Bruce Hayes
UCLA

M90 — Workshop on Stress and Meter
September 21, 2013

Milton, Maxent, and the Russian Method

BACKGROUND ON GENERATIVE METRICS

1. Goals

- Halle and Keyser (1971: 139) defined the research topic thus:

> "When a poet composes metrical verse, he imposes certain constraints upon his choice of words and phrases which ordinary language does not normally obey. The poet and his readers may not be able to formulate explicitly the nature of the constraints that are operative in a given poem; there is little doubt, however, that neither the poet nor the experienced reader would find great difficulty in distinguishing wildly unmetrical lines from lines that are straightforwardly metrical."

- They called for generative grammarians to respond to this challenge: explicating this tacit knowledge in formal rule systems

2. Gradient metricality

- Metricality is often considered gradient: there are perfectly-canonical lines, lines that are "complex" to varying degrees, and unmetrical lines.
- Halle and Keyser's triplet of iambic pentameters (1971, hereafter HK):

(a) Ode to the West Wind by Percy Bysshe Shelley	unmetrical
(b) O wild West Wind, thou breath of Autumn's being	metrical but somewhat complex
(c) The curfew tolls the knell of parting day	metrical and not complex

3. Gradience is extra work for linguists

- In an idealized system, all forms are either perfect or word-salad.
- But such ideal systems are rare at best — and never observed for English iambic pentameter.

4. How should research proceed? HK's Frequency Hypothesis

- HK:157:

> "The more complex the line in terms of [the analysis], the less frequently it occurs."

- I regard this as a *hypothesis about poets*: their verse composition procedure produces a statistical distribution that reflects their complexity intuitions.

- My hunch is that the Frequency Hypothesis is *true*.
  - Why? Because the output of frequency-matching metrical grammars for English roughly matches my own metrical native intuitions.
  - Sooner or later: let's test the hypothesis rigorously, with experiments on living poets.
  - For now: I'll just assume it's right.
- Consequence: we can study gradient metricality indirectly by studying frequency.

5. Making use of the Frequency Hypothesis: HK's analysis of *Beowulf*

- The Old English epic *Beowulf* is composed in:
  - lines
  - … that consist of two half-lines
  - … that contain S (alliterating) and W (non-alliterating) positions.
- HK analyze a taxonomy, with counts, of the line types in *Beowulf*, with line counts carried out by Ann Reed.

6. The Beowulf frequency data

First Half-Line	*Second Half-Line*				
	SW	S	SWW	WSW	WS
SS	SSSW 999	SSS 277	SSSWW 77	SSWSW 67	SSWS 17
S	SSW 665	SS 200	SSWW 21	SWSW 25	SWS 9
SW	SWSW 405	SWS 95	SWSWW 27	SWWSW 17	SWWS 2
WS	WSSW 137	WSS 33	WSSWW 1	WSWSW 19	WSWS 3
SSW	SSWSW 38	SSWS 8	SSWSWW 4	SSWWSW 3	SSWWS 1
WSS	WSSSW 13	WSSS 6	WSSSWW 0	WSSWSW 0	WSSWS 0
SWS	SWSSW 6	SWSS 1	SWSSWW 0	SWSWSW 0	SWSWS 0

- How can we make sense of this quantitative pattern?

7. HK's strategy

- We set up some **hard constraints** — never violated in attested forms.
- We add **soft constraints** — violable, at cost of complexity — hence violations less frequent.

