A MaxEnt analysis of the *Beowulf* meter[1]

Final version, to appear after copy editing in *Phonology*

September 2026

Bruce Hayes Donka Minkova

UCLA

[1] We would like to thank our reviewers and editors for extensive and helpful input. We also thank audience members at the M100 celebration of Morris Halle's career at MIT and in the UCLA Phonology Seminar. We also thank Geoffrey Russom for sharing his essential scansion spreadsheet with us.

At several points the paper refers to a set of Supplementary Materials, consisting of databases, spreadsheet calculations, and scholarly commentary. These may be obtained from the *Phonology* website or from brucehayes.org.

Abstract

The metrical system of the Old English epic *Beowulf* is a longstanding puzzle for the theory of meter. We propose an analysis of this system employing the framework of MaxEnt grammar. Our metrical grammar assigns a probability value to each verse type, which can be checked against their frequencies as they appear in the poem.

The *Beowulf* system, being so distinct from other well-studied systems, offers special insights into the general theory of meter. In *Beowulf*, the metrical feet are modeled on the prosodic inventory of word types in a language, rather than constituting purely rhythmic units like iambs or trochees. Despite this difference, our analysis employs constraints that draw upon, and provide support for, metaprinciples applicable throughout metrics: End Weight, Closure, Rhythmic Simplicity, and Optimal Length.

Two general points emerge. First, for the metrics of *Beowulf* a constraint set consisting solely of inviolable constraints is grossly inadequate, for most of the data patterning is gradient and can only be captured using violable constraints. Second, we suggest that work with pencil and paper alone is unlikely to settle any serious questions of analysis in a system like *Beowulf* — an implemented probabilistic framework like MaxEnt is essential: it offers precise quantitative predictions and allows rigorous statistical comparison of alternative analyses.

Keywords: Old English, metrics, MaxEnt grammars, probabilistic grammars

1. Introduction

The purpose of this article is to propose a metrical analysis of the Old English poem *Beowulf*, cast in a MaxEnt approach to generative metrics. We argue that our analysis is a substantial improvement over earlier work in its match to the data, and sheds light on the evolving theory of metrical constraints.

In this introductory section, we lay out the research problem, providing background on *Beowulf*, on generative metrics, and on MaxEnt grammars as applied to metrics.

1.1 The puzzle of the Beowulf meter

The meter of *Beowulf* is stress-based, but displays few of the rhythmic traits (e.g. constant syllable count, even spacing of stresses) familiar from later European meters such as iambic pentameter. To give an idea of this distinct character, we cite a few lines of *Beowulf* in (1) below. To assist the reader we have annotated the text with a reconstructed phonetic transcription.[2] Each line is shown separated with a gap into its two verses, conventionally known as the "a-verse" and the "b-verse". Boldface is used to indicate alliterating consonants. Above each syllable is given its key phonological class; roughly, **S** denotes primary stress, **s** nonprimary stress, and **x** stressless.

(1) *Beowulf,* lines 47-52 (Shield's funeral)[3]

a. *Text, transcription*, *glosses*

x	**x**	**x**	**x**	**x S x**	**S**	**S s x**
Þā	gȳt	hīe	him	ā**s**etton	**s**egen	g[yl]denne
θɑː	jiːt	hiə	hɪm	əˈ**s**ɛttən	**s**ɛɟn	ˈɟʏlˌdɛnnə
Then	yet	they	it DAT.SG	set PRET. PL.	banner ACC.SG.	golden ACC.SG.

S	**x x**	**S x**	**x x**	**S**	**s x**
hēah	ofer	**h**ēafod,	lēton	**h**olm	beran
hæəh	ˈɔvər	ˈ**h**æəvəd	ˈleːtən	**h**ɔlm	ˈbɛrən
high	over	head ACC.SG.	let PRET.PL.	sea ACC. SG	carry INF.

S x	**x**	**S s**	**x**	**x**	**S x**	**s**[4]
gēafon	on	**g**ārsecg;	him	wæs	**g**eōmor	sefa,
ˈ**j**æəvən	ɔn	ˈ**g**ɑːrˌsɛʤ	hɪm	wæs	ˈ**j**ɔːmər	ˈsevə
give PRET.PL.	onto	sea ACC.SG.	they DAT.PL.	be PRET. 3^{RD} SG.	mournful	spirit NOM.SG.

---

[2] This follows the system in Minkova (2025). Two caveats apply: like all phonetic reconstructions, ours are only well-informed guesses, and Old English evolved considerably over time.

[3] Text from Fulk, et al. (eds.) (2008); [ ] marks an editorial emendation.

[4] The use of a single **S** or **s** over a disyllabic sequence, seen here and for ***s**elerǣdende*, *hæleð*, is explained in §3.4.

S s w	**S**	**S**	**x**	**S x**	
murnende	**m**ōd.	**M**en	ne	cunnon	
ˈ**m**ʊrˌnɛndə	**m**oːd	**m**ɛn	nə	ˈkʊnnən	
mourn PR. PTC. SG.	mind NOM.SG.	man NOM.PL.	not	know (how to) PR.PL.	

S x	**x**	**S x**	**S S s x**
secgan	tō	**s**ōðe,	**s**elerǣdende,
ˈ**s**ɛdʒən	toː	ˈ**s**oːðə,	ˈ**s**ɛləˈræːˌdɛndə,
say INF.	in	sooth ACC. SG.	hall-counsel PR. PTC. PL.

S	**x x**	**S x**	**x**	**x**	**S x**	**x S**
hæleð	under	**h**eofenum,	hwā	þǣm	**h**læste	onfēng.
ˈ**h**æləθ	ˈʊndər	ˈ**h**ɛəvənəm,	hwɑː	θæːm	ˈ**h**læstə	ənˈfeːŋg
hero NOM.PL.	under	heaven DAT.PL.	who	that DAT. SG.	load DAT. SG.	receive PRET. SG.

b. Free translation by Ruth Lehmann (1988)

There a standard stood,   streaming golden,
High overhead.   They let the heaving sea
Bear him on the billows.   They had brooding thoughts,
Mournful spirits.   Those men could not
Tell indeed truly,   though they gave trusted advice,
That warriors over waters   welcomed that burden.

The verses of (1) vary in syllable count from four to seven; the full range in the poem is four to nine. Instead of smoothly alternating stress, one often finds outright clashing stresses, as with 48, verse b (**xxSsx**), and long line-initial strings of unstressed syllables, as in 47a (**xxxxxSx**). At least at first blush, it seems that rather little that the field of metrics has learned from the study of other stress-based verse is applicable to the *Beowulf* meter.

Indeed, the meter of *Beowulf* is felt to be a difficult puzzle. The scholarly literature, partially covered below, includes a great variety of often incompatible proposals.[5] The key problem, we think, is a lack of any means to assess clearly when one analysis performs better than another. Here we propose to use MaxEnt to fill this gap.

1.2 Generative metrics

The goals laid out by Halle and Keyser (1966, 1971) in the founding works of generative metrics remain a near-consensus for research today. The idea is that metrical analysis should characterize what linguistic strings constitute well-formed (i.e., metrical) lines of verse in a given tradition, what strings are unmetrical as lines of verse, and among the metrical lines, which are more

[5] We lack space for a survey here. The founding document of the literature is Sievers (1893), and literature overviews appear in Stockwell and Minkova (1997) and Goering (2020). Newcomers to Old English meter would find Terasawa (2011) particularly useful.

"complex" (i.e., non-canonical, perhaps marginal) and which are less complex (i.e., canonical). The iambic pentameter lines of (2) illustrate these ideas; scansion is shown by slashes.

(2) *Simple, complex, and unmetrical lines: iambic pentameter*

a. And bar- / ren rage / of death's / eter- / nal cold? (Shakespeare, Son. 13)
b. Makes black / night beaut- / eous, and / her old / face new. (Son. 27)
c. Ode to / the West / Wind by / Percy / Bysshe Shel- / ley[6] (construct)

Line (2a) is one instance of a fully metrical, canonical line. Line (2b) is "complex"; here Shakespeare employs unusual metrical options, creating a sense of rhythmic irregularity. Lastly, (2c), coined by Halle and Keyser (1971:171), is an example of a fully unmetrical line, a type avoided by Shakespeare and most other pentameter poets.

The same distinctions are present in the *Beowulf* meter. For *Beowulf*, the basic unit of analysis is not the line but rather the verse, as seen above in (1). Here are some examples analogous to (2a-c).

(3) *Simple, complex, and unmetrical verses for the* Beowulf *meter*

a. **S x S x**
Men ne cunnon 50b
[mɛn nə ˈkʊnnən]
'men know not'

b. **x x S x S s x**
ne geˈfrægn ic ˈfrēondˌlīcor 1027a
[nə jəˈfræjn ɪc ˈfreəndˌliːkər]
'not heard I (of) friendlier'

c. **S x s S**
*fāmig-heals flēat (construct)
[ˈfɒːmijˌhæᵊls flæət]
'foamy-neck (ship) floated'

Example (3a) is standardly scanned as **SxSx**, which represents a pattern seen in 1396 verses in *Beowulf*, over a fifth of the 6364-verse poem. **SxSx** has always been considered the unmarked verse type for *Beowulf*. Example (3b), in contrast, is (as we will show), a unique example of its type, and it is only in the context of a full analysis that it emerges as a possible type, and not a random aberration.

[6] Technical notes: *beauteous* in (2b) counts here as disyllabic; (2c) uses the common option of an eleventh "extrametrical" syllable.

Example (3c) was created by Russom (2022:53) to illustrate a verse type that, under his analysis, is fully unmetrical. Verses of this type are indeed entirely absent from the poem. We will seek to replicate this prediction in our own analysis (§8.2).

The characterization of verses (3a-c) as fully metrical, complex, and unmetrical is dependent on a key hypothesis of metrics put forth by Halle and Keyser (1971:157): that the frequency of a metrical type in a large corpus is directly related to its well-formedness in the poet's metrical grammar. Thus, although we cannot directly access the metrical intuitions of the *Beowulf* poet, under Halle and Keyser's "Frequency Hypothesis" we can use verse frequencies as a proxy for those intuitions. This is standard practice in *Beowulf* research as well as in metrical analysis generally.

1.3 MaxEnt metrics

The framework of **MaxEnt metrics** has been applied in some form to a variety of metrical traditions: Hayes and Moore-Cantwell (2011, Gerard Manley Hopkins); Ryan (2011, Homeric Greek, Finnish, Old Norse, Tamil); McPherson and Ryan (2018, Tommo So). The particular MaxEnt technique we will employ here was introduced in Hayes et al. (2012, Shakespeare and Milton) and further explored by Hayes and Schuh (2019, Hausa songs) and Henriksson (2022, Ancient Greek).

The MaxEnt framework is constraint-based. In using constraints, our work follows much existing generativist work on Old English meter (Getty 2002, Golston and Riad 2001, Cooper 2017, 2019, Russom 1987, 2025). It differs from this earlier work in that it assigns to each constraint a numerical **weight**, which intuitively reflects its strength. The framework defines a **probability** for every candidate verse type. This is calculated using the formula in (11) below, based on the constraint violations, the constraint weights, and a set of candidates. Constraints that bear higher weights have a greater effect in lowering the probability of candidates that violate them.

The probabilities computed by a MaxEnt grammar can be directly compared to the attested verse-type frequencies, enabling an explicit empirical evaluation of the analysis. If we are on the right track, the predicted frequencies (probability times total cases) will form a good match to observed frequencies, as seen for instance in scattergrams like (39) below. Such a match can also be evaluated quantitatively and tested statistically (§8), allowing informed comparisons to rival analyses.

An essential trait of MaxEnt metrics is that, as a descendant of Optimality Theory (Prince and Smolensky 1993/2004) it makes use of a GEN function, in our case a large list of imaginable candidate verse types. We follow the tradition of classifying all the verses as sequences of the symbols **S**, **s**, and **x**, as in (1), and our GEN (§4.1) consists of the logically possible concatenations of these symbols up to a certain length. GEN provides a further criterion for evaluating competing analyses: not only should an analysis match the frequencies of existing sequences, it should also assign a zero or near-zero probability to the missing sequences.

1.4 Roadmap

With this background we can now define our purpose more clearly. Our main goal is to construct a new and improved analysis for the *Beowulf* meter using MaxEnt. As will be seen, we draw heavily on the research of Geoffrey Russom (Russom 1987, 1998, 2017, 2022, 2025), both for the necessary scholarly spadework and for matters of theory: model architecture, metrical representations, some of our constraints, and much of the theoretical rationale for the constraint system. There are in fact many theories of the *Beowulf* meter in the scholarly literature, which we have no space to cover here.[7] Our judgment is that Russom's stands out both for its completeness and its explicitness; it is unique in the degree to which it can be explicitly formalized and thus serve as the basis for MaxEnt work.

We seek to build on Russom's results by constructing the first *Beowulf* analysis that has been tested using the two criteria given in §1.3. Specifically, we quantitatively assess the degree to which the predictions made by the analysis match the frequency of existing verse types, and we require that the probabilities assigned to most or all zero-frequency items in GEN be near zero. The pursuit of these goals has led us to adopt a constraint system that goes beyond Russom's original proposals in various ways, and thus has potential implications for the theory of meter in general.

The remainder of this article is organized as follows. §2 gives background information about relevant phonological properties of Old English. §3 is a basic description of the *Beowulf* meter, with some key concepts crucial to the analysis and a discussion of our data source. §4 covers the details of MaxEnt as applied to meter. §5 and §6 begin the analytic discussion by translating some of Russom's work into MaxEnt, assessing its strengths and weaknesses. Our own proposal, is given in §7. Our model is submitted to various forms of evaluation in §8. The final section addresses a variety of open questions arising from the analysis and concludes.

2. Phonological basics

The discussion in this section confines itself to the essentials. For the broader picture, and access to the research literature, we recommend Terasawa (2011), Goering (2016), and Russom (1987).

The key linguistic elements that determine the metrical status of verses in *Beowulf* include phonological phrasing, morphological structure, syllable weight, and especially, the stress pattern, which we cover first.

---

[7] In response to a reviewer comment, we remark specifically on two alternatives that we rejected because of insufficient empirical coverage. Halle and Keyser (1971), while a breakthrough work in defining the research program, fails to cover secondary stress (missing the known differences between **Ssx** and **Sxx**) and Resolution (wrongly allowing **SxSx** verses in which both **S**'s are phonologically light syllables). Fabb and Halle (2008:264) makes a more radical (over)simplification, i.e. that the rules should ignore all stressless syllables. This neutralizes even more foot types (e.g., **S** and **Sx** are treated identically, even though **SS** is nonexistent and **SxSx** abundant).

2.1 Stress in Old English

In Old English, the first syllable of the stem of any content word bears stress, as seen in (4a,b). In transparent compounds, the stress on the second word is preserved, demoted to secondary stress, as in (4c).

(4) *Stress in Old English stems and compounds*

	OE	*IPA*	*gloss*
a.	*hund*	[ˈhʊnd]	‘dog’
b.	*dēofol*	[ˈdeəvəl]	‘devil’
c.	*ˈþēodˌcyning*	[ˈθeədˌkʏnɨŋ(g)]	‘people-king’

Prefixes show a phonemic distinction between stressless prefixes, as in (5), and stressable prefixes, as in (6).

(5) *Stressless prefixes:* a- be-, ge-

a.	*Ā-ˈrīs*	[əˈriːs]	‘arise’
b.	*be-ˈcweðan*	[bəˈkwɛðən]	‘observe’
c.	*for-ˈgetan*	[fərˈɟɛtən]	‘forget’

(6) *Stressable prefixes:* ˈand-*,* ˈofer

a.	*ˈand-ˌswaru*	[ˈandˌswɑrʉ]	‘answer’
b.	*ˈofer-ˌhȳd*	[ˈɔvərˌhyːd]	‘pride’

For further details on prefixal stress, see Minkova (2008).

Inflectional suffixes are stressless; as in *-on* ‘Pl.’ of *cunnon* in (3a) or *-um* ‘Dat. Pl.’ of *heofenum* in (7). As a first approximation, we will assume that derivational suffixes are stressless, exploring other possibilities in §9.3.

Grammatical words are generally stressless, as in (7).

(7) *Stressless grammatical words*

a. *nē on foldan fæþm* ‘nor on earth’s-bottom’ 1393a
[neː ɔn ˈfɔldən ˌfæðm]

b. ofer sǣwylmas ‘over sea-swells’ 393b
[ɔvər ˈsæːˌwʏlməs]

The literature notes subtle differences in the stress level of various grammatical words; see Kendall (1991:ch. 8-11) and Russom (2022:57-59; 2025). However, here we limit ourselves to the simplified taxonomy above: **S** is primary stress, **s** is subordinated stress, and **x** is stressless.

2.2 Syllable weight

Syllable weight in Old English follows a common pattern: syllables are heavy if they contain a long vowel or end in a consonant, else light. Syllable weight plays a role in Resolution, discussed in §3.4.

2.3 Segmental phenomena

To understand the meter in full, we must take into account several generally-recognized optional segmental processes that affect syllable count. These include, for instance, medial Syncope and syllabification of sonorant consonants. For reasons of length we place the discussion of these processes in Supplementary Materials A.

