of (20a) (as in *Weeds <u>among</u> weeds, or flowers with flowers gather'd*, but in fact Kiparsky (1977:218-221) argues that phonologically clitic function words should not be included. If we exclude such lines from the training data, along with one line with the perhaps non-simplex mismatched iambic word *methinks* (Son. 104), the penalty scores for (45p, q) rise to over 20.

For some of the lines in (45) the author provided a control line, intended to be maximally similar while removing the constraint violation at issue. Taking these controls at face value, we can subtract out their scores in assessing the lines of (45). For the nine lines that have a control in the original source, the differences in score were as follows: (45a) 6.66, e 11.52, f 31.35, g 11.54, i 15.32, j 1.44, k 8.68, l 5.18 s 11.47. Of these only (45j) is close in value (1.44) to its control, Sonnet 73, "As the deathbed whereon it must expire". This line represents an unusual type for Shakespeare (mismatched compound word) and plausibly deserves to be penalized almost as harshly as (45j).

7.3.3 The Youmans word order inversion test

Youmans (1982, 1983, 1989) developed an interesting test for metrical complexity based on the hypothesis that when poets use marked word order, their purpose most often is to conform to metrical constraints. For instance, in Sonnet 6 Shakespeare wrote *Make sweet some vial; treasure thou some place* rather than the normal word order *Make some vial sweet; treasure thou some place*; plausibly, this is because in the normal word order the stressed syllables of *vial* and *sweet* create gross metrical violations. Youmans proposes that Shakespeare's marked word orders usually can be explained in this way, and reorderings that increase metrical tension are relatively unusual. We used Youmans's test here as a further way of checking our Shakespeare grammar (43).

We employed a corpus of 169 lines, created by Youmans in his earlier research on the *Sonnets* and representing all of his rewrites of sonnet lines except those in which the word order change affects rhyme.[29] We phonologically coded these rewritten lines using the same method outlined above, retaining where possible the original codings in all portions of lines that were identical in the Shakespeare and Youmans versions. We then computed the scores of the Youmans versions under the grammar in (43) and compared them with the scores for the Shakespeare originals. Of the 169 lines, Shakespeare's marked-word-order version has a lower penalty score in 107, or 63.3%. The Youmans versions had lower penalty scores in only 20 lines (11.8%) and in 42 lines (24.9%) the two word orders resulted in the same score.[30] These results are not far short of the improvement Youmans claimed (1982:77) on the basis of his own examination of the *Sonnets*.[31] The average metrical "improvement" (reduced penalty score) that Shakespeare obtained by using marked word order in the lines we examined was 3.73.

For the handful of lines where Shakespeare's use of marked word order makes the line substantially more complex, we would agree with Youmans that the explanation is likely to be rhetorical improvement (e.g., in focus marking), as in for example the focused *this* of Son. 18 *So long lives this, and this gives life to thee* (vs. *This lives so long*…).

---

[29] We are grateful to Prof. Gilbert Youmans for sharing this corpus with us.

[30] Of these, 26, or 15.4%, had identical codings in the two versions and thus would receive identical scores under any grammar.

[31] It emerged that just a few constraints were responsible; they are (with percentage of the total effect given): (13a) *STRESS IN W (48.5%), (30) *CONSECUTIVE STRESSLESS IN S (16.5%), (35f) *FALL FROM W(~[$_{IP}$__) (12.2%), (35c) *FALL FROM W(lexical, ~[$_{CG}$__) (6.7%), (20a) *RISE FROM S(lexical) (6.6%), , (17a) *NO FALL FROM S (5.1%), (24d) *RISE FROM S(− +, IP-final) (2.3%), and (37b) *POSTTONIC INVERSION (2.0%).

7.3.4 Evaluating the generality of the grammars

Because the grammars in (43) were trained on the entire verse corpora, one potential concern is that they have *overfit* the data (on the general problem of overfitting, see for example Duda et al. 2001:5). The grammars might have learned relatively accidental or unsystematic properties rather than capturing only significant metrical generalizations.

To address the issue of overfitting, we adopted the commonly used method of *k*-fold cross-validation (e.g., Duda et al. 2001: 483ff.) as follows. We first divided the corpus for each poet into $k = 10$ roughly equal portions, or *folds*. We then retrained the weights of the grammars *k* times, each time using $k - 1$ of the folds as the basis of training and reserving the remaining fold for testing. If the weighting procedure leads to substantial overfitting, the penalty scores of lines in the training folds should be significantly lower than those of the 'unseen' lines in the corresponding testing fold. We assessed this prediction with a two-sample t-test. For neither of the poets did the test come out significant, even under an extremely nonstringent significance criterion ($p > .3$ in each case). The test provides evidence that the grammars embody general and systematic properties of the poets' metrical practices.

A further reason to think that our grammars represent systematic metrical practices of Shakespeare and Milton is that each poet's grammar yields poor results when applied to the verse of the other. For each poet separately, we evaluated the ten non-overlapping folds with both of the grammars in (43) and averaged the penalty scores assigned to the lines in each fold by each grammar. As expected, a poet's grammar assigns much lower penalty scores to his own lines than does the other poet's grammar ($p < .001$ for both poets by paired two-sample *t*-tests).

8. Interpreting the grammars

We now give a sketch of what this method tells us about the metrics of Shakespeare and Milton, following the expository order given earlier in §5. Some of our results confirm well established observations of traditional metrics, acting as a check on the analysis. Others bear on proposals made in the generative metrics literature, and some results are new.

Our 87-constraint UM was organized into families of constraints from which the particular constraints of our grammars were selected. It is not always feasible to establish intuitively what aspect of the data caused a particular member of a constraint family to be selected. Therefore, in what follows we will describe the grammars in general terms, focusing on the principles of metrics that are implemented in the selected constraints.

8.1 Bracketing matches

(a) As we would expect, the grammars penalize lines that fail to end with a phrase break; and the less salient the phrase break, the worse the degree of violation. This can be seen in the weights for (9e,g,h), given above in (43). For example, if Milton ends a line in a P-break (but not an IP-break), this will incur only the penalty for (9e) ALIGN(Line, IP), i.e. .36. If a Milton line ends only in a word break, it would violate (9g) ALIGN(Line, CG) as well and incur the summed penalty of .36 + 2.53 = 2.89. A hypothetical Miltonic line that failed to end even in a word break would receive the summed penalty of all three constraints, .36 + 2.53 + 14.12 = 17.01.[32] A similar pattern holds for Shakespeare.

(b) Both poets show a preference not to place phonological breaks in the middle of the line, confirming the "central pause" idea of traditional metrics. Shakespeare shows a two-constraint additive system ((10e) ALIGN(IP, Line) .85, (10f) ALIGN(P, Line)) .19, such that placing an IP break in the middle of the line is worse than placing a P-break; Milton's grammar is similar but in this case the two constraints are ALIGN(P, Line) .29 and ALIGN(CG, Line) .13.