8. HK's hard constraints

Constraint	*Meaning*
LINE → HALFLINE$_1$ HALFLINE$_2$	A line must consist of two half-lines.
HALFLINE$_1$ → (X) X	The first half line must contain 1 or 2 X positions.
HALFLINE$_2$ → X (W)	The second half line must contain 1 X position or 1 X position + W
X → {S, SW, WS}	X must be one of these three things
*HALFLINE WITH TWO BRANCHING X's	*4 position halfline

9. Half-lines possible under the hard constraints

Halfline 1:	$[S]_x[S]_x$	Expand each X as S
	$[SW]_x[S]_x$, $[WS]_x[S]_x$	Expand first X as two positions
	$[S]_x[SW]_x$, $[S]_x[WS]_x$	Expand second X as two positions
	$[S]_x$, $[SW]_x$, $[WS]_x$	Omit one X position.
Halfline 2:	$[S]_xW$	Expand X as S
	$[SW]_xW$, $[WS]_xW$	Expand X as two positions
	$[S]_x$, $[SW]_x$, $[WS]_x$	Omit W position.

10. Simplifying the problem

- Let's follow HK and consider *only* the candidates that respect the hard constraints; i.e. (9).
- This defines the space of possible line types given in the table of (6).

11. HK's soft constraints

- This involves a bit of framework-shift; HK use licenses instead of constraints.
  - For reasons to appear I will translate the analysis into constraints.
- The soft constraints are:
  - *BRANCHING X IN FIRST HALF-LINE
  - *SHORT HALF-LINE (just one position)

12. A couple of soft constraints I'd like to add in

- *$[WS]_x$ Disfavor lines where X branches as WS (cf. Prince 1989:52)
- *BRANCHING X Disfavor branching X anywhere, not just first half-line.

13. How do we evaluate gradient theories of this sort?

- HK in 1971 followed a rigorous *qualitative* strategy: relative differences in constraint violation should be reflected by relative differences in frequency.

- Today we have the mathematical and computational tools to let us aim for a precise quantitative match.

MAXENT GRAMMARS

14. Quick summary of maxent grammars

- Maxent is a species of **harmonic grammar** (Legendre et al. 1990).
- In **architecture** it is the same as Optimality Theory (Prince and Smolensky 1993/2004), with inputs, candidates created by GEN, an EVAL component consisting of constraints.
- The constraints are not ranked but bear **weights**.
- For each input, the output of the grammar is not one winner but a **probability distribution over candidates**.
- This is done with simple math, not covered here; for a presentation see Hayes and Wilson (2008, §3).
- Following Hayes and Wilson (2008), we model well-formedness as probability.

15. The connection to the Frequency Hypothesis

- Maxent grammars come with a **learning procedure** (Berger et al. 1996) that is guaranted by proof to find the weights that achieve the optimal fit to the data.
- We feed to appropriate software the elements of the corpus (here, (6)) with frequencies and constraint violations.
  - I used the (publicly available) Maxent Grammar Tool (Wilson and George 2009)[1].
- The weighting algorithm in the software finds the weights that best match the frequencies of the data corpus
  - … which, if the Frequency Hypothesis is correct, will best match with metricality.

16. A form of maxent grammar suitable for metrics, using the Full String Set

- Let the input be a simple, single dummy form: **/Input/**.
- Let the candidate set be the **Full String Set** = all conceivable lines the poet might compose — here, all conceivable sequences of bracketed S and W, as in (9).
- We seek a grammar that
  - assigns near-zero probabilities to unattested members of the Full String Set
  - assigns frequency-matching probabilities to the attested members

17. Applying the Full String Set method to the Old English case

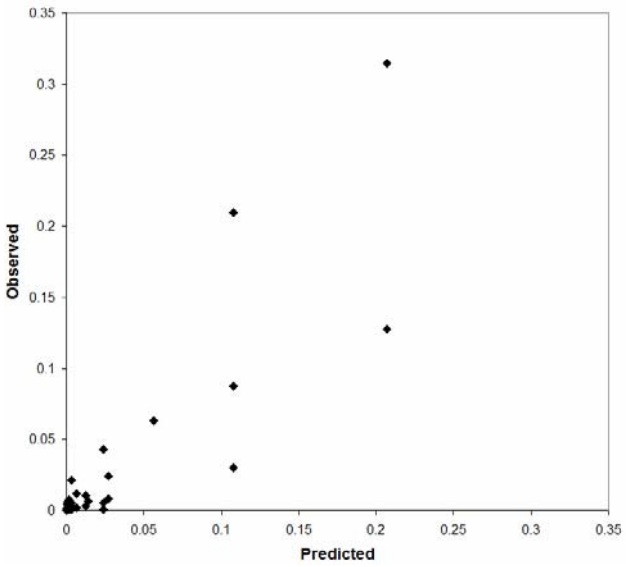
- A shortcut: we give HK's hard constraints infinite weights, so candidates violating them are assigned zero probability and may be ignored.
- Now we can focus on the candidates of (6).