3. Metrical basics

3.1 Constituent structure of the line

The starting point for almost any metrical analysis since Kiparsky (1977) is to posit an abstract constituent structure for the line. There is a clear default pattern found cross-linguistically (Burling 1966, Piera 1980, Hayes 1983), namely a preference for a *binary* hierarchy, in which each constituent at a given level dominates two constituents at the next level down. The structure proposed for *Beowulf* by Russom, which we adopt here in its essentials, is binary at its higher levels. For explicitness we generate it with an ordinary phrase structure grammar, given in (8).

(8) *A phrase structure grammar for lines of* Beowulf

a.	Line	→	Verse Verse
b.	Verse	→	Quarterline Quarterline
c.	Quarterline	→	(Upbeats) Foot
d.	Upbeats	→	{x, xx, xxx, xxxx}
e.	Foot	→	{Sx, Ss, S, Ssx, Sxs, Sxx, Sxxs, x, xx}[8]

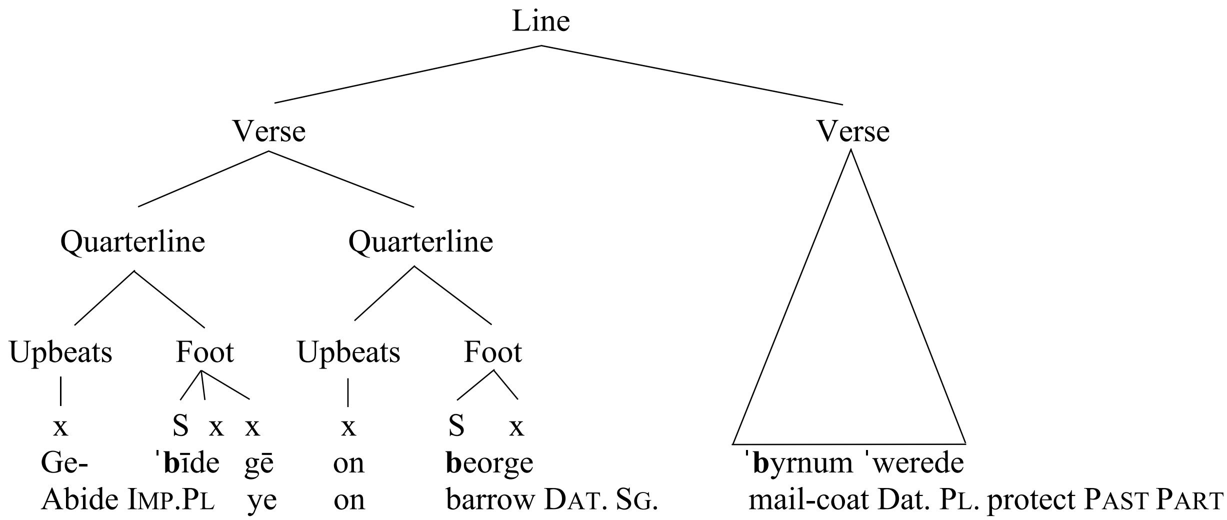
Since we will be working with a constraint-based grammar below, it is useful to construe these phrasal structure rules as inviolable constraints, e.g. "A line must always consist of two verses."

An example of the structure that can be generated by (8) is given in (9); it illustrates various options available for the expansion of the categories Upbeats and Foot.

---

[8] For discussion of this nine-member foot inventory see §6.1.

(9) *Constituency of* Beowulf *line 2529 (details for a-verse only)*



‘Remain ye on the barrow, protected by coats of mail’

The reader is cautioned that line 2529 is not typical; the norm is for Upbeat constituents to be left empty (cf. **SxSx** as the canonical verse type). We chose 2529a as an illustration because both of its Upbeat constituents are overtly realized.

As noted, we refer to the first and second verses of a line as the a-verse and b-verse. It will be seen that some of the constraints reference the location of the verse in the line, imposing greater strictness on b-verses. We also will adopt an abbreviated notation for the constituent structure of individual verses, placing brackets around the feet and separating out upbeats, if any, with a hyphen; thus the a-verse structure in (9) would be notated **x-[Sxx]x-[Sx]**.

3.2 Locating the verse and line boundaries in the text

The verses and lines found in *Beowulf* strongly tend to align with linguistic units like words and phrases (Russom 1987:15). This is unsurprising, because alignment of linguistic and metrical units is a defining characteristic of metrical verse everywhere (Lotz 1960, Kiparsky 1977, Hayes 1989a). In *Beowulf*, the alignment is strict enough that it permits scholars to identify the Line and Verse levels of (9) within the text without substantial disagreement; thus we are safe in adopting the Verse, as all others do, as the basic unit of analysis (but see Appendices F-G for the Line level). The principles of bracketing agreement will be invoked below in our discussion of the constraint *RUN-ON (§6.2).[9]

[9] For discussion of how metrical and linguistic units are aligned in *Beowulf*, see Suzuki (1996:27-35, 384-389), Stockwell and Minkova (1997), Golston and Riad (2001), Getty (2002:136-158), Minkova (2003:22-24), McCully (2022), Russom (1998:3-5, passim, 2017 §1.13, §1.16, §2.8, 2025).

3.3 Alliteration

Beowulf is an instance of alliterative verse, meaning it regulates identity between onsets of stressed syllables at various points in the line; these are marked with boldface in the verse sample of (1). For the basics, see Terasawa (2011); very briefly we can say that alliteration is satified by identity of a specific singleton onset, a prevocalic epenthetic [ʔ-], and certain clusters (sp-, st-, sk-). The locations of the alliterating onset are governed by constraints of the metrical system, as follows: (a) If the a-verse has one stress, it must alliterate; (b) if the a-verse has two stresses, then the first *must* alliterate and the second *may* alliterate; (c) the first stress of the b-verse must alliterate (i.e. using the same alliteration as the a-verse); (d) the last stress of the b-verse is a non-alliterating position.

Below, we take adherence to these principles as a given, focusing on other aspects of scansion. The *Beowulf* system does include interactions of verse type and alliteration, e.g., Bliss (1967), Kendall (1991), Hutcheson (1995), Suzuki (1996), Russom (2017, 2022), but we will not aspire to capture them here.

3.4 Resolution

The *Beowulf* poet often employs a disyllabic sequence of the form *stressed light* + *unstressed* as the metrical equivalent of monosyllabic **S** or **s**. This may be seen in (1), where *ˈsefa*, *ˈ**s**eleˈrǣˌdende* and *ˈ**h**æleð* are respectively scanned as **S**, **SSsx**, and **S**. In other words, the symbols do not denote actual syllables, but something more abstract, "positions." While the normal content of a position is a single syllable, in this special case, a position is filled by a disyllabic sequence. This pattern is called **resolution**, and is a metrical option exercised on occasion throughout in the English metrical tradition. For details, see Supplementary Materials A, and for general discussion Sievers (1893), Bliss (1967), Keyser (1969), Fulk (1992), Russom (1995), Hutcheson (1995, Ch. 3), Terasawa (2011), and Goering (2016). In what follows, when we use the term "position" we mean any one of {**S**, **s**, **x**}, irrespective of whether the **S** or **s** is filled by a resolved sequence. Resolved sequences are metrically distributed in a way that is similar to that of their counterparts **S** and **s**. A full analysis of the *Beowulf* meter would capture the differences, a task we defer to future work (§9.2). Our data compilation treats resolved sequences just like any other instances of metrical **S** or **s**.[10]

3.5 Sample scansions

To sum up our basic description, we give the metrical structures of the verse lines of (1) in (10). These differ in minor ways from Russom's structures, for reasons to be given in §4.1. We have added stress marks to the orthographic text along with brackets to show the parse of the verses into feet. For clarity, we here use $\mathbf{S_r}$ and $\mathbf{s_r}$ to denote **S** and **s** positions filled by resolved sequences; however, our analysis treats them just like any other **S** or **s**.

[10] While Resolution is not our main focus here, it seems clear that assuming something like Resolution is essential to a restrictive theory. If main-stressed light syllables are simply treated as instances of **S**, then the analysis seriously overgenerates, for instance by allowing **SxSx** verses in which both **S** positions are light. Such verses are in fact entirely absent.

(10) *Metrical structure of lines 47-52, from (1)*

Þā gȳt hīe him āˈsetton	**xxx-[xx][Sx]**	ˈsegen ˈg[yl]ˌdenne	**[S][Ssx]**
ˈhēah ofer ˈhēafod,	**[Sxx][Sx]**	lēton ˈholm ˌberan	**[xx][Ssx]**
ˈgēafon on ˈgārˌsecg;	**[Sxx][Ss]**	him wæs ˈgeōmor ˌsefa,	**[xx][Sxs$_{r}$]**
ˈmurˌnende ˈmōd.	**[Ssx][S]**	ˈMen ne ˈcunnon	**[Sx][Sx]**
ˈsecgan tō ˈsōðe,	**[Sxx][Sx]**	ˈseleˈrǣˌdende,	**[S$_{r}$][Ssx]**
ˈhæleð under ˈheofenum,	**[S$_{r}$xx][S$_{r}$x]**	hwā þǣm ˈhlæste onˌfēng.	**[xx][Sxxs]**

3.6 Data source

The analyses below are all based on the data in *Beoshare.xlsx*, an important spreadsheet shared with other scholars by Prof. Russom, in which he offers an explicit prosodic interpretation and syllabic sequence (i.e. of **S**, **s**, and **x**) of every verse in the poem. While we differ occasionally with Russom's assignment of these symbols, we nonetheless adopted them virtually without change. Our purpose in doing this was to avoid the possibility of bias; i.e. of altering the scansions on behalf of our own analysis.

We did edit our copy of *Beoshare* in minor ways. We discarded all verses (97 of them) that Russom marked with a "?" or left unscanned, indicating substantial uncertainty. It is reasonable to let the analysis be established on the basis of the clear cases, especially when these are abundant. We also excluded the 21 "hypermetrical" verses, widely agreed to represent a meter distinct from the main *Beowulf* meter.[11] Lastly, we used a different data format from Russom; we prefer to work with strings consisting of complete verses, whereas in *Beoshare* the contents of upbeat and foot positions are listed in separate columns. These edits are reflected in our own analytical spreadsheet, Supplementary Materials B.

4. MaxEnt grammars and their application to metrics

The type of MaxEnt grammar we employ here was invented for the study of phonotactics by Hayes and Wilson (2008) and applied to meter in Hayes et al. (2012). The grammar type could be called an **inventory grammar**, since it is intended to define an inventory of possible linguistic objects whose class membership may be gradient. Inventory grammars can be used in linguistics to analyze the set of words (phonotactics, morphology) or the set of sentences (syntax), as well the set of lines of metrical verse in a particular tradition. The architecture of an inventory grammar involves a GEN function (§1.3), which consists of all imaginable candidate structures in the domain under study. This is often an infinite set, but for practical purposes is it often feasible to confine one's attention to strings of a certain maximum length.

A MaxEnt grammar expresses gradient membership in the class of well-formed items for every member of GEN by assigning it a probability. An effective MaxEnt grammar will *frequency-*

[11] The hypermetrical verses appear in four clusters that interrupt the normal meter; for locations see Russom (1987:164); and for metrical analyses see Russom (1987:Ch.6) and Hartman (2020).

match the data, such that the simplest, most canonical members of GEN receive the highest probability.

For example, our own preferred MaxEnt grammar in (38) below assigns a probability of 0.223 to canonical **[Sx][Sx]** as a b-verse.[12] This roughly matches the actual frequency (0.232) with which the *Beowulf* poet uses **[Sx][Sx]** in b-verses. In contrast, for verse types thought to be unmetrical, an adequate MaxEnt grammar will assign only a vanishingly small probability. For instance, the absent verse type **[Sxs][S]** ((3c) above), judged to be unmetrical by Cable (1974:89) and Russom (2022:53), is assigned by grammar (38) a probability of 1.82×10^{-10} (as a b-verse; a is even lower). This value is so low that we judge that the grammar should be regarded as marking this verse type as essentially impossible. In fact, MaxEnt never actually assigns zero probabilities, so to use it at all we must be willing to accept vanishingly low probabilities like 10^{-10} as a prediction of unmetricality.

Lastly, the highly unusual verse type of (3b), **xx-[Sx][Ssx]-a**, is assigned by grammar (38) a probability of 0.00003; this is very low (perhaps too low), but it is still orders of magnitude greater than the value assigned to unmetrical **[Sxs][S]**.

In reporting our results we find it more intuitive to convert probabilities to predicted counts by multiplying them by the number of a- and b- verses, depending. Thus, **[Sx][Sx]-b**, assigned the probability 0.223 by grammar (38), is predicted to occur $0.223 \times 3126 = 699$ times (correct is 726); **[Sxs][S]-b** is predicted to occur 6×10^{-7} times (correct is zero), and **xx-[Sx][Ssx]-a** is predicted to occur 0.10 times (correct is 1). We will follow this method in reporting grammar predictions below.

In what follows, we flesh out the mechanisms of a MaxEnt inventory grammar as applied to *Beowulf*. We cover GEN first, then the computation of probabilities.

4.1 GEN

The GEN function provides the set of competing candidates to which probability values are assigned. We deploy our GEN in two copies, one for a-verses and one for b-verses; the grammar ends up computing two probabilities for each verse type, one as an a-verse and one as a b-verse. We also employed two different procedures for formulating GEN.

4.1.1 RUSSOM-GEN

In its purest form RUSSOM-GEN consists simply of the 2025 structures generated by phrase structure grammar (8). For actual analysis, we must pair each verse of *Beowulf* with one of these structures, where there is more than one choice. For instance, 145a *āna wið eallum* ‘one against all’ could be **[Sx]x-[Sx]** or **[Sxx][Sx]**. The pairings we use are taken from *Beoshare.* Russom assigns his pairings based on which structure best matches the linguistic bracketing of the verse; in the present case, **[Sx]x-[Sx]** is chosen because it matches the proclitic character of the syllable

[12] Probability values mentioned in this paragraph appear in the spreadsheet in Supplementary Materials E.

wið 'against'. In contrast, 2096a [*weorðode*][*weorcum*] 'honored with deeds' is paired by Russom with **[Sxx][Sx]**.

4.1.2 AUTHORS-GEN

AUTHORS-GEN is based on employing all possible strings of **S**, **s**, and **x** up to the maximum found in the poem. Such strings by themselves are insufficient, since our constraints apply not to raw strings but rather to the metrical structures generated by (8). Therefore, we must *parse* the strings into metrical structures prior to constraint evaluation.

Our parsing system works as follows. Where there is a choice, it is preferred to parse a syllable as part of a foot rather than as an upbeat; thus, for **SxxSx** our system selects **[Sxx][Sx]** over **[Sx]x-[Sx]**. Further, where **x** positions need to be parsed as upbeats, it is preferred to parse them as Upbeat1 rather than as Upbeat2; hence for **xxxxxSx**, **xxx-[xx][Sx]** is preferred to (say) **x-[xx]xx-[Sx]**. We think these choices make sense *a priori*, since it is widely agreed that upbeat syllables are a non-core, accreted element of the *Beowulf* meter and that they are better tolerated in Upbeat1; for further discussion, see §7.4.2 and §7.4.3.

In our analysis, we have adopted the RUSSOM-GEN for grammars originating with Russom (§5, §6) and AUTHORS-GEN for our own grammar (§7). This gives each model its own best-performing GEN.[13]

4.1.3 Trimbacks

To make the project computationally feasible, we simplified both versions of GEN. For AUTHORS-GEN, we exclude all candidates that have no parse under the grammar of (8). This has the same effect of including a powerful constraint that bans unparseable strings.
This reduces GEN from 19,674 to just 754 candidates.

Our second exclusion, applicable to both GEN functions, was to truncate the eight verses whose total number of {**S**, **s**, **x**} exceeds 8.[14] They are all of the form **xxxxxSxxs**, but were analyzed as **xxxxSxxs**. This facilitated the repetitious statistical computations reported in §8.3. The accuracy cost is small: running the system with the verses included, we obtained virtually identical results.

With the exclusions, AUTHORS-GEN has 754 candidates, RUSSOM-GEN has 1846. The RUSSOM-GEN is larger because, unlike in our system, it posits multiple structural representations of the same sequence of **S**'s and **x**'s, as with **SxxSx** above.

---

[13] For the "Russom-I" grammar (§5) we actually get slightly better results for AUTHORS-GEN, but since performance under either parse is very poor, the choice doesn't matter.

[14] These are: 626b, 722b, 970b, 1298b, 1461b, 1484a, 1585b, 2638b.

4.2 Constraints

A constraint is a function that inputs a candidate and outputs the number of violations. The central question of the whole analysis is what constraints one should use, and this is the subject of §5-§7 below. We assess constraint violations using simple software; this permits rapid evaluation of all candidates in GEN, and also resolves any uncertainty about whether a candidate violates a constraint.

4.3 Weights and weight-fitting

Every constraint has a weight. The MaxEnt formula, given in (11), has the effect that constraints with higher weights have a larger effect in lowering the probability of candidates that violate them.