(c) A widely noted aspect of Milton's verse (e.g. Andrews 1918:216-221, Sprott 1953:112, Steele 1999:107-108) is his exceptionally frequent use of run-on lines and central pauses. In our grammars, this effect appears to be reflected most at the IP level. There are substantially lower weights for Milton than Shakespeare for both ALIGN(Line, IP) (militating against run-ons; S 2.52, M 0.36) and ALIGN(Line, IP) (militating against central pauses; S 0.85, M not selected).

(d) Kiparsky's (1977) suggestion that poets tend to echo Foot boundaries with phonological breaks is also confirmed by the weights given to (10a) ALIGN(IP, Foot) (S: .48), (10b) ALIGN(P, Foot) (S: .41, M: .28), and ALIGN(Foot, P) (M: .15) . Kiparsky's original arguments are indirect, based on their role of such bracket matching in conjoined constraints (§5.5.1), as well as the intuitions of critics. Our results provide new support for Kiparsky's position: the constraints show a direct foot-matching effect.

(e) Lastly, we tested the traditional view that placing large phrase breaks in a way that divides the line 1+9 or 9+1 is disfavored. This view is supported by the weights given to constraints of family (11); see (43) for details.

We offer the following intuitive illustration of the effect of bracketing mismatch constraints. The famous opening quatrain of Shakespeare's Sonnet 129, which bristles with run-on and central-pause violations, accumulates a summed penalty score (bracketing constraints only) of 23.2, the highest of any quatrain in the *Sonnets*. At the opposite extreme is the opening quatrain of Sonnet 34, which incurs a summed bracketing penalty of only 0.37.

(46) a. *Sonnet 129, lines 1-4 (23.2)*

The expense of spirit in a waste of shame
Is lust in action: and till action, lust
Is perjur'd, murderous, bloody, full of blame,
Savage, extreme, rude, cruel, not to trust;

---

[32] We were somewhat surprised not to see ALIGN(Line, P) among the other corresponding alignment constraints. It is possible that this reflects inaccuracies in transcribing P breaks in our corpora; in retrospect we judge that we may have relied too much on the criterion of aligning P breaks at right XP edges (after Selkirk 1986), and might have done better with a system more sensitive to P-phrase length.

b. *Sonnet 34, lines 1-4 (0.37)*

Why didst thou promise such a beauteous day,
And make me travel forth without my cloak,
To let base clouds o'ertake me in my way,
Hiding thy bravery in their rotten smoke?

8.2 Stress matching

The core stress-matching constraint turned out to be (13a) *STRESS IN W, which bears a substantial weight for both poets (S: 2.94, M: 2.95). This constraint is assisted by minor poet-specific constraints of the kind proposed by Jespersen: (15a) *RISE FROM S (M: .42), (17a) *NO FALL FROM S (S: .28, M: .21), and (17b) *NO RISE FROM W (M: .11). A consequence deducible from these constraints is that for either poet a "perfect" line must fill all S positions with stress, which seems intuitively correct; see (44a) for examples.

The principle that LEXICAL STRESS is salient is affirmed by the weights assigned three constraints, all poet-specific: (20a) *RISE FROM S(lexical) (S: 2.59), (26c) *RISE FROM S(lexical, CG-final) (M: 12.86), and (35c) *FALL FROM W(lexical, ~[$_{CG}$__) (S: 1.44).[33]

The principle that the LAST FOOT OF THE LINE is regulated with special strictness is confirmed by the substantial weights given to (22a) *NO RISE FROM W(final foot) (S: 1.68, M: 2.28).

The tacitly-conjoined constraints proposed by Kiparsky (§5.5.1) that forbid SIMULTANEOUS STRESS MISMATCH AND BRACKETING MISMATCH were selected for both grammars. The Milton constraint is (24a), *RISE FROM S(IP-final) (2.01), forbidding any mismatched rising sequence whose right phrase edge is foot-medial. The Shakespeare constraint ((24d), *RISE FROM S(− +, IP-final), 1.93) is analogous but forbids only stressless-stressed sequences. We conclude that the "fallacy of expected values" (§5.5) is *not* manifested by Kiparsky's original proposal. A violation of either (24a) or (24d) imposes a penalty on top of the pure-bracketing constraints ((10a,b)) and pure stress-mismatch constraints ((17a), (20a)) that are concomitantly violated.

The grammars also support Kiparsky's contention that there is a "dialect difference" between Shakespeare and Milton in this area: Shakespeare has *RISE FROM S(− +, IP-final), Milton the more general *RISE FROM S(IP-final), just as Kiparsky claimed. A caution is that this result was not obtained for all the grammars we constructed; see §9 below.

The question of what LICENSES INVERSIONS—prosodic structure or line structure (§5.6)—is resolved in favor of allowing both. For Shakespeare, both relevant constraints have a prosodic

---

[33] That *RISE FROM S(lexical) was not selected for Milton probably reflects a shortcoming already noted in our coding procedure; almost all the violations of this constraint in the Milton material involve grammatical words like *above*, which show special behavior. We can check this by trimming the training data so it include no lines with mismatched grammatical categories. In this case *RISE FROM S(lexical) is selected for Milton and given the substantial weight of 3.07.

environment ((35c) *FALL FROM W(LEXICAL, ~[$_{CG}$__ 1.44, (35f), *FALL FROM W(~[$_{IP}$__ 0.84), whereas Milton invokes a blend, with a penalty for inversions that are not CG-initial ((35h) *FALL FROM W(~[$_{CG}$__), .52) and for lexical inversions that are not Line-initial ((35j) *FALL FROM W(~[$_{Line}$__), .37).

8.3 Extrametrical syllables

The grammars of (43) confirm in broad outline the account of extrametrical syllables given in §5.5.4: extrametrical syllables contribute substantially to the metrical complexity of a line ((32a) *EXTRAMETRICAL, S 2.61, M 3.56); they virtually always require a fall in stress ((32b) *EM WITHOUT FALL, S: 15.04, M: 3.29), and it is strongly preferred that they occur within the same simplex word as the preceding syllable ((32d) *NONLEXICAL EXTRAMETRICAL, S 2.43, M 3.04). The constraint (32c) *STRESSED EXTRAMETRICAL was selected with marginal statistical significance (but high weight, 10.17) for Shakespeare; evidently its effects are largely deducible from (32b) and (32d).

Kiparsky's proposal that extrametricals are incompatible with run-on lines (see §5.5.4) is not supported by the grammars; neither one includes (32e) *EXTRAMETRICAL(~IP-final) or (32f) *EXTRAMETRICAL(~IP-final). Thus from the viewpoint of the grammars under discussion, this would be a case of the fallacy of expected values. However, (32e) was selected as significant in other grammars, discussed below in §9; hence the evidence from the grammars is equivocal.