[1] www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool

- From the frequencies in (6) and the constraints in (11-12), the Maxent Grammar Tool computed these constraint weights:

Constraint	*Weight*
*SHORT HALFLINE	0.65
*EXPANDED X	2.03
*EXPANDED X IN HL1	1.41
*$[WS]_x$	2.15

- Using maxent math one can generate the **predicted probability** of any given line type.
- … and these fit the data reasonably well ($r = 0.87$):



18. What we have so far

- Metricality; gradient metricality
- Frequency Hypothesis — implying we can model well-formedness by modeling frequency
- Maxent grammars as a way of expressing explicit, constraint-based analyses of gradient metricality.

IAMBIC PENTAMETER

19. Procedure of Hayes, Wilson, and Shisko (2012, *Language*)

- Like HK, we assumed the Frequency Hypothesis and used it as the basis for studying metricality.
- Focus: Shakespeare's *Sonnets* and Books IX and X of Milton's *Paradise Lost*
- We set up a representational system: 4 levels of stress, 5 of phrasing, hence 20 phonological symbol types.

- Within this system, the Full String Set is very large:
  - For 10-syllable strings, there are 20^{10}, about 10 trillion.
  - Special software by Wilson used finite state machines to search this set without enumerating it.

20. What constraints?

- We set up a system of 87 constraints — "**UM**, Universal metrics"
- This was based on a general framework for metrics, intended to be broad enough to encompass most of the research literature in generative and traditional metrics.
- Constraints included from:
  - Jespersen 1900; Bridges 1923; Sprott 1953; Halle and Keyser 1966, 1971; Kiparsky 1975, 1977; Tarlinskaja 1976; Hayes 1983, 1989; Youmans 1989; Hanson and Kiparsky 1996; Fabb and Halle 2008

21. Procedure

- Just like in the Beowulf model above: software finds constraints and weights that best match the frequencies of the data — including zero frequency for unattested members of the Full String Set.
- These constraints and weights served as a grammar, which we then used to assign probabilities to all the lines in the corpus.

22. How did it come out?

- Predictions of the grammar fairly decently match our own (imperfect, anachronistic) intuitions — see Hayes/Wilson/Shisko §7.3.1 and below.
- We carried out other tests (e.g. the Youmans word-order-repair test; Youmans 1982, 1983, 1989), and it did reasonably well.

23. What did the project say about existing research in generative metrics?

- Vindication: the published literature did rather well in providing effective constraints for the analysis.

EXPECTED VALUES AND THE RUSSIAN METHOD

24. Is it right to pick out metrical lines against the Full String Set?

- This is not the only possible conceptualization.

25. Altenrative: comparing against expected values

- In general, when you're working with frequencies, it's often good to compare **observed** frequencies with a good estimate of the **expected** frequencies.
- English utterances in general have statistical phonological properties

- E.g. avoidance of clash and lapse (Liberman and Prince 1977, Prince 1981, Selkirk 1984 et seq.)

- Perhaps we should be doing a comparison: how are the phonological properties of verse different from the "normal-phonology" baseline?