The constraint weights could in principle be provided by the analyst, as is often done for constraint rankings in Optimality Theory. However, a more accurate grammar is usually obtained by using a machine-implemented algorithm to *fit* the weights to the data. For MaxEnt, a variety of search procedures exist that are mathematically guaranteed (Della Pietra et al. 1997) to locate the best weights. We fitted our weights in Excel, whose "Solver" utility implements the algorithm described in Fylstra et al. (1998).

4.4 The calculation of probability

The MaxEnt formula is given in the original reference for MaxEnt grammars, Goldwater and Johnson (2003, ex. 1). Suppose we have an input, some candidates, some constraints each bearing a weight, and a table indicating how many times each candidate violates each constraint. Then the probability assigned to a particular candidate x is calculated by (11):

(11) *The MaxEnt formula: probability computed for candidate* x

$$\Pr(x) = \frac{1}{Z}\exp(-\Sigma_i\, w_i \mathrm{f}_i(x)) \qquad \text{where } Z = \Sigma_j \exp(-\Sigma_i\, w_i \mathrm{f}_i(x_j))$$

$\Pr(x)$	probability of x
$\exp(y)$	e to the power of y
Σ_i	sum across constraints
w_i	weight of ith constraint
$f_i(x)$	number of times x violates the ith constraint
Σ_j	sum across candidates.

This math has been argued to be sensible and intuitive (Hayes 2022); for example it tells us that constraints differ in strength and that they can "gang up" in their effects — these are indeed effects that are claimed to be applicable in metrics by Russom (2022:38-39).

MaxEnt weights can be given a precise interpretation: for constraint C with weight w, e^w is the ratio of probabilities assigned to a minimal pair of candidates of which one violates C and the other does not. Some examples: a weight penalty of 0.69 makes a candidate half as frequent; 2.3

about a tenth as frequent; 4.6 a hundredth, and 8.7 makes a candidate 6000 times less frequent — and so likely not to show up in *Beowulf* at all. A weight of 16.1 means ten million times less frequent, which constitutes a close approximation to an outright ban (the MaxEnt formula asymptotes to zero, but never actually reaches it).

5. Applying MaxEnt analysis to the grammar of Russom (2025)

5.1 Russom's constraints

The Russom analysis of the *Beowulf* meter has evolved over time, starting with its original publication in 1987. We focus here on the latest version, Russom (2025:§§1.10-1.14), referring to the 2025 grammar as **Russom-I**. The five constraints of this grammar are as follows.

***SHORT** requires that the total number of *footed positions* be four or more. An example of a candidate that violates *SHORT is **x-[Sx]**. Note that unfooted upbeats do not count for purposes of this minimum: **xx-[xx][S]** is metrically impossible (and absent from *Beowulf*), even though its total number of positions is five.[15]

***LONG** states: "Verses with two long feet are unacceptable." A long foot is defined (p. 31) as any of **Sxx, Ssx**, **Sxs**, and **Sxxs**; i.e. consisting of at least three positions (**S**, **s**, or **x**). An example violation is **[Sxx][Ssx]**.

***HEAVY** states that "heavy [feet] are excluded from first position in long verse patterns." Heavy feet are those which have more than one **S** or **s** position (**Ss, Ssx, Sxs, Sxxs**). Long verse patterns are verses in which the feet contain a total of five or six positions (p. 31). An example violation is **[Ss][Sxx]**.

***MORPHOLOGY** excludes **Sxs** and **Sxxs** as the first foot of a verse, so for instance **[Sxs][Sx]** is a violator. For the morphological rationale of this constraint see Russom (2025:31). In §7.5.1 we offer a comparable constraint with a rhythmic rather than a morphological basis.

***OVERLAP** reflects a key metaprinciple of Russom's approach, that *ease of parsing* is important: the constraint system is designed in part to help the listener detect the metrical structure of a verse.[16] *OVERLAP militates against rival interpretations of a surface form. In particular, it forbids the foot **[Sxsx]** because it sounds too much like the verse **[Sx][Sx]**, and forbids the verse **[Sxx][S]** because it sounds too much like the foot **[Sxxs]**). Given this dual purpose it might be sensible to restate *OVERLAP as two separate constraints, though we will not bother with this task here.

---

[15] The pedigree of *SHORT goes back to Sievers (1893); for a recent defense see Pascual (2013-2014).

[16] Within phonology, a family of constraints likewise motivated by listener intelligibility is MINDIST, proposed by Flemming (2004).

The logically possible foot combinations (rows are Foot 1, columns Foot 2) are shown in (12). Mnemonic letters in the cells mark constraint violations.[17] We also added to the cells the actual counts with which the various foot pairings occur in *Beoshare*; shaded cells have zero attestations. Note that **[x]** and **[xx]** are not included as second feet; we address this gap later on.

(12) *Violations of five constraints from Russom (2025:37)*

F1: \ *F2:*	**[S]**	**[Sx]**	**[Ss]**	**[Sxx]**	**[Ssx]**	**[Sxs]**	**[Sxxs]**
[x]	S	S	S	**19**	**921**	**697**	**146**
[S]	S,O	S,O	S,O	**106**	**354**	**146**	**19**
[xx]	S	**295**	**22**	**2**	**181**	**139**	**24**
[Sx]	S,O	**2043**	**109**	**26**	**124**	**32**	**1**
[Ss]	S,O	**242**	**33**	H	H	H	H
[Sxx]	O	**77**	**18**	L	L	L	L
[Ssx]	**445**	H	H	L,H	L,H	L,H	L,H
[Sxs]	M	H,M	H,M	L,H,M	L,H,M	L,H,M	L,H,M
[Sxxs]	M	H,M	H,M	L,H,M	L,H,M	L,H,M	L,H,M

Legend: S = *SHORT, L = *LONG , H = *HEAVY, M = *MORPHOLOGY, O = *OVERLAP

In the chart, violations and positive counts are in complementary distribution, certainly an encouraging sign for the analysis. On the other hand, the violation-free cells, treated identically by the grammar, vary over a huge range of counts, from 1 to 2043, a possible portent of trouble. We can find a closer evaluation by rendering this analysis as a MaxEnt grammar and see how well it works.

5.2 MaxEnt implementation

To implement the Russom-I grammar, we computed the violation counts for all five constraints for every candidate in GEN. For example, for **[Sxx][Ssx]**, our software returns one violation for *LONG and zero for the other four. These violations are incorporated into a full tableau including all 1846 candidates. We illustrate below with an extract showing just 14 candidates: we give just the ten most frequent well-formed candidates, followed by candidates intended to illustrate violations of Russom's five constraints.[18] We also include the best-fit weights for the analysis, discussed below, and the probabilities generated.

[17] We added some violations implied by Russom's statements but absent from the original chart; these violations are redundant and do not affect the predictions of the analysis.

[18] We annotate candidates as a- or b- verses for purposes of consistency with subsequent analyses; they are immaterial here since none of the five constraints distinguish between a- and b-verses.

(13) *Partial tableau for Russom-I*

		*SHORT	*LONG	*HEAVY	*MORPHOLOGY	*OVERLAP	
	Violation counts:	0	0	0	0	0	
Candidate	*Weight:* *Frequency:*	50.00	50.00	27.67	21.20	33.83	*p*
[Sx][Sx]-b	726						0.0027
[Sx][Sx]-a	656						0.0027
[Sx] x-[Sx]-a	300						0.0027
[Ssx][S]-b	286						0.0027
x-[x][Sxs]-b	278						0.0027
[Sx] x-[Sx]-b	269						0.0027
x-[x][Ssx]-b	227						0.0027
x-[x][Ssx]-a	216						0.0027
xx-[x][Sxs]-b	186						0.0027
[S][Ssx]-b	182						0.0027
…	…						…
xx-[xx][S]-a	0	1					5.2×10^{-25}
[Sxx][Sxx]-a	0		1				5.2×10^{-25}
[Ssx][Sx]-a	0			1			2.6×10^{-15}
[Sxs][S]-a	0				1		1.7×10^{-12}
[x][Sx]-b	0					1	5.4×10^{-18}

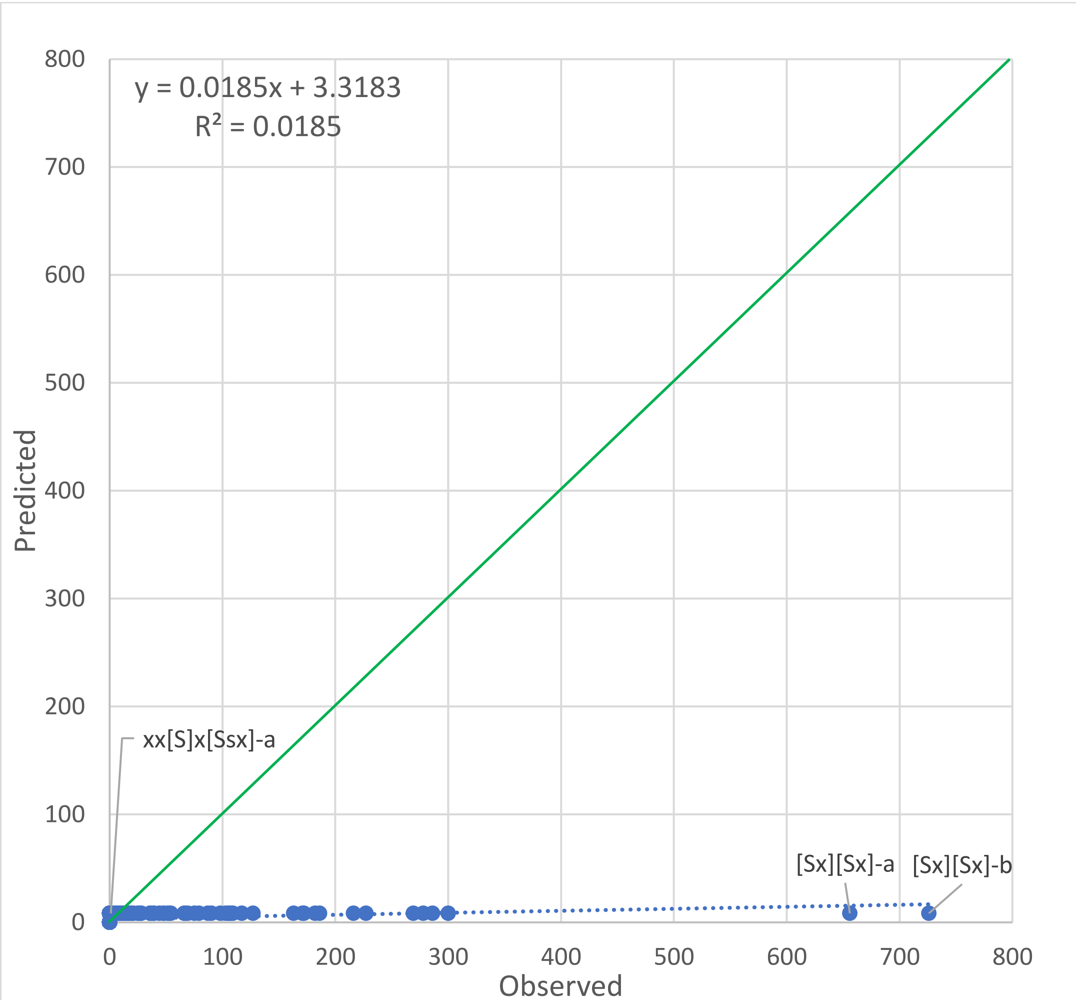
The full tableau may be viewed as a spreadsheet in Supplementary Materials C; it includes the calculations of best-fit weights and probabilities.

With this procedure, the best-fit weights for Russom-I (as found by the Excel Solver) were as given in the second row of (13). The weights of all five constraints are extremely high. These constraints are never violated in real verses, but are frequently violated in the candidates of GEN, so a good model fit is obtained with a very high weight. In fact, the theoretically optimum weight is infinite, but the Solver settles for a merely large number, which will vary from constraint to constraint due to idiosyncrasies of the weight computation.

The very high constraint weights in turn result in very low probabilities for all violating candidates — probabilities so low that by standard MaxEnt practice (§4.4) we would accept them as the practical equivalent of zero.

How well does this model perform? One way to assess this is to form a scattergram, as in (14). Here, the *x*-axis depicts the actual counts of the various verse types (a- and b-verses plotted separately), and the *y*-axis depicts the counts predicted by the analysis. A good model will place the points all close to the diagonal (in green), where predicted and observed coincide; and the linear regression line (dotted blue) will be close to the diagonal as well.

(14) *Scattergram: performance of Russom-I*



In this case, the use of a scattergram is probably overkill, but it does make plain that the model fit of Russom-I is poor, especially in comparison with the later models plotted below in (18) and (39) (for full statistical comparison see (40) below).

We offer the following diagnosis: to a close approximation, this model assigns just two probability values: candidates that violate a constraint receive a probability of very close to zero, and the remaining violation-free candidates all receive the same probability, 0.0027. These probabilities correspond to predicted frequencies of zero and 8.3. Thus all the points in (14) for the (never-penalized) attested verse types fall in a single horizontal row, while all the points for penalized candidates overlap (a single dot) at the origin.

Two of the dots on the chart correspond to cases discussed earlier. Example (3a), **[Sx][Sx]-b**, represents a verse type that occurs 726 times in our corpus and is predicted to occur about 8 times. Example (3b), with **xx-[S]x-[Ssx]**, represents a type unique in *Beowulf*, but is likewise predicted to occur about 8 times. Generative metrists have always been expected to deal with differences of this magnitude.

The Russom-I grammar can also be faulted for not doing justice to Russom's own expertise: his work includes multiple proposals that could help explain the differing data frequencies. These proposals are formalizable as violable constraints, which, deployed in MaxEnt, can produce a grammar that matches frequencies tolerably well (see next section). We judge that what is missing in Russom's work was a formal procedure for deploying violable constraints in a way that permits them to be assessed with the same rigor that he applies to inviolable constraints. Thus, in the following section, we undertake the construction of a "Russom-Augmented" grammar, incorporating new constraints based on ideas in his works, and assessing its performance against the *Beoshare* corpus.

6. A "Russom-Augmented" grammar

6.1 Constraints penalizing non-canonical feet

To begin, we return to the inventory of possible feet introduced in (8e), i.e. {**Sx**, **Ss**, **S**, **Ssx**, **Sxs**, **Sxx**, **Sxxs**, **x**, **xx**}. This list is by no means arbitrary, but reflects a fundamental principle of Russom's theory: the feet are not purely-rhythmic entities like iambs or trochees, but instead are the metrical counterparts of the inventory of possible prosodic words in the language. As Russom puts it (2025:28), "The Old English foot is abstracted from an Old English word. There is an Old English foot pattern for every native word pattern, function words and compounds included." The one exception (**[Sxsx]**) is covered by Russom's constraint *OVERLAP, discussed above. Although Russom's theory of feet may appear novel to some readers, other metrical systems likewise use a foot inventory modeled on words: Russom cites analogues in other old Germanic languages (Russom 1998) and notes similar systems in Czech (Jakobson 1938), and in Old Irish (Travis 1973, Russom 2017:41-45).

In Russom's approach, the feet have unequal status: those feet which correspond to frequent, canonical word patterns in the language serve as the predominant feet of the metrical system. In particular, the foot **[Sx]** has a special status (Russom 2025:386): "The normative Sx foot is abstracted from the normative Old English word, which has a stressed root syllable followed by an unstressed inflection. Other foot patterns are complex to the extent that they differ from the Sx norm."

We will not attempt to assess formally the degree to which other feet differ from **Sx**, but instead simply set up violable constraints that weakly penalize all non-**Sx** feet, letting their weights be determined by the data.

(15) *Eight constraints regulating foot frequency*

*[Ss], *[S], *[Ssx], *[Sxs], *[Sxx], *[Sxxs], *[x], *[xx]

6.2 Other constraints

***RUN-ON**. Our chart (12), following Russom, omits **[x]** and **[xx]** as the second foot of a verse. They are tacitly banned by some yet-unstated constraint, for which we have already seen (§3.2) an appropriate rationale, namely the characteristic agreement of prosodic and metrical bracketing in verse. The essentials are given in Russom (1987:26). In Old English, the function words are proclitic, forming a tight phonological bond with the following word. For almost any metrical tradition, breaking the bond of a clitic and its host at a verse boundary is a fairly radical metrical practice.[19] In particular, it is not tolerated in Old English, which means that putting either **[x]** or **[xx]** (which consist of function words) at the end of a verse is not tolerated. This suggests a constraint banning **[x]** and **[xx]** in verse-final position, which we will call *RUN-ON; it ends up receiving a very high weight (17).

***UPBEAT**. Russom (1987:ex. (21b)) notes, "As a kind of encrustation that obscures the underlying design, the extrametrical word [forming one or more upbeats] will add to complexity." Our constraint assesses one violation for every upbeat syllable in a verse. It treats all upbeats the same, irrespective of which quarterline or verse they occur in.[20] As it turns out, *UPBEAT receives a modest weight and thus is suitable to describing metrical complexity. Longer strings of upbeats receive the same penalty multiple times and so are predicted to be more complex.