8.4 Consecutive S filled by stressless syllables

Both poets evidently avoid consecutive S positions filled by stressless syllables; this is supported by the weight assigned to (30) *CONSECUTIVE STRESSLESS IN S (M 1.34, S 1.37); that is, if the grammars of (43) are correct this is not an instance of the fallacy of expected values.

8.5 Puzzle constraints

The "puzzle constraints" of §5.7 are difficult to explain under the approach to metrics taken here, since they forbid stress in S position. Thus it would be gratifying if the approach of maxent metrics could make them disappear as epiphenomenal, the consequence of other, better motivated constraints. This did not happen; the constraint (37b) *POSTTONIC INVERSION received a substantial weight for both poets (S 1.70, M: 2.02) and tested as highly significant.

We explored one possibility for eliminating this constraint, adding the phonological constraint *CLASH forbidding consecutive stressed syllables. (This is violated by all lines that violate (37b).) This did not help; the constraint remained in the grammar, with a similar weight.[34]

---

[34] A reviewer for *Language* asked us what might constitute a *counterexample* to a maxent theory of metrics. By this we assume is meant not just the mathematical framework but the substantive principle (5) and its deduced consequences. Kiparskyan "puzzle" constraints, which ban stress in S, certainly do look like a counterexample, though we can hope that a deeper future understanding might yet resolve the puzzle.

8.6 Summary

To sum up what we learn from the grammars (43), we believe that on the whole they support the findings of past research by metrists—both generativist and traditional—on these two bodies of verse. In particular, the grammars provide at least some support for the following general points: (a) Bracketing agreement—independently of stress—is important, and includes a tendency for agreement with the foot boundaries and not just line boundaries. (b) Stress is regulated with special strictness in three contexts: when it is lexical, when it is phrase-final, and when it is line-final. (c) Contrariwise, mismatched stress is licensed when it follows a break, which can be either prosodic (e.g., $[_{IP}$) or metrical ($[_{Line}$).

8.7 Stress maximum constraints

There is one conspicuous gap in the general pattern of support for previous theoretical proposals: we find no support for any kind of stress maximum constraint. No matter what combination of data coding (AS, BH) and constraint selection strategy (bottom-up, top-down) we used, not a single member of the stress maximum constraint family was selected for our grammars. The most likely reason, illustrated earlier in (28), is that the descriptive work attributable to stress maximum constraints is already done by other constraints that have broader empirical support. In other words, stress maximum constraints appear to embody the fallacy of expected values.

Since stress maximum constraints have played an important role in the generative metrics literature, we sought to test further whether it is appropriate, as our calculations suggest, to do without them. Specifically, we experimented with the strategy of giving the stress maximum constraints a “head start”. We set up an initial grammar whose constraints were all and only the stress maximum constraints of our UM (i.e., (29)). We then let the maxent system add to this grammar whatever additional constraints from the UM were justified by the gain criterion. As before, we did this for both poets and both data coders, a total of four tests. Each time, we found that the maxent system installed additional constraints on top of the stress maximum constraints, and that when grammar learning was complete, none of the stress maximum constraints with which we had started tested as significant.

It also seemed possible that the stress maximum constraints were doing poorly because we posited so many of them—perhaps each one can be replaced by another, so that no one constraint tests as significant. To check this, we produced very small grammars consisting of just one single stress maximum constraint, selected from the two best performing ones for any combination of coder and poet. Starting with these very small stress maximum grammars, we again let the system select from the full UM, and again the newly selected constraints functionally replaced the stress maximum constraints.

We were curious to see which constraints were doing the work of the stress maximum constraints. A method that proved effective is based on the fact that in the course of weighting, the maxent system computes the expected number of violations for a constraint, given the other constraints and weights. For stress maximum constraints, we most often find that the expected number of violations is actually lower than the observed number, which is why the stress maximum constraints did not make it into our grammars—under such circumstances they would actually harm the fit of the model to the data. However, when we also take out certain other constraints, then the expected violation counts for stress maximum constraints rises above the observed level, which tells us that the removed constraint is doing work that the stress maximum constraint is doing. Using this technique, we find that the principal constraints that "usurp the role" of the simple stress maximum constraint (29a) are (13a) *STRESS IN W and (35f) *FALL FROM W(~[$_{IP}$__).

8.8 Gradience based on the Prosodic Hierarchy

Hayes (1989) hypothesized that English metrics involves various continua that are interpretable under the assumption of a strictly-layered Prosodic Hierarchy. Constraints applying with special strictness at the ends of phrases are stricter at higher-ranking phrases, and constraints assigning special license for inversion give more license at the beginning of higher-ranking phrases. This claim is plausibly extendible to the pure-bracketing constraints discussed here (§5.2, §8.1): at foot and especially line boundaries, the higher-ranking the matching prosodic break, the better; and medially to lines and especially feet, the higher-ranking the matching prosodic break, the worse.

As explained above in §8.1, the way we detect such cases in maxent is that a positive weight gets assigned to more than one member (W, CG, P, IP) of a particular phrasing-based constraint family, so that the summing of weights produces a cumulative effect. This pattern is indeed observed here and there in our data, as follows:

(47)	*Poet*	*Phenomenon*	*Constraint family*	*Phrasal categories of selected constraints*
	S, M	Align line breaks at prosodic breaks	(9e-h)	W, CG, IP
	S	No breaks medial to line	(10e-h)	P, IP
	M	No breaks medial to line	(10e-h)	CG, P
	S	No breaks medial to foot	(10a-d)	P, IP
	S	No breaks medial to initial foot	(11a-d)	CG, P
	M	No breaks medial to initial foot	(11a-d)	CG, P, IP
	M	No breaks medial to final foot	(11e-h)	P, IP

However, none of the constraints based on phrase-final strictness ((24)) or phrase-initial license for stress ((35)) show the sort of culminativity that Hayes claimed for them. To be sure, large disparities of frequency relating to phrase rank are visible in the data (see the frequency count columns in the Appendix below), but these disparities do not cash out as culminative weightings in a maxent grammar. It is possible that larger data samples or more accurate coding of phrasal structure could change this picture.