26. The Russian method[2]

- Collect a **prose sample** — ideally, written by the poet himself — to serve as a baseline for comparison with verse.
  - For instance: a set of sentences that happen to be 10 syllables long.
- References: Kolmogorov and Proxorov 1968, Gasparov 1971, Taranovsky 1971, Tarlinskaya and Teterina 1974, Tarlinskaja 1976.
- And not just Russians: Devine and Stephens (1976), Biggs (1996), Hall (2006), Kevin Ryan (three hours from now)

27. Maxent as a formal framework for applying the Russian method

- We develop a maxent grammar whose sole purpose is to make *correct guesses about whether a particular line is verse or prose*.
  - Metrical = highly likely to be verse
  - Unmetrical = highly likely to be mere prose
- We use the same constraints as before.
- Some prose lines will be metrical by accident — but few enough to make little difference.

28. This is a question of theory, not methodology

- We can construe the analytical approaches as hypothesized about what poets are subconsciously trying to achieve:
  - **Full String Set hypothesis**: verse composition = selection of strings that stand out from the Full String Set as metrically probable.
  - **"Russian hypothesis"**: verse composition = selection of strings that stand out from the *population of characteristic phonological utterances of the language* as metrically probable.
- To help decide, let's analyze the same data from both perspectives.

MY MILTON-BY-RUSSIAN-METHOD PROJECT

29. Milton verse corpus

- As in Hayes/Wilson/Shisko 2012 Books 9-10 of *Paradise Lost* (2293 lines)
- I recycled the prosodic transcriptions (stress, phrasing) of this corpus.

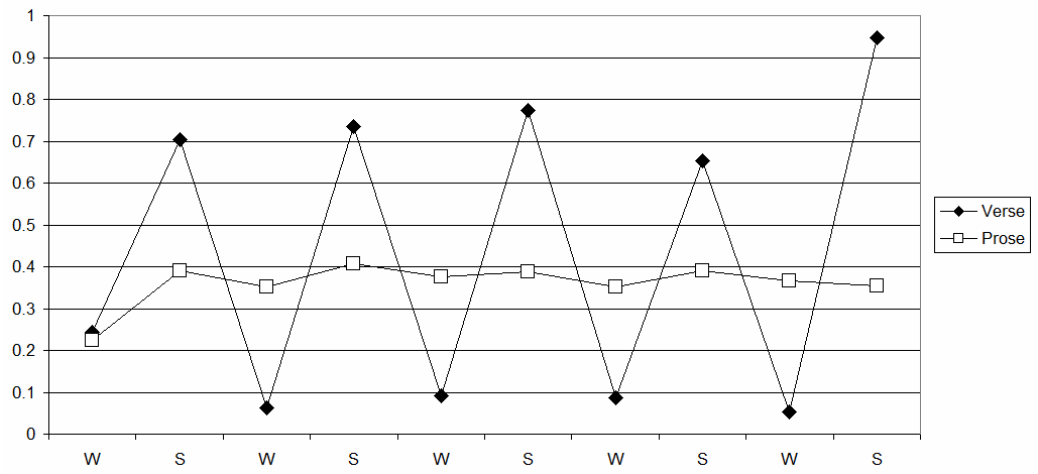
[2] This quick summary oversimplifies the work of the Russian scholars, notably in neglecting their work in actually *computing* a plausible prose model instead of sampling it.

30. Milton prose corpus

- Milton wrote a great deal of prose, easily accessed on-line.
- I downloaded the full corpus, broke it into sentences, and randomized sentence order.
- Finding pseudolines: for each sentence I
  - selected the first 10 syllables of a sentence if that ended in a word break
  - else the first 11 if that ended in a word break
  - else discarded the sentence
- I annotated the corpus for prosody in the same way I did earlier for the verse corpus.
- Corpus size: 1000 lines.