PARSEHELP. Like *OVERLAP, PARSEHELP is founded in Russom's idea that the structure of a verse should be readily parseable. A potential parsing issue is the ambiguous status of a verse-initial sequence of one or more **x**'s: if one hears **x** or **xx** at the beginning of a verse, should it be parsed as Upbeat material or as an initial foot (**[x]** or **[xx]**)? Russom's idea (2022:60) is that the metrical system gives the listener a probabilistic cue: if the string of **x**'s is *long*, then it is likely to consist of upbeat syllables plus a stressless initial foot (**[x]** or **[xx]**), as in **xxx-[xx][Sx]**, a pattern attested 48 times. If the string is *short*, then it becomes more plausible that it could be an upbeat followed by a stressed foot, as in **x-[Sx][Sx]**, with 14 attestations.

To formalize this idea in MaxEnt, we let *UPBEAT serve as the general penalty for upbeats; then, for verses in which the first foot is **[x]** or **[xx]**, we partially *undo* the effect of *UPBEAT by assessing a credit for every pretonic **x**. In MaxEnt, the assignment of credit rather than blame can be done by using negative weights. The best-fit weight for PARSEHELP does indeed come out negative in the weight-fitting process.

***A3INB**. The relevant pattern is noted by Sievers (1893), Russom (2022:37), and Terasawa (2011:37): a b-verse is not allowed to be a token of Sievers's "Type A3". Translated into our terminology, this means that specifically in b-verses, the first foot cannot be **[x]** or **[xx]** when the second foot is **[Sx]** or **[Ss]**. Here are sample violations of this constraint: **[xx][Sx]-b**, **x-[xx][Ss]-b**, **xx-[xx][Sx]-b**. **[Sx][Sx]-b** is not a violation because its first foot is neither **[x]** nor **[xx]**; **[x][Ssx]-b** is not because its second foot is neither **[Sx]** nor **[Ss]**.

To conclude this section, we give explicit prose statements of the four new constraints:

---

[19] For instance, the mature Shakespeare strands proclitics at the end of lines (Kiparsky 1975:600-602), but only rarely; the constraint against this in Hayes et al.'s MaxEnt grammar (2012:699, 714) bears a substantial weight.

[20] Russom uses "extrametricals" and "anacrusis" to distinguish different kinds of upbeats. Since these terms have multiple usages in the literature, we prefer to avoid them, letting the relevant distinctions be stated in the constraints themselves.

(16) *Constraints added to form Russom-Augmented*

a. *RUNON. Penalize any sequence of proclitic syllables at the end of a verse (one violation per syllable).

b. *UPBEAT. Assess one violation for every syllable parsed as an upbeat.

c. PARSEHELP. If the first foot of a verse is **[x]** or **[xx]**, assess a credit for every upbeat syllable.

d. *A3INB. Assess a penalty if, in a b-verse, the first foot is either **[x]** or **[xx]**, and the second foot is either **[Sx]** or **[Ss]**.

Our Russom-Augmented grammar is formed by adding the constraints of (15) and (16) to the five already listed in (13), a total of 17 constraints.

6.3 Implementing Russom-Augmented in MaxEnt

Using the same method as before, we calculated the best-fit weights for Russom-Augmented against the data of *Beoshare* (Supplementary Materials D) and obtained the values in (17). The last two columns are for statistical testing and will be covered below.

(17) *Best-fit weights for Russom-Augmented*

Constraint	*Violation count*	*Weight*	*ΔLL*	*p*
a. *Repeated from Russom-I*				
*SHORT	0	50.00	2075.4	< 0.0001
*LONG	0	50.00	5.8	0.0007
*HEAVY	0	27.67	221.5	< 0.0001
*MORPHOLOGY	0	21.20	229.9	< 0.0001
*OVERLAP	0	33.83	44.6	< 0.0001
b. *New: Foot constraints (15)*				
*[Ss]	457	2.31	1984.8	< 0.0001
*[Sxx]	248	4.19	3305.6	< 0.0001
*[xx]	663	1.20	286.6	< 0.0001
*[x]	1783	−1.48	544.7	< 0.0001
*[S]	1090	−0.56	58.7	< 0.0001
*[Ssx]	2045	2.15	1262.9	< 0.0001
*[Sxs]	1014	2.65	1723.5	< 0.0001
*[Sxxs]	190	4.31	2790.7	< 0.0001
c. *New: from (16)*				
*RUNON	0	14.10	1227.9	< 0.0001
*UPBEAT	3524	1.59	7967.0	< 0.0001
PARSEHELP	580	−0.99	227.5	< 0.0001

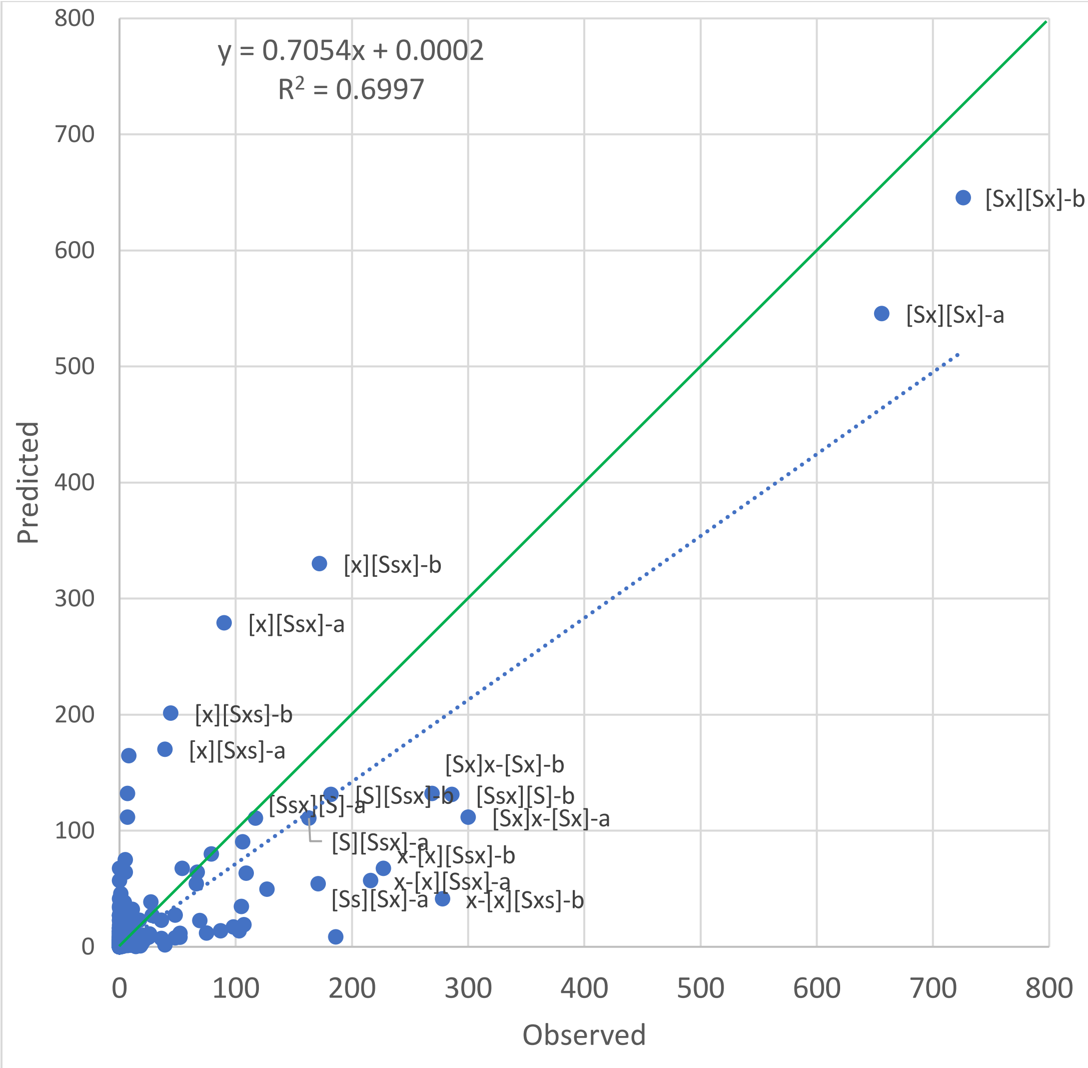
*A3inB	0	22.80	376.3	< 0.0001

The weights are what we might have expected: *RunOn and *A3inB get very high weights, since they are both inviolable and useful. *Upbeat and ParseHelp govern tendencies and get modest weights; ParseHelp as a "credit" receives a negative weight, as anticipated; and its magnitude is, crucially, lower than that of *Upbeat (else verses would counterintuitively become ever better as more upbeats are added). The weights of the constraints against non-canonical feet (17b) are likewise modest, since all feet are basically well-formed and the differences only affect metrical complexity.[21]

We again assess the performance of the model visually, by plotting a scattergram on the same basis as (14).

[21] Not too much meaning should be read into the precise weights for non-canonical foot constraints. Their general effect is to match frequency of the foot types, but in the actual weights, this is sometimes obscured by the presence of other constraints.

(18) *Scattergram: performance of Russom-Augmented*



Clearly, Russom-Augmented constitutes a big improvement over Russom-I: with violable constraints, the model can differentiate in a sensible way among existing verses. Notably, the two canonical verse types **[Sx][Sx]-a** and **[Sx][Sx]-b** are given by far the highest probabilities, and other well-attested types are generally given probabilities well above zero. If we return to our comparison of (3a) (**[Sx][Sx]-b]**) with (3b) (**xx-[S]x-Ssx]-a**), we now find that the predictions of the grammar are within shouting distance of being correct: **[Sx][Sx]-b]** is predicted to occur 646 times (actual value 726) and **xx-[S]x-[Ssx]-a** is predicted to occur 2.7 times (actual value 1). [22]

[22] To be fair to Russom-I, our Russom-Augmented includes two additional inviolable constraints, *RUNON and *A3INB. If we make a non-stochastic grammar consisting of just the seven inviolable constraints given so far, the scattergram will look almost the same as (14), except that the uniform predicted counts for violation-free verses will rise to about 10.6 for a-verses and 11.8 for b-verses.

6.4 Statistical testing

It is normal to assess a MaxEnt model's constraints with a statistical test, checking whether they contributes significantly to the accuracy of the outcome. Following Hayes et al. (2012), we employed a likelihood ratio test (Field 2026:982), based on the difference in log likelihood between two grammars, one of which is a subset of the other. We formed 17 subset grammars by deleting one constraint at a time from the full set, then calculating the change in log likelihood that results (ΔLL in (17) above). The test gives the probability that a constraint could have the effect that we see simply at random (because the text of *Beowulf* is only a sampling of the full population of lines generated by the poet's grammar). As (17) shows, all 17 constraints pass this test.

We note in passing that this outcome is not a forgone conclusion. Both in mining Russom's work for constraints, and in trying to devise our own, we have encountered surprises in the form of constraints that we *thought* would help the analysis, but turned out not to test as significant — or even came out with zero weight, indicating uselessness. To give one example, the constraint *LAPSE (here formulated as banning two **x** in a row) has a strong pedigree in both phonology (Prince 1983, Selkirk 1984 et seq.) and metrics (Golston and Riad 1997), so it is a perfectly reasonable hypothesis to consider. But when we added it to our proposed grammar (38) and computed its best-fit weight, it turns out to be zero — all the useful analytic work that *LAPSE might have done is already better done by existing constraints.

6.5 Prospect

We have found that the Russom-Augmented grammar could be improved further by formalizing descriptive generalizations found elsewhere in Russom's work. We do not report these findings here, because we find the further we proceed, the more outré our adaptations become, and the more we risk "putting words in Russom's mouth." Instead, we present our own grammar, to be called AUTHORS-Grammar. This grammar draws on Russom's, but we also appeal to other ways to ground our constraints in general principles.

7. Developing a new grammar

7.1 What is a principled constraint?

A widely shared goal of generative metrics (Russom 1987, 2025; Hanson and Kiparsky 1996; Hayes et al. 2012; Blumenfeld 2015, 2016) is that all the constraints should be grounded in metaprinciples that regulate what can be a constraint. We invoke four.

End Weight. Heavy entities (e.g. heavy syllables, polysyllabic words, long metrical units) are placed later. This principle is acknowledged by Russom (2025:384) and studied comprehensively in Ryan (2019). The constraints based on End Weight in AUTHORS-Grammar are (36a-c) below.

Closure. Hayes (1983:373) states this as "Correspondence to a metrical pattern tends to be lax at the beginnings of units; strict at the ends." The principle dates back at least to Smith (1968) and has an extensive literature; e.g. Ryan (2017) on Latin and Old Norse; and Strang (1970:326), McCully (1996, 2022), and Russom (2017:Ch.1; 2025:384) on Old English. In the present context, the key structure is (9): closure implies that b-verses could be subject to tighter restrictions than a-verses. A simple form of evidence for closure in *Beowulf* is that **[Sx][Sx]**, the canonical verse, is more common in b than in a (726 vs. 656). The reason, we suggest, is that many other members of GEN are penalized more harshly in b; much of the probability mass thus freed up is awarded to **[Sx][Sx]**. The closure-based constraints in AUTHORS-Grammar are (30a-b), (32c), and (37a-b).

Rhythmic Simplicity. This idea has an extensive history in metrics; see e.g. Kiparsky (1977), Piera (1980), Prince (1989), and for a broader perspective Lerdahl and Jackendoff (1983). In simple rhythmic forms, we expect even spacing of beats, preference for binary over ternary intervals (**[Sx]** better than **[Sxx]**), and avoidance of extrametrical material (here, upbeats). Some constraints based on Rhythmic Simplicity that we propose below are (30a-b), (19), and (34).

Optimal Length. Metrical units can have preferred lengths independent of their internal structure. See definition and discussion in §7.4.1. Specific cases are (22), (28a, b), and (32a-c).

These principles have guided us in the construction of our own grammar. Of the 25 constraints of our grammar (listed below in (38)), eleven are retained from Russom-Augmented (17) (dropping six), and fourteen are new.

7.2 Constraints inherited from Russom

The retained constraints include the eight foot constraints discussed above under (15), as well as the following three.

(20) *RUN-ON (16a)

(21) *OVERLAP (§5.1)

*OVERLAP requires that we also retain Russom's metaprinciple (§5.1) concerning parseability.

(22) *SHORT (§5.1) — renamed *SHORTFOOTSEQUENCE

To be clear on how violations of *SHORTFOOTSEQUENCE are counted, we assign it the following definition: total the positions in the two feet (i.e. excluding upbeats). If the total is less than four, assign violations corresponding to the shortfall. The constraint is renamed in order to distinguish it from (28a) below.

7.3 Moraic structure

We suggest that the effects of both End Weight and Optimal Length emerge most clearly if we adopt a theory of **moraic structure** (Hyman 1985, Hayes 1989b). Moras have been used to elucidate Old English meter in a variety of ways (Sievers 1893; Golston and Riad 2001; Cooper 2017, 2019); our own approach differs from all of them. For us, the moraic structure is a *metrical* interpretation of the strings of **S**, **s**, **x** used for scansion; specifically, **S** and **s** receive two moras, **x** receives one.

These assignments are related to the moraic structure of Old English phonology (Lass and Anderson 1975:§1.3, Hogg 1992; Goering 2016, 2023), yet they are not identical. **S** and **s** positions are most often filled with a phonologically bimoraic syllable, or else with a resolved sequence, of which the first syllable must be phonologically monomoraic (§3.4). However, our system treats **S** and **s** as metrically bimoraic, even in the exceptional cases where phonologically they are monomoraic.[23] Our moraic system also mismatches the phonology in treating stressless syllables as light even when they are phonologically bimoraic (*ūt* 'out', *mīn* 'my'). We are not the first to suggest that a metrical system can employ weight criteria different from the phonology; see Kiparsky (1989:313-14), Golston (2021), and Ryan (2019: Ch.4, 2022 and references there).

Following these principles, the mora counts for the nine foot-types in our analysis emerge as in (23).

(23) *Assigning mora counts to the nine foot types*

Foot	*Mora count*	*Foot*	*Mora count*
[x]	1	**[Sxx]**	4
[xx]	2	**[Ssx]**	5
[S]	2	**[Sxs]**	5
[Sx]	3	**[Sxxs]**	6
[Ss]	3-4		

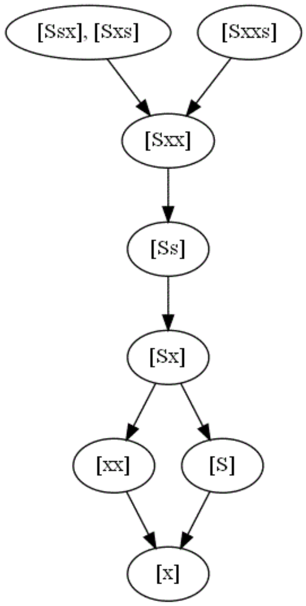
The evidence supporting hierarchy (23) comes from the fact that feet assigned different mora counts are distributed differently. The detailed arguments are given in table (24), with the overall results summarized in the Hasse diagram of (25).

[23] For example, in **[Ssx]** feet, **s** is often a phonologically light syllable, as in 2197a *on ðām lēod**scipe*** 'in that country'. Yet, according to our test of upbeat counts (§7.4), **[Ssx]** with phonologically light **s** behaves just like **[Ssx]** with phonologically heavy **s** — i.e. as pentamoraic.