9. Comparison with other coding and learning schemes

As noted above, we developed grammars using two data coders (BH, AS) and two methods of constraint selection (bottom up, top down). In §8, we reported the bottom-up BH grammar as the most representative; in (48) below we give the others, including all constraints that were selected under any of the eight grammars. The reader may check for the representativeness of the grammar described above by comparing its values, given in italics, with the others.[35]

(48) *Grammars learned with two data coders and two regimes of constraint selection*

		Shakespeare				*Milton*			
		Bottom-up		*Top-down*		*Bottom-up*		*Top-down*	
		BH	*AS*	*BH*	*AS*	*BH*	*AS*	*BH*	*AS*
(9)	c. ALIGN(Foot, CG)	*ns*	ns	ns	0.32	*ns*	ns	ns	ns
	d. ALIGN(Foot, W)	*ns*	ns	ns	ns	*0.15*	0.33	0.15	0.33
	e. ALIGN(Line, IP)	*2.52*	2.61	2.52	2.23	*0.36*	0.28	0.36	0.28
	g. ALIGN(Line, CG)	*3.09*	2.28	3.09	2.11	*2.53*	2.11	2.53	2.11
	h. ALIGN(Line, W)	*11.40*	11.69	6.70	6.93	*14.12*	15.70	8.24	8.51
(10)	a. ALIGN(IP, Foot)	*0.48*	0.29	ns	ns	*ns*	ns	ns	ns
	b. ALIGN(P, Foot)	*0.41*	0.39	0.52	ns	*0.28*	0.18	0.28	0.18
	e. ALIGN(IP, Line)	*0.85*	0.73	1.00	0.96	*ns*	ns	ns	ns
	f. ALIGN(P, Line)	*0.19*	0.16	0.15	0.48	*0.29*	0.54	0.29	0.53
	g. ALIGN(CG, Line)	*ns*	0.14	ns	0.14	*0.13*	0.07	0.13	0.07
(11)	a. *]$_{IP}$ - FOOT$_1$	*ns*	ns	ns	ns	*1.19*	0.67	1.19	ns
	b. *]$_{P}$ - FOOT$_1$	*0.81*	ns	0.82	0.21	*0.80*	1.22	0.80	1.50
	c. *]$_{CG}$ - FOOT$_1$	*0.67*	0.65	0.67	0.56	*0.18*	0.17	0.18	0.17
	e. *]$_{IP}$ - FOOT$_5$	*2.58*	2.65	2.86	2.65	*2.98*	2.90	2.98	2.90
	f. *]$_{P}$ - FOOT$_5$	*ns*	ns	ns	ns	*0.30*	0.31	0.30	0.31
(13)	a. *STRESS IN W	*2.94*	2.93	2.93	2.93	*2.95*	3.01	2.95	3.01
(15)	a. *RISE FROM S	*ns*	ns	ns	ns	*0.42*	0.44	0.42	0.47
(17)	a. *NO FALL FROM S	*0.28*	0.20	0.28	0.20	*0.21*	0.22	0.21	0.21
	b. *NO RISE FROM W	*ns*	0.07	ns	0.07	*0.11*	0.15	0.11	0.17
(20)	a. *RISE FROM S(lexical)	*2.59*	2.32	2.56	2.11	*ns*	ns	ns	ns
(22)	a. *NO RISE FROM W(final foot)	*1.68*	1.52	1.69	1.52	*2.28*	2.23	2.28	2.22
	b. *FALL FROM W(final foot)	*ns*	1.05	ns	1.05	*ns*	ns	ns	ns
(24)	a. *RISE FROM S(IP-final)	*ns*	0.55	ns	0.59	*2.01*	1.86	2.01	1.90
	d. *RISE FROM S(− +, IP-final)	*1.93*	ns	2.21	ns	*ns*	ns	ns	ns
(26)	c. *RISE FROM S(lexical, CG-final)	*ns*	ns	ns	ns	*12.86*	1.47	6.99	ns
(30)	*CONSECUTIVE STRESSLESS IN S	*1.34*	1.34	1.34	1.33	*1.27*	1.33	1.27	1.33
(32)	a. *EXTRAMETRICAL	*2.61*	2.54	2.61	1.99	*3.56*	3.48	3.56	3.48
	b. *EM WITHOUT FALL	*15.04*	3.45	8.84	3.47	*3.29*	2.55	3.31	2.55
	c. *STRESSED EXTRAMETRICAL	*10.17*	ns	5.45	ns	*ns*	ns	ns	ns
	d. *NONLEXICAL EXTRAMETRICAL	*2.43*	2.47	2.43	2.47	*3.04*	3.08	3.03	3.08
	e. *EXTRAMETRICAL(~IP-final)	*ns*	0.69	ns	1.06	*ns*	ns	ns	ns
(35)	c. *FALL FROM W(lexical, ~[$_{CG}$__)	*1.44*	1.55	1.43	1.41	*ns*	ns	ns	ns
	f. *FALL FROM W(~[$_{IP}$__)	*0.84*	0.79	0.86	0.76	*ns*	ns	ns	ns

[35] We list all constraints whose significance value as described in §7.1 was greater than .01. "c1-c8" indicate the corresponding data columns of (48). (11a) c7 = .015; (11b) c4 = .11; (11c) c6 = .012, c8 = .012; (17b) c2 = .13, c5 = .015, c7 = .015; (24a) c2 = .037, c4 = .034; (32c) c1 = .055, c3 = .054; (32e) c2 = .015; (35c) c1 = .064, c2 = .042, c3 = .068, c4 = .081; (35h) c6 = .011.

	h. *FALL FROM W(~[$_{CG}$__)	*ns*	ns	ns	ns	*0.52*	0.35	0.52	0.37
	j. *FALL FROM W(~[$_{Line}$__)	*ns*	ns	ns	ns	*0.37*	0.39	0.37	0.38
(37)	b. *POSTTONIC INVERSION	*1.70*	1.61	1.69	1.59	*2.02*	1.94	2.02	1.93
(39)	a. *STRESSLESS CG	*3.39*	2.67	3.39	2.67	*3.22*	2.62	3.22	2.62
	b. *STRESSLESS P	*ns*	ns	ns	ns	*1.44*	1.67	1.44	1.67
	d. *EXTENDED LAPSE	*14.73*	12.71	8.45	8.07	*2.34*	2.62	2.34	2.62
	e. *WORD-INITIAL LAPSE	*20.81*	20.65	9.59	9.21	*20.27*	20.10	9.43	9.34

In general it appears that our illustrative grammar is a typical one. One of the alternative grammars (AS top-down) selected constraint (32e) *EXTRAMETRICAL(~IP-final) for the Shakespeare data; this is why this constraint should not necessarily be regarded as a fallacy-of-expected-values case (§8.3). Other variation strikes us as minor: sometimes, different members of the same constraint family are selected to do a particular part of the descriptive work, and of course the weights vary to some degree. In general, it appears that our different data coders and selection methods led to similar conclusions about the metrics of these poets.