31. How does a prose sample differ from the verse corpus in aggregate properties?

- Stress profiles (fraction stressed, by metrical position in line)



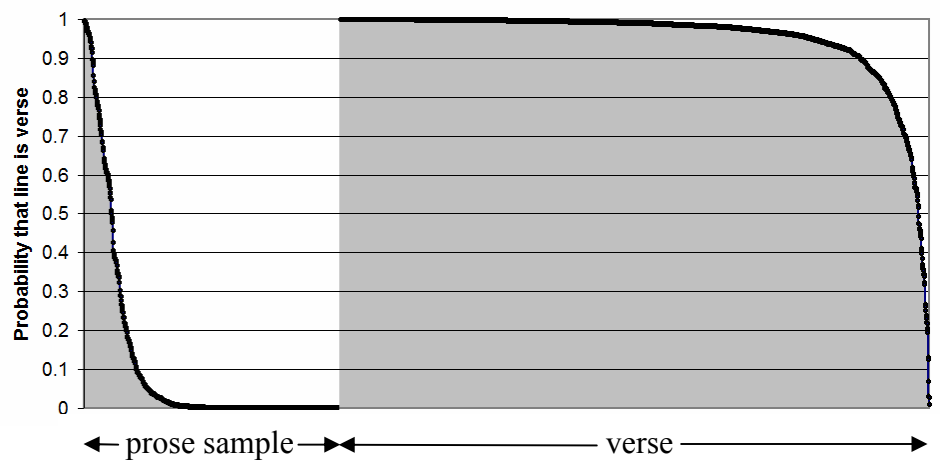
  - Plainly, a sensible "comb" with peak in 10 for verse; mostly flat-line for prose sample.
  - Position 1 in sample is low — this will be important later on.

32. Maxent for the Russian method: a binary choice

- We represent Verse with 1 and Prose with 0.
- No custom software needed: maxent with just two candidates is just a notational variant of Logistic Regression, which I did with R (R Core Team 2013).
- For both Full String Set Method and Russian Method I used a grammar selected "bottom up" from the 87-constraint UM— accrete grammar by adding best-performing constraint.

33. Is the Russian-Method grammar effective at separating verse from prose?

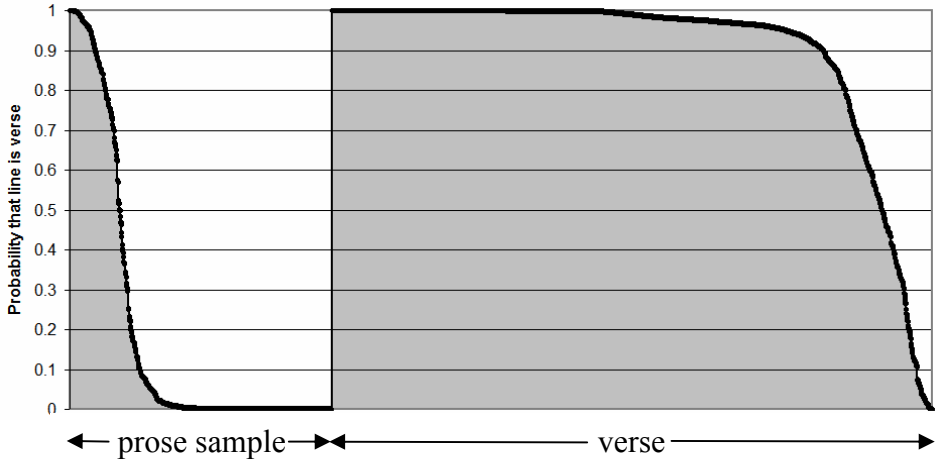
- We can check quickly with an Excel sort: first prose, then reals, sorted by descending probability:



- The system assigns most prose lines low probability as verse, and most verse lines high probability, as desired.

34. Does the Full String Set Method (HWS) also separate verse from prose?

- Reassuringly, yes.