(24) *The arguments for assigning mora counts to feet*

Foot	*Assumed mora count*	*Comparison*	*Basis*
a. **[x]**	1	Lighter than **[S]**	No direct argument, but "subtracting equals from both sides": **[Sx]** > **[xx]**, therefore **[S]** > **[x]**.
		Lighter than **[xx]**	No direct argument, but subtracting equals from both sides: **[Sxx]** > **[Sx]**, therefore **[xx]** > **[x]**.
b. **[xx]**	2	Lighter than **[Sx]**	The verse **[xx][Sx]** is very rare (8 cases); the verse **[Sx][Sx]** is common (1382 cases); the former is not long enough to form a normal verse. See (28a) *SHORTVERSE.
		Lighter than **[Sx]**	In b-verses, **[xx]** is abundant before **[Ssx]**, **[Sxs]**, and **[Sxxs]** (1089 cases). **[Sx]** is rare in this context (11 cases). See (32c) *IMBALANCE2-4-B.
c. **[S]**	2	Lighter than **[Sx]**	In b-verses, **[S]** can precede **[Sxx]**, **[Ssx]**, **[Sxs]**, **[Sxxs]** (371 cases), but **[Sx]** is rare in this context (12 cases). See (32c) *IMBALANCE2-4-B.
		Lighter than **[Sx]**	**[S]** can pair with preceding **[Ssx]** (445 cases); for **[Sx]** this is rare (4 cases). See (32a) *IMBALANCE4-3.
d. **[Sx]**	3	Lighter than **[Sxx]**	**[Sx]** is common (205 cases) before the heavy feet **[Sxx]**, **[Ssx]**, **[Sxs]**, **[Sxxs]**; **[Sxx]** is rare in this context (5 cases). See (31b) *IMBALANCE3-4.
e. **[Ss]**	3	Lighter than **[Sxx]**, **[Ssx]**, **[Sxs]**, **[Sxxs]**	Can pair with a preceding **[Sxx]** (55 cases) or **[Ss]** (25 cases). For the four comparison feet the totals are 5 and 0 respectively. See (32b) *IMBALANCE3-4.
	3	Same as **[Sx]**	It takes same number of upbeat syllables; see (26a) and constraints (28a,b) *SHORTVERSE, *LONGVERSE.
	3	Same as **[Sx]**	Cannot be preceded by **[xx]** in b-verses; i.e. the **s** of **[Ss]** doesn't count as an **s** for (36a) B-VERSEHAS2S.
	3	Same as **[Sx]**	**[xx][Ss]** verses (a or b) are very rare (1), just like **[xx][Sx]** verses (8), explainable if both are pentamoraic. See (28a) *SHORTVERSE.
	but: 4	Heavier than **[Sx]**	Cannot pair with following **[Sxx]**, **[Ssx]**, **[Sxs]**, **[Sxxs]** (zero vs. 205 cases). See (32b) *IMBALANCE3-4.
f. **[Sxx]**	4	Lighter than **[Ssx]**, **[Sxs]**, **[Sxxs]**	**[Sxx]** often occurs as Foot 1 to **[Sx]** or **[Ss]** (779 cases), unlike heavier feet (12 cases total). See (32a) *IMBALANCE4-3.
g. **[Ssx]**, **[Sxs]**	5	Lighter than **[Sxxs]**	No direct argument, but "adding equals to both sides": **[Ss]** < **[Sxs]**, therefore **[Sxs]** < **[Sxxs]**; also **[Sx]** < **[Sxx]**, therefore **[Sxs]** < **[Sxxs]**.
h. **[Sxxs]**	6		

(25) *Hasse diagram of relative weight*



Clearly, the **[Ss]** foot is aberrant: it acts as tetramoraic in one context, but mostly as trimoraic. We can note the following generalization: (24e) shows that [Ss] is trimoraic as a second foot and tetramoraic as a first foot. Earlier work (e.g. Fulk et al. 2008:Appendix C; Russom 2017:§3.6) relates the puzzle of [Ss] to the fact that **Ss** compounds tend to reduce their clashing second member to **x**. For instance, *ful**tum*** [ˈfʊltʊm] 'help, army force' is a reduced form of *ful* [ˈfʊl] 'full' + *tēam* [ˈtɛəm] 'team'. Other than noting these clues, we have no principled account of **[Ss]** behavior and will resort to ad hoc measures in the analysis below.

7.4 The distribution of upbeats

Turning to the actual application of moras in our analysis, we put forth what we believe to be a novel empirical generalization, governing verses in which the first foot is either **[x]** or **[xx]**. Because of other constraints, the second foot will necessarily be one of **[Sx]**, **[Ss]**, **[Sxx]**, **[Sxs]**, **[Ssx]**, or **[Sxxs]**.[24] Further, we include the cases in which upbeats appear before the first foot. We set up the descriptive term *pretonic sequence* to mean the concatenation of the upbeats and the stressless first foot. For example, in **xx-[xx][Ssx]**, the pretonic sequence has four syllables.

The empirical question is whether there is any connection between the content of the second foot and the typical length of the pretonic sequence. This can be checked by forming a graph that counts the number of verses that have a pretonic sequence of any particular length, with a separate line on the graph for each of the six stressed foot types. To match the scaling of the

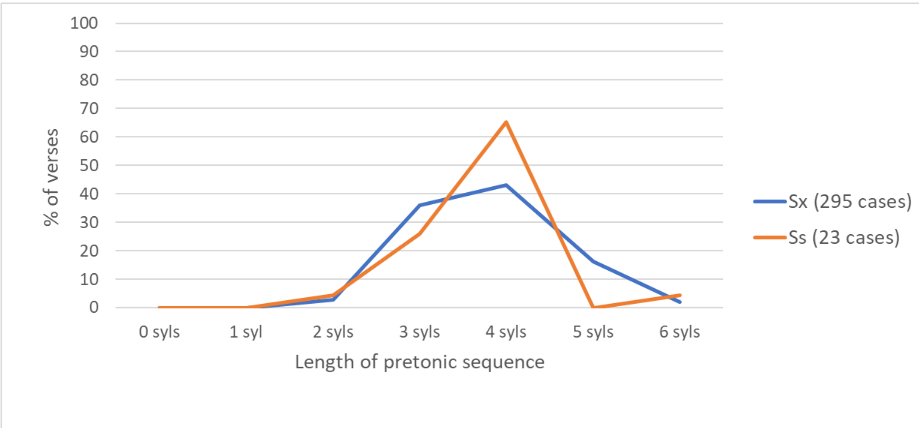
---

[24] [x] and [xx] are not allowed as second feet because of *RUN-ON. [S] is not allowed because combining it with [xx] or [x] violates *SHORTFOOTSEQUENCE. Lastly, [Sxx] is rare as a second foot in this context (see (36c) below), but the 21 cases pattern the same as the other long feet and so we include it here.

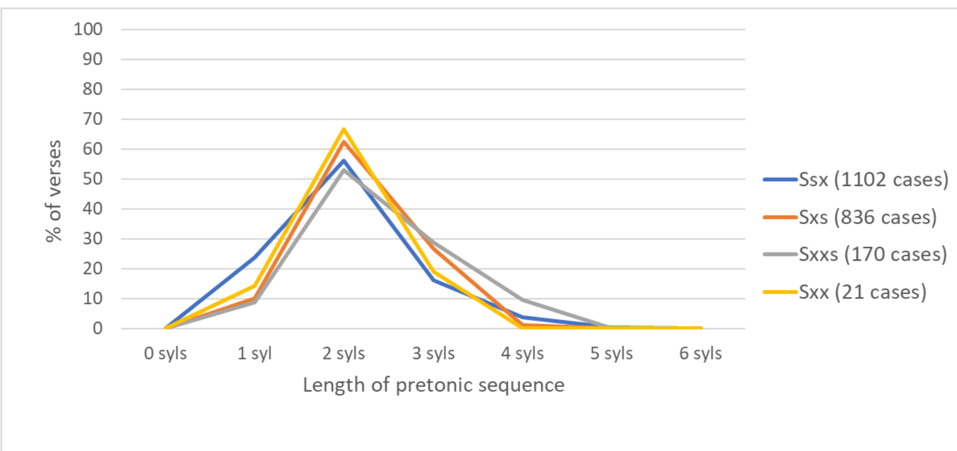
various cases we give percentages rather than raw values, with the totals given in the legend. Looking ahead to our conclusion, we give not one single graph, but rather divide the plots for different second feet according to our proposed mora counts.

(26) *Length of pretonic sequence in syllables, depending on second foot*

a. Shorter second feet (**[Sx]**, **[Ss]**)



b. Longer second feet (**[Sxx]**, **[Ssx]**, **[Sxs]**, **[Sxxs]**)



Clearly, the shorter feet behave differently from the longer ones: short feet have, on the average, about four syllables in the pretonic sequence (3.81), whereas the longer feet have about two (2.11). The difference in the distributions is highly significant (one-tailed Mann–Whitney U test, U = 48,293.5, $p < .001$). There seems to be a tradeoff: the greater length of the longer feet is compensated by a shorter pretonic sequence.

It is reasonable to construe this compensation in moraic terms: the feet **[Sxs]** and **[Ssx]** prefer, roughly, two fewer pretonic syllables than **[Sx]**, which makes sense in terms of moraic equivalence: the extra moras in the feet are balanced by fewer moras in the pretonic sequence. We can fit **[Ss]** into the system under the assumption that this is one of the contexts (see (24e)) where **[Ss]** behaves like three moras. Note that the small difference in the distribution of upbeat counts for **[Sx]** and **[Ss]** does not test as significant (one-tailed Mann–Whitney U test, U = 3324.5, p = 0.432).

To be sure, the trade-offs are not observed at the finest-grained level; strict balance would predict that tetramoraic **[Sxx]** and hexamoraic **[Sxxs]** should have different distributions from pentamoraic **[Ssx]** and **[Sxs]**. In principle we could complicate the analysis below (see (28a,b)) to cover these details, but as will be seen, we have opted for analytic simplicity.

The next step is to account for the pattern of (26) in the formalized AUTHORS-Grammar. First, we need to cover the concept of **optimal-length constraints**.

7.4.1 Optimal-length constraints

In many metrical systems, the length of a line is *deducible from its structure* (i.e. tree structure, as in (9)). For example, imagine a system in which a line consists of two cola, a colon of two feet, and a foot of two syllables. Here, the length of the line is deducible as eight syllables (2 × 2 × 2) — but nothing in the formal description need actually count to eight; it is simply the product of the number of daughters found at each level. One need only count up to two, and the rest is multiplication.

However, not all length patterns are deducible in this way. For example, it has been shown that the lines in some pentameter systems combine a two-foot colon and a three-foot colon, in either order; for Spanish see Piera (1980), for Modern English Hayes (1988), for Middle English Youmans and Li (2002), for Italian and Occitan Duffell (2018:121-126). However, the cola cannot be combined *freely*; if they are allowed to vary between two and three feet context-freely, we would derive *3+3 and *2+2. It follows that the required line length of five feet must result from a constraint that actually counts out five units.[25]

Here is a second example. For French verse, it has been argued by Biggs (1996) that the six-syllable half-line used in the decasyllable and the alexandrine has two structures: either three iambic feet or two anapests; again, the total of six syllables must result from its own constraint and is not deducible from the hierarchy.

Where a constraint specifically governs the required length of a higher-level constituent, rather than letting it be deduced from its content, we will call it an **optimal-length constraint**. We assume that the possibility of such idiom-specific constraints should be included in the general theory of meter.

Since Sievers accounts of *Beowulf* have often employed optimal-length constraints: the verse is required to have at least four positions irrespective of constituency: all three of **[S][Ssx]**, **[Sx][Sx]**, and **[Ssx][S]** satisfy this minimum. Russom (2022) obtained this generalization without

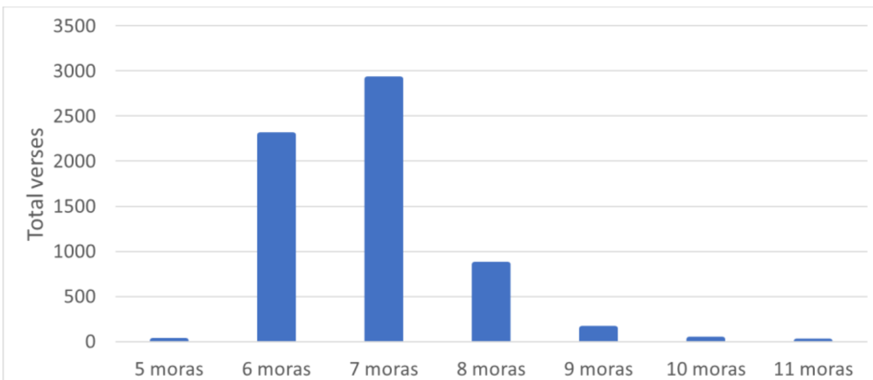
[25] Phonologists have long resisted constraints that count higher than two or three; see e.g. McCarthy (2003), and for a skeptical view Paster (2019). In any event there is no guarantee that metrics respects the same limits as phonology. As a reviewer notes, both Paster's examples and ours count suprasegmental units.

actually employing an optimal-length constraint; Russom (2025) rejoins the mainstream with the optimal-length constraint *SHORT, retained here as *SHORTFOOTSEQUENCE.

7.4.2 Upbeat counts as regulated by optimal-length constraints

We next apply the concept of optimal-length constraints to address the novel empirical generalizations about upbeat counts given in §7.4; i.e. that upbeat counts are in a trade-off relation with the length of the following foot. Our suggestion is that these reflect optimal-length constraints defined *at the verse level*: there is a desired target for verse length, defined in mora count, and the *Beowulf* poet tacitly manipulated upbeats so as to approximate this target statistically. The pattern can be clearly seen when we plot all 6364 verses by their mora count, as in (27):

(27) *Histogram: the number of verses displaying each possible mora count*



From (27), it might seem that the desired mora count should be seven moras. Interestingly, since the constraints interact in subtle ways, it turns out to work almost as well to let the target be six moras, and we adopt it for reasons of intuitiveness and elegance, for six is the length in moras of the default *Beowulf* verse type, **[Sx][Sx]**.[26]

We find that a more accurate grammar can be obtained if we separately penalize verses that have too many moras and those that have too few. The latter seems to be the more serious metrical fault, as the weights in (38) will demonstrate. We posit the two constraints in (28).

(28) a. *SHORTVERSE

Assess a penalty for every mora by which a verse falls short of length six.

b. *LONGVERSE

Assess a penalty for every mora by which a verse exceeds length six.

[26] For an earlier proposal that the verse of *Beowulf* reflects a six-mora optimum, see Hanna (2010). Cooper (2017, 2019), working on non-Beowulfian Old English meter, suggests that a prototypical *line* has 12 moras.

The primary justification for these constraints is that they can account for the patterning of upbeats, but they have effects elsewhere: *SHORTVERSE helps account for the rarity of pentamoraic **[xx][Sx]** as a verse type; and *LONGVERSE assists the constraints of (32) below in penalizing conjunctions of heavy feet.

Our treatment extends a key insight of Russom: the *Beowulf* poet imposes length requirements at more than one level. By design, constraint (22) *SHORTFOOTSEQUENCE (Russom's *SHORT) counts only syllables *within a foot*. However, in studying upbeats, we found that there are also constraints that regulate the length of *the whole verse*; these include *SHORTVERSE. The different bases of these constraints is illustrated in (29); weights are taken from (38) below.

(29) *SHORTVERSE vs. *SHORTFOOTSEQUENCE

Candidate	*SHORTVERSE (*weight* 3.22) "Verses should not fall short of six moras."	*SHORTFOOTSEQUENCE (*weight* 13.96) "The feet together must not fall short of four positions."	*Frequency*
xx-[xx][S]		1	0
[xx][Sx]	1		8

The candidate **xx-[xx][S]** fully obeys *SHORTVERSE, since it has six moras. Yet it is completely impossible for *Beowulf*, for it has just three foot-internal syllables and thus violates the powerful *SHORTFOOTSEQUENCE. In contrast, **[xx][Sx]** complies with *SHORTFOOT SEQUENCE, but has only five moras, violating *SHORTVERSE. It is a *possible* verse type, but owing to the *SHORTVERSE violation, it is rare (8 cases). Note that adding upbeats to **[xx][Sx]** produces much better candidates: **x-[xx][Sx]** occurs 106 times, **xx-[xx][Sx]** 127 times.

In sum, to understand the *Beowulf* meter it is essential to distinguish constraints on *verses* from constraints on *foot pairings* — both are needed.

7.4.3 Further constraints on upbeats

Our constraints *SHORTVERSE and *LONGVERSE regulate the overall count of upbeats. But they are indifferent to where these upbeats get placed, and for this purpose further constraints are needed. We suggest that these further constraints reflect the metaprinciple of Closure (§7.1): an upbeat is more disruptive when it is an upbeat to the second foot (i.e. occurs in the second quarter-line; see (9)). It is more disruptive still if the violation occurs in a b-verse.[27] The necessary constraints are given in (30).

[27] Here are our counts. a-Verse, Upbeat 1: 827; a-Verse, Upbeat 2: 81; b-Verse, Upbeat 1: 424; b-Verse, Upbeat 2: 44.