10. Application of inductive constraint learning

The model we are working with has the capacity to invent its own constraints (Hayes and Wilson 2008, §4), using feature bundles that define natural classes. When we implemented the metrical constraints above in software, we expressed them using an ad hoc feature system invented for the purpose.[36] In principle, any sequence of feature matrices defining natural classes (up to a user-specified length) can be tested as a metrical constraint, and the principle of maximum likelihood used to select a grammar from such constraints, just as Hayes and Wilson (2008) did for phonotactics.[37]

We tried running this system on our metrical data and found the results extremely difficult to interpret. The constraints do not often match the metrical literature, and significantly, roughly a third of them do not even mention the metrical position (S, W etc.) in which the relevant phonological configuration occurs. We conjecture that this uninterpretability results from the fact that the system is indiscriminately trying to learn both the prosodic phonology of English and the system of metrics as if they were one single system. In contrast, the native speaker of English who is acquiring an appreciation of English poetry comes preequipped with full knowledge of the language's phonology, and is tacitly aware that the task is to establish how this phonological material reflects the rhythm of the meter. In the work described here, we in effect forced the system to focus on the metrical task by prefabricating the constraints.

---

[36] Our features classify units consisting of pairings of syllables with metrical slots. The feature [Strong] distinguishes S from W position, [Stress] distinguishes stressed from stressless syllables, [Rise] distinguishes syllables that have less stress than the following syllable, [Fall] is defined analogously, and [+Word], [+CG], [+P], [+IP] are assigned to syllables at the right edges of the phrasal categories indicated.

[37] We used a modified version of the system (Wilson, 2010) that uses the principle of gain, described above, to pick new constraints.

11. Conclusions

We have suggested that most of the research literature on this form of verse can be interpreted as embodying the fundamental principle of resemblance in (5), amplified by just a few principles of particular salience, such as lexical, phrase-final, and line-final. We have fleshed out these ideas with explicit constraints and tested them by forming maxent grammars for our Shakespeare and Milton corpora. The resulting metrical grammars are, to our knowledge, the first that assign well-formedness values to lines based on the full integration of all factors that affect metricality. These scores moreover have an explicit interpretation under the theory of probability, and are obtained using a rational criterion, maximum likelihood. The penalty scores assigned to actual lines under our grammars strike us as mostly reasonable, are generally high for lines judged in earlier work to be unmetrical, and to a fair degree pass the "Youmans test" of accounting for word order.

The maxent approach has also made it possible to assess proposals in the research literature more rigorously than has been done before. We focused in particular on the question of whether complex (tacitly conjoined) constraints suffer from the fallacy of expected values. The results of our inquiry varied. In the case of stress maximum constraints, we consistently found that their effects could be accounted for more economically with other, simpler constraints. Phrase-final constraints, though tacitly conjoined, nevertheless appear to have explanatory merit. The conjoined constraint banning consecutive unstressed S positions likewise passed muster, and our results for Kiparsky's claim that extrametricality and run-on status are incompatible were equivocal.

Constraints neglected in earlier work—perhaps because they are so far from being exception-free—also emerged as important in our study. In particular, there appears to be modest pressure for the prosodic boundaries of the line to match foot boundaries. This provides a new empirical argument for Kiparsky's (1977) claim that feet are real and that bracket matching at the foot level forms part of the metrical system.

12. Issues for further research

12.1 Corpus size

Our work is based on relatively small data corpora, roughly 2100 lines for each poet. The task of scanning all of the lines and annotating every syllable for stress and phrasing proved to be quite time-consuming. Yet, as Kiparsky (1977:191) points out, there is a great advantage of using an extremely large corpus, such as the complete Shakespeare oeuvre. For instance, only corpora of this size are likely to bear reliably on the status of Milton's violations of (35c) *FALL FROM W(lexical, ~[$_{CG}$__), which are quite rare (perhaps a couple dozen in all of his work), but common enough to be intuitively detectable as characteristically Miltonic.[38] Thus, we judge that future work in metrics ought to make use of natural language processing techniques which would

[38] The poet Gerard Manley Hopkins expressed his admiration of them in correspondence, and imitated them in his poetry (Abbott 1935:38; Kiparsky 1977:203). The examples in our corpus are 9.840 *Beyond / all past / exam- / ple and / future* and perhaps also 8.1061 *Of Phi- / liste- / an Da- / lilah, / and wak'd.*

automate scansion, syntactic parsing, phrasal stress assignment, and the annotation of phonological phrasing (see Shih 2010). If such technology can be made reliable enough, it would become feasible to construct explicit metrical analyses of the entire English verse canon, a goal to which we think metrics ought to aspire.

12.2 Controlling for ordinary-language phonology

As noted in §5.8, the prosodic pattern of verse must be interpreted against the backdrop of the prosodic patterns of the language in which it is written. We have taken a conservative approach, including in our model only a few, mostly robust constraints of English phonology ((39)). A model that included more information about the statistical tendencies of English might yield different conclusions. A technique to consider is one pioneered by scholars of the "Russian" school, who evaluated verse against a baseline prose model gathered or synthesized from contemporary prose; see for instance Gasparov (1980, 1987), Tarlinskaja (1976) and for Western applications and extensions Bailey (1975) and Hall (2006). The prose model approach faces a fundamental difficulty: how to select the lines of the model on a principled basis. For instance, for iambic pentameter, should they be completely random ten-syllable sequences, or sequences demarcated by phrase breaks, or sequences that have stress in certain positions?[39] A maxent approach might yield insight.

12.3 Metrics based on corpus evidence

We focused on the oeuvres of Shakespeare and Milton because so much careful analytic work has been done on their verse. Hence, our research is based entirely on what can be learned from close scrutiny of our corpora. In generative linguistics, a proposed analysis thus obtained is normally further tested by constructing novel examples and submitting them to native speakers in elicitation. Could this be done here, in particular with modern verse readers?

In fact, we are skeptical that anything like native-speaker intuitions about Shakespeare or Milton's verse could be gathered from modern readers, however experienced. The vocabulary of Shakespeare and Milton is rather demanding, and as a result virtually all modern readers first engaged with this verse at an age that follows the critical period for language acquisition. In addition, virtually all contemporary readers have extensive experience with other forms of verse, including other forms of iambic pentameter. Thus we think that even the most sensitive modern readers could not rightly be considered to have any better than "L2" (second language) command over the target varieties, and would thus fall short of normal scientific standards for selection of language consultants. We also think it likely that some of the more sophisticated readers might have their intuitions clouded by conscious theorizing. A more promising route for future research would be to study the verse of living poets willing to be interrogated about their well-formedness intuitions.

[39] Gasparov (1987, 324) suggests we should "compile a list of line-long word combinations which comply with the rules of the given meter." This may be possible for some verse traditions, but for English iambic pentameter, where there are hardly any inviolable rules (§3), the procedure is circular.