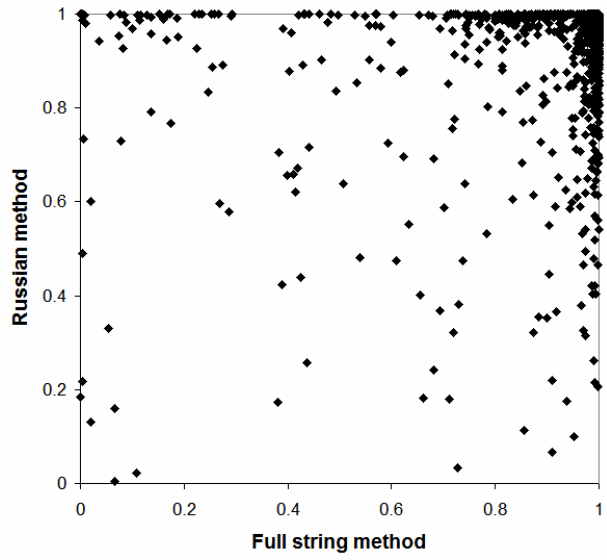


- Not *quite* as well as the Russian-method grammar — which was, after all, designed with this very purpose in mind.

35. Does the Russian method yield conclusions about gradient metricality similar to Full String Set Method?

- No!
- The correlation of harmony values among the lines of the Milton verse corpus is a pathetic 0.17.
- Even when I massage the data (by fitting a logistic curve to the Full String Set Method harmonies), it increases only to 0.33.

- Here is a scattergram showing how different the predictions of the two methods are.



36. Might we conjecture which approach is doing better?

- We can look at the **maximal difference** cases:
  - Full String Set Method likes the line a lot more than Russian Method does.
  - Russian Method likes the line a lot more than Full String Set Method does.
- My own reader's intuition is that where Full String Set Method regards the line as straightforward and noncomplex, so do I.
- Caution: I may have chosen my examples tendentiously; you can download all the data for your perusal if you like at www.linguistics.ucla.edu/people/hayes/Papers/Predictions.txt.

37. Lines that Full String Set Method thinks are simple, Russian Method complex

	Full String Method Harmony	*Full String Method Probability*	*Russian Method Harmony*	*Russian Method Prob.*
Where Tigris at the foot of Paradise	4.65	**0.998**	0.141	**0.465**
In offices of Love, how we may lighten	5.71	**0.993**	1.289	**0.216**
Wouldst thou admit for his contempt of thee	4.4	**0.998**	1.345	**0.207**
Yet are but dim, shall perfectly be then	8.34	**0.910**	2.628	**0.067**

38. Lines that Russian Method thinks is simple, not Full String Set Method

	Full String Method Harmony	*Full String Method Probability*	*Russian Method Harmony*	*Russian Method Proba-bility*
Fooled and beguiled, by him thou, I by thee	13.43	**0.058**	−6.504	**0.999**

Out of my sight, thou Serpent, that name best	12.74	**0.110**	−6.563	**0.999**
One Heart, one Soul in both; whereof good proof	12.36	**0.153**	−7.375	**1.000**
Thrones, Dominations, Princedoms, Virtues, Powers	11.91	**0.221**	−9.386	**1.000**

39. A more substantive comparison, with diagnosis

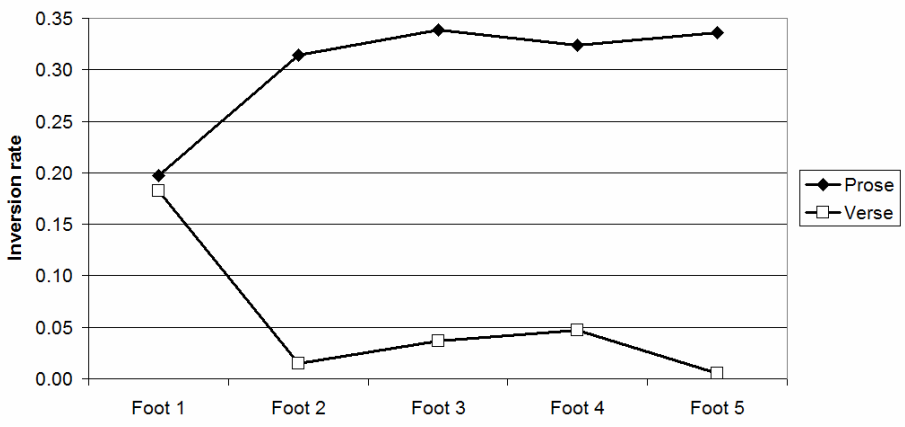
- Among the *very favorite* lines of Russian Method are lines where the first foot is "inverted" (phonological trochee in metrical iamb position).