(30) a. *UPBEAT2

Assess a penalty for every upbeat in the second quarter-line of a verse.

b. *UPBEAT2INB

Assess a penalty for every upbeat in the second quarter-line of a b-verse.

In the tableau calculations, these constraints "gang" in the sense of Jäger and Rosenbach (2006): an upbeat in the second quarter line of a b-verse incurs the summed penalties of both constraints.

Additionally, we penalize one further configuration, namely upbeats occurring before a stressed Foot1. As Russom (2017:63) observed, these are rare: by our count only 92 such syllables are found.

(31) *PRETONICUPBEAT1

Assess a penalty for every upbeat syllable preceding a stressed Foot1.

The rarity of upbeats in this position has long been recognized in the literature (Sievers 1893, Cable 1971:32-45).

7.5 Constraints on foot pairings

We turn now to the use of moras to govern foot pairing. Russom's earliest work (1987) argued for a "trade-off" pattern: a long foot on one side of a verse tends to be compensated by a shorter foot on the other side. In his 2025 analysis (§5.1), this length trade-off derives from the combined action of *SHORT and *LONG. Our account largely follows Russom, but seeks to treat the data gradiently. Indeed, we find that a better fit is possible by assuming not inviolable constraints, as with Russom, but a set of quite-strong but not inviolable constraints. When combined, these suffice to assign zero predictions to quite a few cells of the chart of foot combinations; but when just one of the foot-pairing constraints is violated, we find that a few examples actually occur.

For our constraints we use the same system of moras that we used above to regulate upbeat counts. Specifically, our balance constraints take the form "If Foot1 has *n* or more moras, then Foot2 must have fewer than *m* moras." They are explicitly stated in (32), where μ(Foot) designates a foot's mora count. Each constraint assigns one violation if its requirement is not met. The patterning of these constraints in the data is given in (33).

(32) *Imbalance constraints*

a. *IMBALANCE4-3

If μ(Foot1) > 4, then μ(Foot2) must be < 3.

See red box in (33).

b. *IMBALANCE3-4

If μ(Foot1) > 3, then μ(Foot2) must be < 4.

For purposes of this constraint, **[Ss]** as Foot1 is counted as 4 moras.

Blue box in (33).

c. *IMBALANCE2-4-B

In a b-verse, if μ(Foot1) > 2, then μ(Foot2) must be < 4.

Orange box in (33).

(33) *Scope of the* IMBALANCE *constraints*

		x	**S**	**xx**	**Sx**	**Ss**	**Sxx**	**Ssx**	**Sxs**	**Sxxs**
		1	*2*	*2*	*3*	*3*	*4*	*5*	*5*	*6*
x	*1*	2	3	3	4	4	5	6	6	7
S	*2*	3	4	4	5	5	6	7	7	8
xx	*2*	3	4	4	5	5	6	7	7	8
Sx	*3*	4	5	5	6	6	7	8	8	9
Ss	***4***	4	5	5	6	6	7	8	8	9
Sxx	*4*	5	6	6	7	7	8	9	9	10
Ssx	*5*	6	7	7	8	8	9	10	10	11
Sxs	*5*	6	7	7	8	8	9	10	10	11
Sxxs	*6*	7	8	8	9	9	10	11	11	12

The *IMBALANCE constraints are at the heart of our analysis and require commentary.

No directionality. The constraints are stated in "left-to-right" fashion, i.e. "If Foot1 … then Foot2 …". This is arbitrary; we could just as well have stated them right-to-left.

Not statable as a sum. The *IMBALANCE constraints define a detailed pattern over the logical space of foot combinations. In rough terms, this pattern is focused on combinations that have an excessively large *total* number of moras, as in the lower right region of the chart. But our actual analysis is stated on specific natural classes of foot combinations, hugging the empirical ground. It can be observed in particular that the total of 7 sometimes occurs inside our boxes, sometimes outside them; and the same is true for 8.

Moreover, we have explicitly tried the alternative of invoking a sum; i.e. a constraint against the moras of the two feet jointly exceeding some threshold. This turns out to perform far less well.[28]

[28] The alternative that works best is to penalize every mora which causes the sum to exceed seven. Such a grammar receives a log probability score of −18599.6, far less accurate than the −18040.9 received by a comparable grammar employing the *IMBALANCE constraints of (32).

In other words, although our strategy of taking a total sum of moras worked well for regulating overall verse length (§7.4.2), in the domain of foot-pairing, it is better to follow Russom's original approach, setting up constraints that penalize particular combinations of feet.

Imbalance and closure. The orange box for IMBALANCE2-4-B extends over more territory than its partner, *IMBALANCE3-4 (blue box; applicable to both verses). This is arguably a closure effect, a stricter ban for line-final verses.

Treatment of [Ss]. The obvious "sore thumb" of the analysis is our fiat declaration in (32b) that in this one single context, **[Ss]** behaves as if it were tetramoraic; otherwise, in the four contexts where it makes a difference, it behaves as trimoraic (see (24e)). We have experimented with avoiding the "sore thumb" by assigning **[Ss]** a fractional moraic value, like 3.5. With appropriate integer cutoffs for the IMBALANCE constraints, this works empirically, and overcomes the **[Ss]** weight contradiction. However, we are reluctant to advance the idea of fractional moras with just these data.

As noted above, our analysis attempts to go beyond Russom's results by predicting not just the foot combinations that are absent, but also the combinations that are rare. This is possible because the constraints of (32) have "medium-strong" values: when just one is applicable, the analysis predicts rarity, but when two are applicable, the analysis predicts (essentially) absence. Here are examples. The foot pairings **[Sx][Sxx]-b**, **[Sx][Ssx]-b**, **[Sx][Sxs]-b**, and **[Sx][Sxxs]-b** violate only *IMBALANCE2-4-B, which bears a medium-strong weight (3.69). As such, it correctly allows violations, but only rarely; there are 12 total verses of these four types and the analysis predicts 11.9. The foot pairings **[Ssx][Sx]**, **[Ssx][Ss], [Sxs][Sx]**, **[Sxs][Ss]**, **[Sxxs][Sx]**, and **[Sxxs][Ss]** violate only *IMBALANCE4-3, weight 3.68. Here again violations are possible but rare (12 total, predicted 11.9). Yet the analysis continues to make near-absolute predictions where appropriate. Any verse pairing that violates *more than one* of (32a-c) will receive a harmony penalty of at least 7.37. None of these pairings is attested at all, and none receive a predicted count any higher than 0.05. In sum, thanks to the principle of constraint ganging, our mora-based extension of Russom's balance system not only continues to predict what verse-pairings will be absent, but also covers the low-frequency fringe.

7.5.1 Iambic clash

A puzzling mystery in the *Beowulf* meter revolves around the verse type Sievers called "E", **[Ssx][S]**. In our charts in (35) below, which give the attestations of all foot combinations, **[Ssx][S]** is *surrounded* by impossible combinations — why should it be well-formed? The answer presumably resides in a constraint set that somehow rules out everything surrounding **[Ssx][S]**, but not **[Ssx][S]** itself.

Most of the needed constraints are already in place. **[Ssx]** as a first foot cannot pair with a second foot heavier than **[S]** because of *IMBALANCE4-3. It cannot pair with **[x]** or **[xx]** because of *RUN-ON. Going in the other direction, **[S]** as a second foot cannot be paired with the first feet **[x]**, **[S]**, **[xx]**, **[Sx]**, or **[Ss]** because of *SHORTFOOTSEQUENCE, and it cannot be paired with **[Sxx]** because of *OVERLAP. This leaves just two remaining cases to deal with, namely **[Sxs][S]** and its near-neighbor **[Sxxs][S]**. For Russom (2025), these are ruled out by *MORPHOLOGY (§5.1).

While this constraint would serve our purpose, we think it worth exploring an alternative constraint motivated on rhythmic grounds.

Our attention is attracted to the particular sequence **xsS** that these bad verses contain: *stressless*, *weak stress*, *strong stress*. We will call this sequence an **iambic clash**, since the **xs** part of the sequence has an iambic stress pattern and its **s** clashes with the following **S**. Several observations suggest that iambic clash is a particularly dysrhythmic sequence. First, it is strongly avoided in modern English verse — strikingly, poets avoid sequences like *intènse líght* irrespective of how it is scanned: it is bad in both SWS position and WSW position.[29] Further, while dactylic and anapestic verse often contain **Ssx** sequences, **Sxs** is absent (Kiparsky 1977, Prince 1989); thus while *Oh say can you see by the dáwn's eàrly líght* with **Ssx** is normal, Kiparsky's construct ... *by the dáwn's intènse líght*, with **xsS**, represents a token of a verse type that apparently does not exist; we suggest that the reason is that iterated **Sxs** creates a sequence of iambic clashes: **S x s S x s S x s S x s** …. Violations of iambic clash that arise in ordinary English speech are often repaired phonologically by the "Rhythm Rule" (??*thirtèen mén* = **xsS** → *thìrteen mén* = **sxS**; Liberman and Prince 1977). Lastly, Breiss and Hayes (2022) argue that iambic clash violations tend to be avoided by English speakers when they form words into utterances: while sequences like *intènse líght* are certainly grammatical in English, they occur less frequently than would be expected on statistical grounds.

The above is our defense of the view that iambic clashes are dysrhythmic, and we conjecture that avoidance of such clashes is encoded as part of the metrical grammar for *Beowulf*, as in (34):

(34) *IAMBICCLASH

Penalize verses that contain the sequence **xsS**.

Our analysis roughly follows Cable (1991:147-51), who was the first to suggest clash as a basis for banning ***[Sxs][S]** and ***[Sxxs][S]**. Cable does not specifically mention *iambic* clash, defining it merely as ***sS**. Our own ***xsS** is more accurate, since **[Ss][Sx]** and **[Ss][Ss]** are both metrical verse types.

7.5.2 B-VERSEHAS2S

It will be helpful to preview a constraint here outside its logical context. The constraint B-VERSEHAS2S penalizes any b-verse with just one of the symbols {**S**, **s**}. As elsewhere, the **s** of **[Ss]** must be counted as acting like **w** (see Russom 2017:§3.11). The constraint penalizes b-verses such as **[xx][Sxx]**. We cover this constraint in §7.5.4 below.

7.5.3 Full coverage of well-formed foot pairings

We are now in a position to create charts analogous to (12), used by Russom (2022) to illustrate his grammar. We will deploy two charts, since a and b verses respect different constraints; each

---

[29] The observation is made for James Thomson's verse by Hammond (1991:248). In the 2000-line samples employed by Hayes et al. (2012) there are no instances in Milton and just one (Sonnet 107:8) in Shakespeare.

includes all 81 logically possible foot-pairings and how they are dealt with in the analysis. We include corpus counts, so we can assess constraints that are violated, but only rarely. In the full AUTHORS-Grammar ((38) below), the rarely-violated constraints receive medium-strong weights.

Our charts, like Russom's, are intended to cover only foot-pairings, abstracting away from the upbeats. In our full grammar, however, upbeats are dealt with (§7.4.2)

(35) *How the proposed analysis regulates foot pairings*

a. *Legend*

	Reference	*Weight, from* (38)
a = *OVERLAP	§5.1	26.17
b = *RUN-ON	(16a)	33.11
c = *SHORTFOOTSEQUENCE	(22)	13.96
d = *IMBALANCE4-3	(32a)	3.68
e = *IMBALANCE3-4	(32b)	4.03
f = *IMBALANCE2-4-B	(32c)	3.69
g = *IAMBICCLASH	(34)	19.27
h = B-VERSE HAS2S	§7.5.2	4.73

b. *a-verses*

	[x]	**[S]**	**[xx]**	**[Sx]**	**[Ss]**	**[Sxx]**	**[Ssx]**	**[Sxs]**	**[Sxxs]**
[x]	b,c	c	b,c	c	c	**2**	**90**	**39**	**11**
[S]	b,c	a,c	b,c	a,c	a,c	**8**	**172**	**62**	**5**
[xx]	b,c	c	b	**295**	**23**	**14**	**418**	**196**	**45**
[Sx]	b,c	a,c	b	**663**	**67**	**25**	**130**	**36**	**2**
[Ss]	b,c	a,c	b	**171**	**21**	e	e	e	e
[Sxx]	b	a	b	**423**	**52**	e	**5**e	e	e
[Ssx]	b	**130**	b	**4**d	**7**d	d,e	d,e	d,e	d,e
[Sxs]	b	g	b	d,g	d,g	d,e,g	d,e,g	d,e,g	d,e,g
[Sxxs]	b	g	b	d,g	d,g	d,e,g	d,e,g	d,e,g	d,e,g

c. *b-verses*

	[x]	**[S]**	**[xx]**	**[Sx]**	**[Ss]**	**[Sxx]**	**[Ssx]**	**[Sxs]**	**[Sxxs]**
[x]	b,c,h	c,h	b,c,h	c,h	c,h	**1**h	**172**	**44**	**4**
[S]	b,c,h	a,c	b,c,h	a,c	a,c	**98**	**182**	**79**	**12**
[xx]	b,c,h	c,h	b,h	h	h	**4**h	**422**	**557**	**110**
[Sx]	b,c,h	a,c	b,h	**733**	**5**	**1**f	**9**f	**1**f	**1**f
[Ss]	b,c,h	a,c	b,h	**67**	**4**	e,f	e,f	e,f	e,f
[Sxx]	b,h	a	b,h	**301**	**3**	e,f	e,f	e,f	e,f
[Ssx]	b	**315**	b	d	**1**d	d,e,f	d,e,f	d,e,f	d,e,f
[Sxs]	b	g	b	d,g	d,g	d,e,f,g	d,e,f,g	d,e,f,g	d,e,f,g
[Sxxs]	b	g	b	d,g	d,g	d,e,f,g	d,e,f,g	d,e,f,g	d,e,f,g

The chart gives a more nuanced view of the verse frequencies. All zero-frequency cells either violate a single highly-weighted constraint, such as *b* *RUN-ON (33.11), or a combination of two medium-strong constraints, such as *e* *IMBALANCE3-4 (4.03) + *f* *IMBALANCE2-4-B (3.69). The thinly-populated cells, in the completed analysis, generally violate a medium-strong constraint.[30] Hence there is good qualitative agreement between observations and predictions. However, the key test will be quantitative agreement, which is examined in §8 below.

This completes the core of the analysis, with constraints to cover the matching of feet within the verse and to regulate upbeats. Below, we explore a few other constraints that prove to make a substantial difference to the accuracy of the grammar.

7.5.4 Constraints based on End Weight

The principle of End Weight has a long pedigree, including in metrics. For a recent comprehensive study, see Ryan (2019). The basic idea is that there is a fundamental tendency to place "heavy" elements later in the line or phrase. A unit can count as heavy by multiple criteria: higher mora count (e.g. heavy syllables), higher syllable count (cf. binomials like *ladies and gentlemen* vs. *men and* women; Benor and Levy 2006), or higher word count (Heavy NP Shift; e.g. Bresnan et al. 2007). The units of meter, like feet and verses, can be heavier or lighter and tend to obey the End Weight principle: for instance, when pentameter systems group the line into two cola, it is typical for the two-foot colon to come first and the three-foot colon second (Piera 1980, Hayes 1988, Youmans 1989).

For *Beowulf*, the effect of End Weight is seen in the avoidance of endings that are excessive in **x** positions. The constraint B-VERSEHAS2S, discussed in §7.5.2, could be considered a constraint of this kind. To it we add simple bans on line final **x** and verse-final **xx**:

---

[30] Chart (35) also includes low-frequency cells that violate constraints other than the ones (35) is intended to illustrate. For instance, **[x][Sxx]-a** violates (28a) *SHORTVERSE and (36c) *FINAL XX; and as a result the analysis predicts 0.2 cases. For reasons of space we do not provide similar diagnoses and assessments for all low-frequency cells, but these are readily obtainable by inspecting the full tableaux in Supplementary Materials E.

(36) *Constraints based on the principle of End Weight*

a. B-VERSEHAS2S

Penalize a b-verse that lacks two positions of the form **S** or **s**; **[Ss]** is counted as **[Sx]**.

b. *FINAL x

Penalize a line that ends in **x**.

c. *FINAL xx

Penalize a verse that ends in **xx**.

7.5.5 Constraints based on Closure

We have already invoked the metaprinciple of Closure in three constraints: (30a) *UPBEAT2 (second quarter-line of the verse tolerates upbeats less readily); (30b) *UPBEAT2INB (same for b-verses); (32c) IMBALANCE2-4-B (stricter penalty, and broader scope, for imbalance in b-verses). The constraints in (36) that we classified under End Weight might also be taken as manifesting Closure.

One further possibility to consider is whether the feet themselves might obey closure in their distribution: we could imagine that marked feet avoid verse-final position and/or b-verses, and likewise that unmarked feet are preferred in the same locations. This is actually true to some extent for the canonical foot **[Sx]**, which prefers verse-final position.[31] However, this distribution is already the consequence of constraints previously stated, and would not justify adding another closure-based constraint to the grammar. Similarly, the unusual foot **[Sxx]** (just 7.5% of all feet) avoids line-final position, which might be seen as a closure effect; but we have already attributed this avoidance to *FINAL xx (36c), a constraint based on End Weight. The distribution of most of the other foot types is likewise already accounted for by existing constraints.