12.4 Does maxent correctly model well-formedness?

A further issue concerns whether the formulae of maxent (see (2)) correctly model the way constraint violations determine well-formedness. Maxent is "rational" in the sense that it combines constraints by a criterion that best matches the patterns in the data. As such, it has the best claim to our attention at this phase of research. Yet the real world might be more complicated: perhaps in the metrical grammars that were internalized by Shakespeare and Milton, violation of a strong constraint counted for more—or for less—than would be justified under the maxent criterion.

These possibilities may be related to examples given earlier: the scores for (45) above include violations of a powerful constraint, whereas those for (44c) pool violations of weaker constraints. If we could somehow determine whether one class of line is consistently worse or better than the other, it would shed light on these questions.[40] (Our grammar of (43) assigns slightly higher average penalties to (44c).) Unfortunately, for the reasons given in the previous section, we think it is not practical to address this issue, which would be better pursued with research on the verse of living poets.

12.5 Modeling variation within and across poets

Shakespeare and Milton did not create their metrical systems *de novo*; rather, these systems each emerged from an existing tradition of iambic pentameter verse composition. A richer model than what is proposed here would characterize a poet's internalization of the ambient metrical tradition, as well as his own instantiation and refinement of it. This could be done with a hierarchical model (e.g. Dudík et al. 2007) in which the overall metrical tradition is expressed as a higher-level distribution over constraint weights, and the particular poet's practice at a given point in his career as a sample either from this higher level distribution directly or from a poet-level distribution lying between the metrical tradition and the poet's particular works.

Appendix: Violation counts of all constraints

Classified by poet and data coder.

		Shakespeare		*Milton*	
		BH	*AS*	*BH*	*AS*
(9)	a. ALIGN(Foot, IP)	2,455	2,453	3,010	3,020
	b. ALIGN(Foot, P)	5,105	5,209	5,448	5,461
	c. ALIGN(Foot, CG)	6,839	6,830	7,439	7,719
	d. ALIGN(Foot, W)	8,380	8,316	9,146	9,092
	e. ALIGN(Line, IP)	302	277	1,279	1,235
	f. ALIGN(Line, P)	151	128	701	806
	g. ALIGN(Line, CG)	7	14	50	72
	h. ALIGN(Line, W)	0	0	0	0

[40] A *Language* reviewer suggested that we should simply follow the research literature in assuming that the lines of (45) are "unmetrical" and those of (44c) "complex but permitted". But the literature includes no such comparative judgments of these lines.

(10)	a. Align(IP, Foot)	195	275	678	687
	b. Align(P, Foot)	1,146	1,255	1,469	1,363
	c. Align(CG, Foot)	2,996	3,001	3,362	3,381
	d. Align(W, Foot)	9,232	9,227	8,899	8,900
	e. Align(IP, Line)	826	953	2,008	2,030
	f. Align(P, Line)	3,175	3,273	3,943	3,663
	g. Align(CG, Line)	6,624	6,535	7,209	7,237
	h. Align(W, Line)	15,503	15,503	15,134	15,125
(11)	a. $*]_{IP}$ - Foot_1	18	82	18	19
	b. $*]_{P}$ - Foot_1	50	104	77	56
	c. $*]_{CG}$ - Foot_1	211	266	320	332
	d. $*]_{W}$ - Foot_1	1,917	1,911	1,767	1,763
	e. $*]_{IP}$ - Foot_5	4	5	8	8
	f. $*]_{P}$ - Foot_5	243	256	170	157
	g. $*]_{CG}$ - Foot_5	661	651	583	596
	h. $*]_{W}$ - Foot_5	1,554	1,553	1,619	1,624
(13)	a. *Stress in W	1,233	1,192	1,232	1,145
	b. *Stressless in S	2,673	2,772	2,714	2,688
(15)	a. *Rise from S	428	430	303	279
	b. *Fall from W	558	537	658	659
(17)	a. *No Fall from S	2,605	2,622	2,694	2,605
	b. *No Rise from W	2,865	2,860	2,887	2,786
(19)	a. *Stress Mismatch(− +)	353	363	264	243
	b. *Stress Mismatch(+ −)	463	465	569	577
(20)	a. *Rise from S(lexical)	3	4	13	11
	b. *Fall from W(lexical)	137	148	255	276
(22)	a. *No Rise from W(final foot)	215	227	133	133
	b. *Fall from W(final foot)	14	6	12	9
	c. *Stressless in S(final foot)	194	226	121	131
(24)	a. *Rise from S(IP-final)	8	13	5	5
	b. *Rise from S(P-final)	103	98	41	35
	c. *Rise from S(CG-final)	281	278	179	165
	d. *Rise from S(− +, IP-final)	1	5	2	3
	e. *Rise from S(− +, P-final)	58	72	24	21
	f. *Rise from S(− +, CG-final)	267	269	170	159
(26)	a. *Rise from S(lexical, IP-final)	0	1	0	0
	b. *Rise from S(lexical, P-final)	0	2	0	0
	c. *Rise from S(lexical, CG-final)	1	2	0	3
(29)	a. *Stress Max in W	110	124	74	80
	b. *Stress Max in W(IP-bounded)	92	101	53	60
	c. *Stress Max in W(P-bounded)	52	55	32	35
	d. *Stress Max in W(CG-bounded)	26	23	8	11
	e. *Stress Max in W(W-bounded)	0	0	1	0
	f. *Stress Max in W(− + −)	22	43	32	43
	g. *Stress Max in W(− + −, IP-bounded)	12	32	20	28
	h. *Stress Max in W(− + −, P-bounded)	10	18	14	21
	i. *Stress Max in W(− + −, CG-bounded)	5	11	3	7
	j. *Stress Max in W(rising-lexical)	1	1	1	1
	k. *Stress Max in W(falling-lexical)	3	3	5	6
	l. *Stress Max in W(rising-lexical, IP-bounded)	1	1	1	1

	m. *STRESS MAX IN W(falling-lexical, IP-bounded)	1	1	3	3
	n. *STRESS MAX IN W(rising-lexical, P-bounded)	1	1	1	1
	o. *STRESS MAX IN W(falling-lexical, P-bounded)	1	0	2	1
	p. *STRESS MAX IN W(rising-lexical, CG-bounded)	1	1	1	1
	q. *STRESS MAX IN W(falling-lexical, CG-bounded)	0	0	2	1
(30)	*CONSECUTIVE STRESSLESS IN S	158	165	169	151
(32)	a. *EXTRAMETRICAL	162	165	73	73
	b. *EM WITHOUT FALL	0	2	1	2
	c. *STRESSED EXTRAMETRICAL	0	2	0	1
	d. *NONLEXICAL EXTRAMETRICAL	14	16	4	4
	e. *EXTRAMETRICAL(~IP-final)	17	12	48	45
	f. *EXTRAMETRICAL(~P-final)	9	9	33	32
(35)	a. *FALL FROM W(lexical, ~[$_{IP}$__)	27	29	128	130
	b. *FALL FROM W(lexical, ~[$_{P}$__)	12	11	56	74
	c. *FALL FROM W(lexical, ~[$_{CG}$__)	1	1	5	6
	d. *FALL FROM W(CG-level, ~[$_{IP}$__)	72	74	171	171
	e. *FALL FROM W(CG-level, ~[$_{P}$__)	44	41	75	98
	f. *FALL FROM W(~[$_{IP}$__)	191	182	324	320
	g. *FALL FROM W(~[$_{P}$__)	124	111	157	182
	h. *FALL FROM W(~[$_{CG}$__)	74	67	55	60
	i. *FALL FROM W(lexical, ~[$_{Line}$__)	33	37	60	65
	j. *FALL FROM W(~[$_{Line}$__)	237	224	239	219
(37)	a. *POSTTONIC INVERSION(lexical)	4	5	6	8
	b. *POSTTONIC INVERSION	11	12	9	10
(39)	a. *STRESSLESS CG	99	189	109	201
	b. *STRESSLESS P	29	91	9	11
	c. *STRESSLESS IP	1	63	2	4
	d. *EXTENDED LAPSE WITHIN WORD	0	0	4	3
	e. *WORD-INITIAL LAPSE	0	0	0	0

References

Abbott, Claude Colleer. 1935. *The letters of Gerard Manley Hopkins to Robert Bridges*. London: Oxford University Press.

Adams, Stephen. 1997. *Poetic designs: An introduction to meters, verse forms, and figures of speech*. Peterborough, Ontario: Broadview Press.

Andrews, Clarence Edward. 1918. *The writing and reading of verse*. New York: D. Appleton and Company.

Attridge, Derek. 1982. *The rhythms of English poetry*. Harlow, Essex, UK: Longman.

Bailey, James. 1975. *Toward a statistical analysis of English verse*. Lisse: Peter de Ridder Press.

Barnes, Mervin and Helmut Esau. 1978. English prosody reconsidered. *Language and Style* 11:212-222.

Berger, Adam L., Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22:39–71.

Bjorklund, Beth. 1978. *A study in comparative prosody: German and English iambic pentameter*. Stuttgart: Hans-Dieter Heinz.

Boersma, Paul and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Rutgers Optimality Archive 970.

Bridges, Robert. 1921. *Milton's prosody: with a chapter on accentual verse*. London: Oxford University Press.

Chen, Matthew. 1979. Metrical structure: evidence from Chinese poetry. *Linguistic Inquiry* 10: 371-420.

Chen, Stanley and Ronald Rosenfeld. 2000. A survey of smoothing techniques for ME models. *IEEE transactions on speech and audio processing* 8:37-50.

Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Della Pietra, Stephen, Vincent J. Della Pietra, and John D. Lafferty. 1997. Inducing features of random fields. *IEEE transactions on pattern analysis and machine intelligence* 19:380–393.

Duda, Richard O., Peter E. Hart, and David G. Stork. 2001. *Pattern classification*. 2nd ed. New York: John Wiley & Sons.

Dudík, Miroslav, David M. Blei, and Robert E. Schapire. 2007. Hierarchical maximum entropy density estimation. In *Proceedings of the 24th International Conference on Machine learning*, Corvallis, OR, pp. 249-256.

Duffell, Martin. 2002. The Italian line in English after Chaucer. *Language and Literature* 11:291-305.

Duffell, Martin. 2008. Some observations on English binary metres. *Language and Literature* 17: 5-20.

Fabb, Nigel and Morris Halle. 2008. *Meter in poetry: A new theory*. Cambridge: Cambridge University Press.

Friedberg, Nila. 2002. Metrical complexity in Russian verse: a study of form and meaning. Ph.D. dissertation, Department of Linguistics, University of Toronto.

Friedberg, Nila. 2006. Constraints, complexity, and the grammar of poetry. In B. Elan Dresher and Nila Friedberg, eds., *Formal approaches to poetry: Recent developments in metrics*. Berlin: Mouton de Gruyter, pp. 212-231.

Furnivall, Frederick James and John James Munro. 1910. *Shakespeare; life and work*. Cassell.

Gasparov, Mikhail L. 1980. Quantitative methods in Russian metrics: Achievements and prospects. In G. S. Smith, ed., *Russian Poetics in Translation. Vol. 7: Metre, Rhythm, Stanza, Rhyme*. Colchester: University of Essex.

Gasparov, Mikhail L. 1987. A probability model of verse. *Style* 21: 322-358

Goldwater, Sharon and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the Workshop on Variation within Optimality Theory, Stockholm University*, 2003.

Golston, Chris. 1998. Constraint-based metrics. *Natural Language and Linguistic Theory* 16: 719-770.

Hall, Daniel Currie. 2006. Modeling the linguistics-poetics interface. In B. Elan Dresher and Nila Friedberg, eds., *Formal approaches to poetry: Recent developments in metrics*. Berlin: Mouton de Gruyter, pp. 233-249.

Halle, Morris and S. Jay Keyser. 1966. Chaucer and the study of prosody. *College English* 28:187-219.

Halle, Morris, and S. Jay Keyser. 1971. *English stress: Its form, its growth, and its role in verse,* Harper and Row, New York.

Hammond, Michael. 1991. Poetic meter and the arboreal grid. *Language* 67: 240-259.

Hanson, Kristin. 2006. Shakespeare's lyric and dramatic styles. In B. Elan Dresher and Nila Friedberg, eds., *Formal approaches to poetry: Recent developments in metrics*. Berlin: Mouton de Gruyter, pp. 111-134.

Hanson, Kristin and Paul Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72: 287-335.

Hayes, Bruce. 1983. A grid-based theory of English meter. *Linguistic Inquiry* 14, 357-393.

Hayes, Bruce. 1988. Metrics and phonological theory. In Frederick Newmeyer, ed., *Linguistics: The Cambridge survey*. Cambridge: Cambridge University Press, 220-249.

Hayes, Bruce. 1989. The Prosodic Hierarchy in meter. In Paul Kiparsky and Gilbert Youmans, eds., *Rhythm and meter*, Academic Press, Orlando, FL.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce. 2008. Two Japanese children's songs. Ms., Department of Linguistics, UCLA, Los Angeles, CA. http://www.linguistics.ucla.edu/people/hayes/251metrics/papers/TwoJapaneseChildrensSong2008Version.pdf

Hayes, Bruce and Claire Moore-Cantwell. 2001. Gerard Manley Hopkins's sprung rhythm: corpus study and stochastic grammar. *Phonology* 28:235-282.