There is, however, one case for which adding closure-based constraints on foot distribution may be justified, at least in the sense that it follows from no other principles and improves the fit of the grammar. This concerns the arguably marked foot type **[Ss]** (it forms just 3.6% of all feet). **[Ss]** is less common as a second foot than as a first foot, and also less common in b-verses than in a-verses. The specific counts are a-verse: Ft1 192, Ft2 140; b-verse: Ft1 71, ft2 13. Thus, we venture the constraints of (37).

[31] The counts are a-verse Ft1 923, Ft2 1556; b-verse Ft1 750, Ft2 1101. As can be seen, however, there is no closure effect for the b-verse position.

(37) *Closure constraints for* ***[Ss]***

a. *[Ss]ASFOOT2

Penalize an **[Ss]** foot if it occurs as the second foot of its verse.

b. *[Ss]INB-VERSE

Penalize an **[Ss]** foot if it occurs in a b-verse.

The lowest count, i.e. 13 for foot 2 of b verses, is approximately expressed by the ganging of the two constraints.

8. Assembling and evaluating the analysis

In the preceding sections, we have set up the constraints needed for a grammar of the meter of *Beowulf*. The full AUTHORS-Grammar is completed by fitting the constraint weights against the data; see Supplementary Materials E. The results are given in (38), which gives the basic information for each constraint. The last two columns report the same statistical testing that we applied to the Russom-Augmented grammar in §6.4.

(38) *AUTHORS-Grammar*

Legend: P = Parsing, OL = Optimal Length, RS = Rhythmic Simplicity, Cl = Closure, EW = End Weight, ΔLL = change in Log Likelihood

Constraint	*Weight*	*Ref.*	*Meta-prin-ciple*	*Total viols.*	*ΔLL*	*p*
[Ss]	1.13	(15)	§5.1	446	151.6	< 0.0001
[Sxx]	−0.63	"	"	937	68.9	< 0.0001
[xx]	0.75	"	"	2084	152.3	< 0.0001
[x]	3.18	"	"	363	925.1	< 0.0001
[S]	1.58	"	"	1063	375.4	< 0.0001
[Ssx]	−1.35	"	"	2057	253.8	< 0.0001
[Sxs]	−0.51	"	"	1014	26.8	< 0.0001
[Sxxs]	−0.11	"	"	190	0.5	0.31[32]
OVERLAP	26.17	§5.1	P	0	403.2	< 0.0001
*RUN-ON	33.11	(16a)	§6	0	1368.7	< 0.0001
*SHORTFOOTSEQUENCE	13.96	(22)	OL	0	62.4	< 0.0001
*SHORTVERSE	3.22	(28a)	OL	12	199.3	< 0.0001
*LONGVERSE	1.27	(28b)	OL	5164	1330.5	< 0.0001
*UPBEAT2	0.84	(30a)	RS, Cl	125	39.4	< 0.0001
*UPBEAT2INB	0.97	(30b)	RS, Cl	44	15.8	< 0.0001
*PRETONICUPBEAT1	2.47	(31)	RS	92	511.0	< 0.0001
*IMBALANCE4-3	3.68	(32a)	OL	12	282.0	< 0.0001
*IMBALANCE3-4	4.03	(32b)	OL	5	185.0	< 0.0001
*IMBALANCE2-4-B	3.69	(32c)	OL, Cl	12	290.5	< 0.0001
*IAMBICCLASH	19.27	(34)	RS	0	195.7	< 0.0001
B-VERSE HAS 2 S	4.73	(36a)	EW	5	371.3	< 0.0001
*FINAL X	0.69	(36b)	EW	2003	85.4	< 0.0001
*FINAL XX	2.29	(36c)	EW	153	369.8	< 0.0001
*[Ss]ASFOOT 2	1.14	(37a)	Cl	183	62.6	< 0.0001
*[Ss]INB-VERSE	1.45	(37b)	Cl	84	90.2	< 0.0001

We suggest that the non-significant value for **[Sxxs]** arises because virtually all the work it might do is already performed by other constraints; the grammar would work just about as well without it (see Δ*LL*). We are reluctant to remove **[Sxxs]** from the grammar since it belongs to an otherwise well-motivated constraint family.

[32] The non-significant value for **[Sxxs]** arises because virtually all the work it might do is already performed by other constraints; the grammar would work just about as well without it (see Δ*LL*). We are reluctant to remove **[Sxxs]** from the grammar since it belongs to an otherwise well-motivated constraint family.

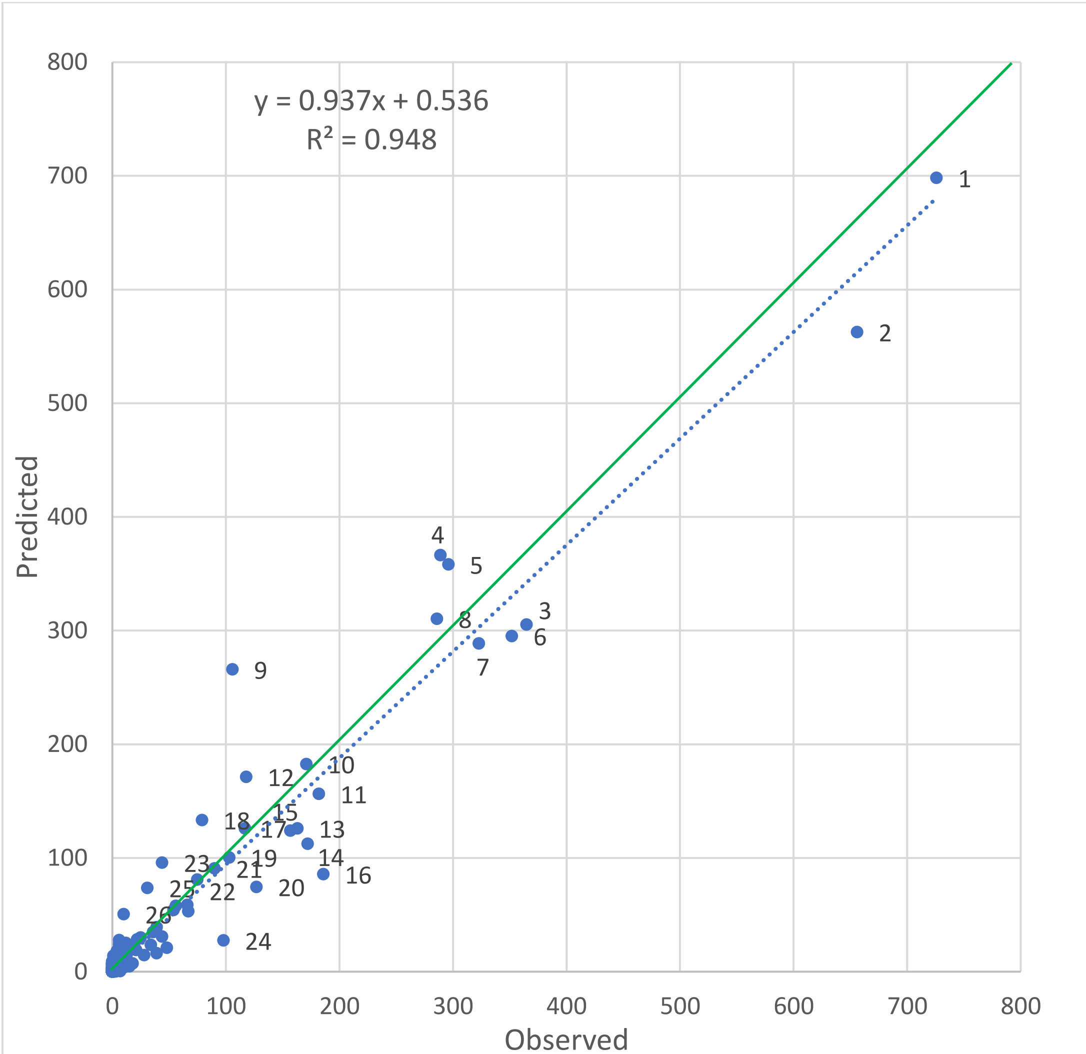
The grammar of (38) is intended to be a comprehensive account of the frequency of every verse type included in GEN. To assess this claim, we appeal to the scattergram of (39) below, to the significance values attached to the individual constraints, and to the further checks given in §8.1-§8.3. We have also found it helpful to look at how the analysis treats individual verse types. This is hard to do on an intuitive basis, but we find that the analysis is clarified if we inspect the tableau rows corresponding for particular verse types; these may be found in Supplementary Materials E.

8.1 Statistical testing

The goodness of fit achieved by AUTHORS-Grammar is illustrated in scattergram (39). As before, the green line represents the ideal pattern "predicted = observed", and the blue line is the fitted regression line.

(39) *Evaluating AUTHORS-Grammar: predicted vs. observed probabilities*

Key: 1=[Sx][Sx]-b, 2=[Sx][Sx]-a, 3=xx-[Sxs]-b, 4=[Sxx][Sx]-b, 5=[Xx][Ssx]-b, 6=[Sxx][Sx]-a, 7=[xx][Ssx]-a, 8=[Ssx][S]-b, 9=x-[xx][Sx]-a, 10=[Ss][Sx]-a, 11=[S][Ssx]-b, 12=[Sx][Ssx]-a, 13=[S][Ssx]-a, 14=[x][Ssx]-b, 15=[xx][Sxs]-a, 16=x-[xx][Sxs]-b, 17=[Ssx][S]-a, 18=[S][Sxs]-b, 19=x-[xx][Ssx]-b, 20=xx-[xx][Sx]-a, 21=[x][Ssx]-a, 22=x-[xx][Ssx]-a, 23=[x][Sxs]-b, 24=[S][Sxx]-b , 25=[Sx][Sxs]-a , 26=[xx][Sxx]-a



Visually, we feel that (39) compares favorably with (18), which evaluates the Russom-Augmented grammar. By the crude metric of correlation, it also performs quite a bit better (*r* = 0.948 vs. 0.705).

Further model comparisons are given in table (40) below.

(40) *Model comparison*

	AUTHORS-Grammar + AUTHORS-GEN	AUTHORS-Grammar + Russom-GEN	Russom-Augmented	Russom-I
Log likelihood	−18040.9	−20384.9	−22913.1	−36973.3
LL no grammar	−36936.5	−42611.2	−42611.2	−42611.2
LL saturated model	−17404.1	−18443.8	−18443.8	−18443.8
"Distance covered"	96.7%	92.0%	81.5%	23.3%
Number of constraints	25	25	17	5
AIC	36131.8	40819.8	45858.2	73956.6

The log likelihood of the data (a standard accuracy metric; Field (2026:1037) under AUTHORS-Grammar is −18040.9. We compare this value with the logical minimum of −36936.5 (no grammar all; all candidates equiprobable) and the logical maximum of −17404.1, obtained with a saturated model (one in which every attested verse type gets its own constraint). Thus, by employing the 25 constraints of AUTHORS-Grammar we cover 96.7% of the distance between no grammar and the logical maximum. Table (40) also gives the same figures for the two Russom grammars we explored earlier; note that for these, the values for the no-grammar and saturated-grammar conditions are different since there is a different GEN (§4.1.1). Lastly, to make the comparison fair, we also give the performance of our own constraint set using Russom's GEN; we do this because, all else being equal, a larger GEN will produce a lower log likelihood.

The final row of (40) gives the Akaike Information Criterion (AIC), a metric for model comparison that corrects for number of parameters; lower is better, and differences of at least 10 are held to be clearly meaningful (Burnham and Anderson 2002:70). From the values given we conclude first that even when we control for the fact that though it uses more constraints, the AUTHORS-Grammar achieves a closer fit to the data; and second that, as mentioned above, our grammar works better with the GEN we proposed in §4.1.2, rather than the original Russom GEN.

In making these comparisons we acknowledge that the Russom-Augmented grammar might perform much better if it were further augmented to have 25 constraints like our own model. However, comparison of the two approaches must depend not just on accuracy, but on the ways in which the constraint systems reflect explanatory metaprinciples of meter, a comparison that depends not just on the analysis of *Beowulf* but also on typological work and theorizing. Hence, we leave further development of the Russom grammar to future work. The point at hand is that the MaxEnt system permits grammars to be compared for their accuracy in a way that was not hitherto possible.

8.2 Russom's test verses

In the interest of rigor, Russom (2022) constructed a list of 38 imaginable verse types that violate his primary constraints and (encouragingly, for his account) are unattested in *Beowulf*. These imaginary cases include, for example, ***SxsS**, discussed above.[33] We think this is a good test and report here how AUTHORS-Grammar performed on it.

We cover first the case of **[Ss][Ssx]**. This is a verse type absent from *Beowulf*, a fact that the Russom grammars explain by its violation of *HEAVY. In our AUTHORS-Grammar, **[Ss][Ssx]** (as an a-verse) incurs one violation of *IMBALANCE3-4 and two of *LONGVERSE. These violations contribute to a large Harmony penalty, 6.34, from which is computed the predicted probability of 0.00032. This translates to a predicted value of just 1.0 cases, a value we find reasonably close to the desired value of zero.[34]

[Ss][Ssx]-a is in fact the worst case. For the remaining 34 cases of zero-frequency verse types put forth by Russom, AUTHORS-Grammar does better, assigning lower probabilities and lower expected counts (most of them very close to zero).

8.3 Checking against overfitting

Since our grammar has quite a few constraints, we must check if it is *overfitted*, a critique discussed in the modeling literature (e.g. Burnham and Anderson 2002:18). First, as shown in (38), every constraint but one (see fn. 32) passes the likelihood ratio test, meaning its weight is unlikely to be as high as it is just from sampling error. Second, we applied a common check against overfitting, five-fold cross validation (Witten and Frank 2005). Here, we repeatedly weight the constraints using just 4/5 of the data, then use these weights to make predictions about the remaining fifth. Doing this five times and combining the results, we obtain predictions in which no data have been used to predict their own probabilities. The resulting log likelihood is −18059.0, a drop of 18.1 from the original value. This drop seems modest; note for comparison that almost all of the constraints in (38) produce improvements that are considerably larger. Thus, the degree of improvement we obtain by testing against training data is not much more than that obtained from our weakest constraints. For this reason, we judge that our model is not seriously overfitted.

9. Possible improvements and outlook

Could our grammar be made better? To be sure, there are further constraints that could improve the model fit. For instance, scattergram (39) includes a rather conspicuous outlier, **x-[xx][Sx]-a**, which is predicted to be 2.5 times more common than it really is. If we added a constraint DON'T BE **x-[xx][Sx]-a**, with a suitable weight, it would put **x-[xx][Sx]-a** directly on the diagonal of the scattergram and improve the model's log likelihood. But we would see such a constraint as violating widely-adopted standards in metrics; i.e. to seek constraints that have some generality as well as grounding in general principles. We would rather consider the low observed frequency of **x-[xx][Sx]$_{\mathbf{a}}$** to be, pending a better account, purely accidental.

---

[33] Since we use a different system for parsing metrical positions into feet (§4.1.2), a handful of Russom's cases emerge as marginally attested under our system. These are **[Ssx][Ss]** (8 cases), **[Ssx][Sx]** (4 cases), and **[Sxx][Ssx]** (5 cases).

[34] As a b-verse, **[Ss][Ssx]** violates several additional constraints, so the expected count is only 0.007.

Indeed, since our grammar is already 96.7% of the way to the logical maximum accuracy (see (40)), we would be mildly surprised if it turned out that there were still any non-ad-hoc constraints left that would improve the model, given its GEN function and the constraints it already includes. A more meaningful way to improve our system would be to expand its scope; we suggest three ways below.

9.1 Improving the foot-parsing system

We noted in §8.1 that our own crude theory of foot-parsing, "minimize upbeats" (§4.1.2), produces better model fit than Russom's more nuanced approach, which pays attention to the linguistic bracketing of the syllables. Russom's approach is supported by intriguing distributional patterns (Russom 2017:17-19), and we believe that in general terms it is probably right. Future work would do well to explore novel parsing systems that like Russom's respect linguistic bracketing but perform better empirically.

Another area deserving of improvement is the *ad hoc* serial arrangement of our system, in which strings like **SxSx** are first parsed, then evaluated. Since MaxEnt is a parallelist framework, it would be a more principled approach to make parsing part of GEN: we could include all logically possible parses of all possible strings of {**S**, **s**, **x**} and let the grammar decide which parses are optimal. This goes beyond the computational capacities currently at our disposal.

9.2 Other ways to enrich GEN

GEN could also be enriched by adding to its set of terminal symbols. For instance, we could augment our current three-symbol set to five: {**S**, $\mathbf{S_r}$, **s**, $\mathbf{s_r}$, **x**}, where $\mathbf{S_r}$ and $\mathbf{s_r}$ represent resolved sequences (§3.4). This would enable us to include constraints regulating the distribution of Resolution (see e.g. Russom 1987:§4.8). We might also elaborate the symbol set to encode positions that alliterate, or to express the actual phonological weight of the syllables, supplementing the metrical weight described in §7.3. Lastly, GEN might reflect the fine-grained taxonomies of syllable weight discussed by Russom (2022:57-58). Naturally, every expansion of GEN increases the computational challenge, since even under our eight-syllable verse length limit, the size of our GEN exceeds n^8, where n is the number of distinct symbols.

9.3 Philological improvements

For this study we adopted the stress patterns for Old English posited by Russom in *Beoshare*. This includes the stressing of three suffixes for which scholarly opinion is divided. Specifically, *-ian* 'infinitive', *-ode* 'past', and *-lice* '-like (adj.)' are considered by Russom, as well as Cable (1991) and Fulk (1992), to have been stressless. Other scholars (Dresher (1985:38-48), and Gąsiorowski (1997:82-3)) posit weak stress; e.g. *-ian* was [-ˌiən]. Our analysis is based on Russom's interpretation. However, as an experiment, we created an amended copy of

Beoshare in which these suffixes were treated as stressed.[35] Strikingly, this produced a major improvement in the log likelihood of the model, from −18040.9 to −17604.9. This result, we feel, lends credibility to the stressed-suffix view, in that it reflects a more orderly pattern in the data. This suggests that an inquiry assessing the consequences of affix stress in more general terms might be fruitful.

9.4 Expanding to the line level

Is a line just the concatenation of a possible a-verse and b-verse, or are there line-level regularities?[36] Space does not permit full discussion, but our pilot modeling has found the following. (a) Bliss (1967) is correct in claiming that pairings of metrically-simple verses with other metrically-simple verses are avoided (e.g. **[Sx][Sx] [Sx][Sx]** is a very unlikely line), and that pairings of metrically-complex verses with other metrically-complex verses are likewise avoided. (b) We replicated Golston's (2009) finding that prosodically-identical verses (same **S**/**s**/**x** sequence) in the same line are avoided; (c) There are modest statistical tradeoffs of length and **S**-count between the two verses of a line. See Supplementary Materials F and G.

9.5 Summary

In this study we have attempted to create probabilistic grammars that closely match the metrical patterning of the *Beowulf* text. Taking the work of Russom as our starting point, we sought an improved constraint set, founded on the metaprinciples of §7.1. Some specific innovations we have suggested are: (a) We adopted a meter-specific system of mora count (§7.3), used to govern verse length (§7.4) and foot-pairings (§7.5). (b) We suggested that *IAMBICCLASH, a constraint widely attested in Modern English verse, is applicable to the *Beowulf* meter as well. (c) We amplified an important point made in Russom's original work: Russom demonstrated that the *Beowulf* poet employed constraints that are "upbeat-blind," respecting only the contents of the feet. Here, we revealed even greater virtuosity: the poet was upbeat-blind for the purpose of regulating foot-pairings, but upbeat-sensitive in regulating overall verse length — for comparison see (29).

A point that emerges from this study is the absolute necessity of a stochastic framework for the formal study of this meter. The force of our argument comes from comparison: the first grammar we implemented, Russom-I, represents the earlier tradition of generative metrics in proposing only constraints that are inviolable. As we showed (see §5.2 and (40)), this grammar performs poorly, neglecting the bulk of the intricate patterning evident in the data. Our work is not the first to demonstrate large gaps in coverage resulting from the neglect of gradience and violability; see §1.3.

---

[35] Some details: (a) We exclude occurrences of these suffixes immediately following a stressed light syllable. (b) We remain agnostic about whether the stressed pronunciations were present in the speech of the *Beowulf* poet or were poetic archaisms (Ringe and Taylor 2014:161), analogous to modern English *o'er*.

[36] We would like to thank two of the reviewers for *Phonology* for pointing out this question to us.

The remaining point we wish to make is methodological, arising from our use of MaxEnt. The AUTHORS-Grammar we describe here is the endpoint of a long series of exploratory grammars. In testing novel metrical analyses, we benefited from the fact that in MaxEnt the consequences of any proposal are immediately assessable by checking what they imply for log likelihood. Indeed, we sometimes found (§6.4) that a novel constraint that we thought was plausible actually failed to improve model fit. In this sense, we were no longer "flying blind." In MaxEnt methodology, data inform us clearly about what are the most helpful generalizations and structural insights.

References

Benor, Sarah, and Roger Levy (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82:233-278.

Biggs, Henry (1996). A Statistical Analysis of the Metrics of the Classic French Decasyllable and the Classic French Alexandrine. UCLA PhD Thesis.

Bliss, A. J. (1967). *The Metre of Beowulf*. Oxford: Blackwell.

Breiss, Canaan, and Bruce Hayes (2020). Phonological markedness effects in sentence formation. *Language* 96:338-370.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen (2007) Predicting the dative alternation. *Cognitive foundations of interpretation*, ed. by G. Boume, I. Krämer, and J. Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Burling, Robbins (1966). The Metrics of Children's Verse: A Cross-Linguistic Study. *American Anthropologist*, 68.6, 1418-1441.

Blumenfeld, Lev (2015). Meter as faithfulness. *Natural Language & Linguistic Theory*, 33, 79-125.

Blumenfeld, Lev (2016). Generative Metrics. An Overview. *Language and Linguistics Compass*, 413–430.

Burnham, Kenneth P., and David R. Anderson (eds). (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York, NY: Springer New York. 2002.

Cable, Thomas (1971). Constraints on Anacrusis in Old English Meter. *Modern Philology*, 69, 2:97-104.

Cable, Thomas (1974). *The Meter and Melody of* Beowulf. Chicago: University of Illinois Press.

Cable, Thomas (1991). *The English Alliterative Tradition.* University of Pennsylvania Press.

Cooper, Andrew (2017). A unified account of the Old English metrical line. PhD dissertation. Stockholm University. ORCID iD: 0000-0003-2499-1788.

Cooper, Andrew (2019). A Quantitative Analysis of the Old English Verse Line. In Plecháč, Petr, Barry P. Scherr, Tatyana Skulacheva, Helena Bermúdez-Sabel, and Robert Kolár (eds). *Quantitative Approaches to Versification*. Institute of Czech Literature of the Czech Academy of Sciences, 53-75.

Della Pietra, Stephen, Vincent J. Della Pietra & John D. Lafferty (1997). Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19:380–393.

Dresher, Bezalel E. (1985). *Old English and the theory of phonology*. Routledge. New York: Garland.

Duffell, Martin (2018). *Chaucer's Verse Art in its European Context.* Tempe, AZ: ACMRS.

Fabb, Nigel and Morris Halle (2008) *Meter in Poetry: A New Theory*. CUP.

Field, Andy (2026) *Discovering Statistics using R and RStudio*. London: Sage.
Flemming, Edward (2004) Contrast and perceptual distinctiveness. In Bruce Hayes, Donca Steriade, and Robert Kirchner, eds., *Phonetically-Based Phonology*. CUP, 232-276.
Fulk, Robert D. (1992). *A History of Old English Meter.* Philadelphia: University of Pennsylvania Press.
Fulk, Robert D., Robert E. Bjork & John D. Niles (eds.) (2008). *Klaeber's 'Beowulf and the fight at Finnsburg'*, 4th edn. Toronto: University of Toronto Press.
Fylstra, Daniel, Leon Lasdon, John Watson, and Allan Waren. 1998. Design and use of the Microsoft Excel Solver. *Interfaces* 28, no. 5:29-55.
Gąsiorowski, Piotr. (1997) *The phonology of Old English stress and metrical structure*. Frankfurt: Peter Lang.
Getty, Michael (2002). *The Metre of Beowulf: A Constraint-Based Approach*. New York: Walter de Gruyter.
Goering, Nelson (2016). The linguistic elements of Old Germanic metre: Phonology, metrical theory, and the development of alliterative verse. DPhil Thesis, Oxford University, Oxford.
Goering, Nelson (2020) Eduard Sievers' *Altgermanische Metrik* 125 years on. In Rodrigo Pérez Lorido et al. (eds.) *Of ye Olde Englisch language and textes: New perspectives on Old and Middle English language and literature*, 139-161.
Goering, Nelson (2023). *Prosody in Medieval English and Norse*. Oxford: Oxford University Press.
Goldwater, Sharon and Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. In Spenader et al. (eds.) *Proceedings of the workshop on variation within Optimality Theory*, 111-120.
Golston, Christopher (2009). Old English feet. In T.K. Dewey/Frog (eds), *Versatility in Versification: Multidisciplinary Approaches to Metrics*. Berkeley Insights in Linguistics and Semiotics 74. New York: Peter Lang, 105–122.
Golston, Christopher (2021) A Quantitative Tetrameter for Proto-Indo-European. *Synchrony and Diachrony of Ancient Greek: Language, Linguistics and Philology* 112:439.
Golston, Chris and Tomas Riad (1997) The phonology of classical Arabic meter. *Linguistics* 35: 111-132.
Golston, Christopher & Riad, Thomas (2001). Scansion and alliteration in *Beowulf*. *Jahrbuch für Internationale Germanistik*, 35:1. 77–105.
Halle, Morris & Samuel Jay Keyser (1966). Chaucer and the study of prosody. *College English* 28.187-219.
Halle, Morris & Samuel Jay Keyser (1971). *English Stress: Its Form, Its Growth, and Its Role in Verse*. New York: Harper & Row.
Hanson, Kristin & Paul Kiparsky (1996). A parametric theory of poetic meter. *Language*, 72.2. 287-335.
Hartman, Megan E. (2020). *Poetic Style and Innovation in Old English, Old Norse, and Old Saxon*. Walter de Gruyter GmbH & Co KG.
Hammond, Michael (1991). Poetic meter and the arboreal grid. *Language* 67:240-259.
Hanna, Patrizia Noel Aziz (2010). Layers of versification in *Beowulf*. *Anglia-Zeitschrift für englische Philologie* 127(2):238-260.
Hayes, Bruce (1983). A grid-based theory of English meter. *LI*, 14. 3, 357-393.
Hayes, Bruce (1988) Metrics and phonological theory. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge Survey*, CUP, 220-249.

Hayes, Bruce (1989a). The Prosodic Hierarchy in meter. In Paul Kiparsky and Gilbert Youmans (eds.) *Rhythm and Meter*. San Diego, Calif.: Academic Press. 201-260.
Hayes, Bruce (1989b). Compensatory lengthening in moraic phonology. *LI* 20:253-306.
Hayes, Bruce (2022) Deriving the Wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics* 8:474-494.
Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* 39. 379–440.
Hayes, Bruce & Claire Moore-Cantwell (2011). Gerard Manley Hopkins' sprung rhythm: Corpus study and stochastic grammar. *Phonology*, 28:2, 235-282.
Hayes, Bruce, Colin Wilson & Anne Shisko (2012). Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88:691-731.
Hayes, Bruce & Schuh, Russell G. (2019). Metrical structure and sung rhythm of the Hausa rajaz. *Language*, 95:253–299. https://doi.org/10.1353/lan.2019.0043.
Henriksson, Eric (2022). Greek Meter: An Approach Using Metrical Grids and Maxent. PhD thesis, University of Helsinki http://urn.fi/URN:ISBN:978-951-51-7990-6
Hogg, Richard (1992). *A Grammar of Old English*, Vol I: Phonology. Oxford: Blackwell.
Hyman, Larry (1985). *A theory of phonological weight*. Dordrecht: Foris.
Hutcheson, Rand B. (1995). *Old English Poetic Metre*. Cambridge: D.S. Brewer.
Jäger, Gerhard & Anette Rosenbach (2006). The winner takes it all – almost: cumulativity in grammatical variation. *Linguistics* 44. 937–971.
Jakobson, Roman (1938). K popisu Máchova verse. In Jan Mukařovský (ed), *Torso a tajemství Máchova díla*. Borovy: Praha, 207-278.
Kendall, Calvin B. (1991). *The Metrical Grammar of Beowulf*. CUP.
Keyser, Samuel Jay (1969). Old English Prosody. *College English* 30:331-356.
Kiparsky, Paul (1975). Stress, syntax, and meter. *Language* 51:576-616.
Kiparsky, Paul (1977). The rhythmic structure of English verse. *LI*, 8.2, 189-247.
Kiparsky, Paul (1989). Sprung rhythm. In Paul Kiparsky & Gilbert Youmans (eds.) *Rhythm and Meter* 305-340. San Diego, Calif.: Academic Press.
Lass, Roger & John Anderson (1975). *Old English Phonology.* CUP.
Lehmann, Ruth (1988). *Beowulf*: *An Imitative Translation.* Austin: University of Texas Press.
Lerdahl, Fred and Jackendoff, Ray (1983). *A generative theory of tonal music*. MIT Press, Cambridge, MA.
Liberman, Mark, and Alan Prince (1977). On stress and linguistic rhythm. *LI*, 8.2, 249-336.
Lotz, John (1960). Metric typology. Thomas Sebeok (ed.) *Style in Language,* Cambridge, MA: MIT Press, 135-148.
McCarthy, John (2003) OT constraints are categorical. *Phonology* 20:75-138.
McCully, Christopher B. (1996). Domain-end phenomena and metrical templates in Old English verse. In McCully, C.B. and J.J. Anderson (eds.) *English Historical Metrics,* CUP, 42-58.
McCully, Christopher. B. (2022). Raiding the demotic: verse as evidence for speech prosody in Old and Middle English. *English Language & Linguistics*. 26.3. 513-532.
McPherson, Laura & Kevin M. Ryan (2018). Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94.1. 119-156.
Minkova, Donka (2003). *Alliteration and sound change in early English.* CUP.
Minkova, Donka (2008). Prefixation and stress in Old English. *Word Structure* 1.1. 21–52.
Minkova, Donka (2025). Historical Phonology. In Laura Wright & Raymond Hickey (eds). *New Cambridge History of the English Language*, Vol. 1. CUP, 385-467.

Pascual, Rafael J. (2013-2014) Three-Position Verses and the Metrical Practice of the Beowulf Poet. *SELIM* 20:49–79.
Paster, Mary (2019). Phonology counts. *Radical: A journal of phonology* 1:1-39.
Piera, Carlos José (1980). Spanish Verse and the Theory of Meter. PhD dissertation, University of California, Los Angeles.
Prince, Alan S. (1983) Relating to the grid. *Linguistic inquiry* 14:19-100.
Prince, Alan. (1989). Metrical forms. In Paul Kiparsky and Gilbert Youmans (eds). *Rhythm and meter*, 44-80. Orlando, FL: Academic Press.
Prince, Alan & Paul Smolensky (1993/2004). Optimality Theory: Constraint interaction in generative grammar. *Optimality Theory in phonology: A reader*, 1-71. Wiley Online Library DOI:10.1002/9780470756171.
Ringe, Donald, and Ann Taylor (2014). *The Development of Old English*. Oxford: OUP.
Russom, Geoffrey (1987). *Old English meter and linguistic theory*. CUP.
Russom, Geoffrey (1995). Constraints on resolution in *Beowulf*. In M.J.Toswell (ed.) *Prosody and Poetics in the Early Middle Ages*. University of Toronto Press. 147-163.
Russom, Geoffrey (1998). *'Beowulf' and Old Germanic metre*. CUP.
Russom, Geoffrey (2017). *The evolution of verse structure in Old and Middle English poetry*. CUP.
Russom, Geoffrey (2022). The Word-Foot Theory of Old English Meter, Version II. *Journal of English and Germanic Philology* 121. 34-64.
Russom, Geoffrey (2025). *The Syntax of Beowulf: Word Order, Poetic Meter, and Formulaic Technique in the Old English Verse Clause*. London: Uppsala Books.
Ryan, Kevin. M. (2011). Gradient weight in phonology. PhD dissertation. University of California, Los Angeles.
Ryan, Kevin M. (2017). The stress–weight interface in metre. *Phonology* 34 3:581-613.
Ryan, Kevin M. (2019). *Prosodic weight: Categories and continua*. Oxford University Press.
Ryan, Kevin M. (2022). Syllable weight and natural duration in textsetting popular music in English. *English Language & Linguistics*, 26.3, 559-582.
Selkirk, Elizabeth O. (1984) *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.
Sievers, Eduard (1893). *Altgermanische Metrik*. Halle: Niemeyer.
Smith, Barbara Herrnstein (1968) *Poetic closure: a study of how poems end*. Chicago: University of Chicago Press.
Stockwell, Robert & Donka Minkova (1997). Prosody. In Robert Bjork and John Niles (eds.) *A* Beowulf *Handbook*. Lincoln: University of Nebraska Press. 55-85.
Strang, Barbara (1970). *A history of English*. London: Methuen.
Suzuki, Seiichi (1996). *The metrical organization of* Beowulf*: Prototype and isomorphism.* Berlin: Walter de Gruyter.
Terasawa, Jun (2011). *Old English Metre: An Introduction*. Toronto: University of Toronto Press.
Travis, James (1973). *Early Celtic Versecraft: Origin, Development, Diffusion* Ithaca, NY: Cornell University Press.
Witten, Ian H. & Eibe Frank (2005). *Data mining: practical machine learning tools and techniques*. 2nd edn. San Francisco: Morgan Kaufmann.
Youmans, Gilbert (1989). Milton's meter. In Paul Kiparsky and Gilbert Youmans (eds.) *Rhythm and Meter*. San Diego, Calif.: Academic Press, 341-379.

Youmans, Gilbert & Xingzhong Li (2002). Chaucer: Folk Poet or Littérateur. In Donka Minkova & Robert Stockwell (eds.) *Studies in the History of the English Language. Mouton de Gruyter, Berlin*, 153-176.