Hayes, Bruce and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379-440.

Jakobson, Roman. 1933. Über den Versbau der serbokroatischen Volksepen. *Archives néerlandaises de phonétique expérimentale* 7-9: 44-53. Reprinted in Jakobson (1966).

Jakobson, Roman 1952. Slavic epic verse: studies in comparative metrics. Oxford Slavonic papers 3: 21-66. Reprinted in Jakobson (1966).

Jakobson, R. 1966. *Selected writings IV: Slavic epic studies*. The Hague: Mouton.

Jespersen, Otto. 1900, English translation 1933. Notes on meter. In *Linguistica*, 249-274. Copenhagen: Levin and Munksgaard.

Keller, Frank. 2000. Gradience in grammar: experimental and computational aspects of degrees of grammaticality. Ph.D. Thesis, University of Edinburgh.

Kelly, Michael H. 2004. Word onset patterns and lexical stress in English. *Journal of Memory and Language* 50: 231-244.

Kelly, Michael H. and J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14: 389-403.

Keshet, Ezra. 2005. Relatively optimal textsetting. Ms., Department of Linguistics and Philosophy, MIT. Downloaded 20 May 2012 from http://www.ezrakeshet.com/Home/papers.

Kiparsky, Paul. 1968. Metrics and morphophonemics in the *Kalevala*. In Charles Gribble, (ed.), *Studies presented to Roman Jakobson by his students*. Cambridge, MA: Slavica.

Kiparsky, Paul. 1975. Stress, syntax, and meter. *Language* 51: 576-616.

Kiparsky, Paul. 1977. The rhythmic structure of English verse. *Linguistic Inquiry* 8: 189-248.

Koelb, Clayton. 1979. The iambic parameter revisited. *Neophilologus* 63: 321–329

Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Report CU-CS-465-90. Computer Science Department, University of Colorado at Boulder. (Online: www.cs.colorado.edu/department/publications/reports/docs/CU-CS-465-90.pdf.)

Lerdahl, Fred and Ray Jackendoff. 1983. *A generative theory of tonal music*. Cambridge: MIT Press.

Liberman, Mark and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

Magnuson, Karl, and Frank G. Ryder. 1970. The study of English prosody: an alternative proposal. *College English* 31. 789-820.

Magnuson, Karl and Frank G. Ryder. 1971. Second thoughts on English prosody. *College English*. 33. 198-216.

McCarthy, John J. and Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology* 79–153.

Nespor, Marina and Irene Vogel. 1986/2007. *Prosodic phonology.* Dordrecht: Foris. 2nd. ed. Berlin: Walter de Gruyter.

Oras, Ants. 1960. *Pause patterns in Elizabethan and Jacobean drama: An experiment in prosody*. Gainesville: University of Florida Press.

Piera, Carlos. 1980. Spanish verse and the theory of meter. Doctoral dissertation, UCLA. http://www.linguistics.ucla.edu/general/dissertations/Piera_Diss_UCLA1980.pdf

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker. 2010. Harmonic grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27: 77-117.

Prince, Alan. 1989. Metrical forms. In Paul Kiparsky and Gilbert Youmans, eds., *Rhythm and meter*, 44-80. Orlando, FL: Academic Press.

Prince, Alan and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical report, Rutgers University Center for Cognitive Science.

Schipper, Jakob. 1910. *A history of English versification*. Oxford: Clarendon Press.

Selkirk, Elizabeth O. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aronoff and Marie-Louise Kean, eds., *Juncture*. Saratoga, CA: Anma Libri, 107-29.

Selkirk, Elizabeth O. 1984. *Sound and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, Elizabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3:371–405.

Shih, Stephanie. 2010. Corpus and computational tools for generative metrics. Paper given at the 2010 Annual Meeting of the Linguistic Society of America, Baltimore.
Sitter, John. 2007. Pope's versification and voice. In Pat Rogers, ed., *The Cambridge Companion to Alexander Pope*. Cambridge: Cambridge University Press, 37-48.
Smolensky, Paul. 1995. On the internal structure of the constraint component of UG. Colloquium given at UCLA, April 7, 1995. Available on Rutgers Optimality Archive, #86.
Smolensky, Paul, and Geraldine Legendre. 2006. *The harmonic mind*. Cambridge: MIT Press.
Sprott, S. Ernest. 1953. *Milton's art of prosody*. Oxford: Blackwell.
Steele, Timothy. 1999. *All the fun's in how you say a thing: an explanation of meter and versification*. Athens: Ohio University Press.
Swiger, Tami. 1994. Japanese folk and children's songs. Ms., Department of Linguistics, UCLA, Los Angeles, CA. http://www.linguistics.ucla.edu/people/hayes/251metrics/papers/ Swiger1994JapaneseFolkAndChildrensSongs.pdf
Tarlinskaja, Marina. 1967. Aktsentnye osobennosti angliiskogo sillabo-tonicheskogo stikha. [The accentual features of English syllabo-tonic verse]. *Vop.yaz* 3:81-91.
Tarlinskaja, Marina. 1976. *English verse: theory and history*. The Hague: Mouton.
Tarlinskaja, Marina. 1987. *Shakespeare's verse: iambic pentameter and the poet's idiosyncrasies*. New York: Peter Lang.
Tarlinskaja, Marina. 1989. General and particular aspects of meter: literatures, epochs, poets. In Paul Kiparsky and Gilbert Youmans, eds., *Rhythm and meter*, Academic Press, Orlando, FL, 121-1545.
Tarlinskaja, Marina. 2006. What is "metricality"? English iambic pentameter, in Elan Dresher and Nila Friedberg, *Formal approaches to poetry: Recent developments in metrics*. Berlin: Mouton de Gruyter, pp. 53–76.
Tarlinskaja, Marina and Lilya M. Teterina. 1974. Verse—Prose—Meter. *Linguistics* 129: 63-86.
Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in optimality theory*. Cambridge, MA: MIT Press.
Wasserman, Larry. 2004. *All of statistics: A concise course in statistical inference*. New York: Springer.
Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf and Patti J. Price. 1991. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91: 1707-1717.
Wilson, Colin. 2010. Emergent stringency bias in phonotactic learning. Ms. Baltimore: Department of Cognitive Science, Johns Hopkins University.
Wilson, Colin and Marieke Obdeyn. Under review. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms., Johns Hopkins University.
Youmans, Gilbert. 1982. Hamlet's testimony on Kiparsky's theory of meter. *Neophilologus* 66: 490-503.
Youmans, Gilbert. 1983. Generative tests for generative meter. *Language* 59: 67-92.
Youmans, Gilbert. 1989. Milton's meter. In Paul Kiparsky and Gilbert Youmans, eds., *Rhythm and meter*, 341-379. Orlando, FL: Academic Press.