Line	*Full String Method Harmony*	*Full String Method Probability*	*Russian Method Harmony*	*Russian Method Proba-bility*
Bitter ere long back on itself recoils;	12.25	0.168	−9.02	**1.000**
Motion, each act won audience ere the tongue,	12.22	0.172	−9.00	**1.000**
Goddess humane, reach then, and freely taste.	11.87	0.228	−13.48	**1.000**
Human, to put on Gods, death to be wished,	11.83	0.235	−8.74	**1.000**
Constant, mature, proof against all assaults,	11.66	0.267	−9.24	**1.000**

- This is almost surely an error for the Russian Method: most metrists would agree that inversion in pentameter is a source of complexity — even its most common location, the first foot.
  - German and Russian pentameter outright forbid inversion of the first foot when it is "lexical" (two syllables in same word, as in examples above).

40. Where did Russian Method go wrong? My conjecture

- Take another peek at the prose lines in (31).
- Because English sentences often begin with a proclitic (like *and*, *but*, *for*, *the*, *he*), the first syllable of a prose sample line is frequently stressless.
- … and so initial "inversions" (phonological trochee filling metrical iamb) are *underrepresented* in the prose sample.
- What about verse? For Milton — like all English poets —inversion is *most common* in the first foot.
- The two trends create a "pinching" pattern (graph):



- So, as far as weight-setting goes, *the first foot is already just fine* — no constraints are needed to derive a deviation from prose, because there is no deviation.
- Which casts doubt, perhaps, on the whole idea that prose norms serve as baselines for metricality.

41. A simpler case

- It's widely felt that big internal phrase breaks contribute metrical complexity
  - Look again at (38), the set of lines Russian Method does not penalize enough; internal breaks abound here.
- Simple comparison of the violation rates show that Milton's verse is richer in **large line-internal phrase breaks** than his prose
  - for Intonational Phrase breaks: .88 per line in verse, .52 per line in prose
- Yet in the HWS grammar, constraints banning big line-internal phrase breaks actually got big weights. How can this be?
- Answer: Line-internal big phrase breaks in verse may be overrepresented relative to prose, but *underrepresented* relative to the Full String Set.
- A possible explanation for the complexity of internal breaks:
  - The Full String Set provides the right backdrop for evaluating the constraints; Milton's prose does not.

CONCLUSIONS

42. Reviewing the core question

- **Full String Set hypothesis**: verse composition = selection of strings that stand out from the Full String Set as metrically-probable.
- **"Russian hypothesis"**: verse composition = selection of strings that stand out from the population of characteristic phonological utterances of the language as metrically-probable.
- So far, modeling based on the Full String Set Hypothesis seems to be getting a better approximation to metrical well-formedness intuitions.
- It remains to be seen if this is reason to reject the Russian Hypothesis, or reflects errors in my implementation of it.

References

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

Biggs, Henry. 1996. A statistical analysis of the metrics of the classic French decasyllable and the classic French Alexandrine. Ph.D. thesis, University of California, Los Angeles.

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

Devine, A. M. and Laurence Stephens (1976) The Homeric hexameter and a basic principle of metrical theory. *Classical Philology* 71:141-163.

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

Gasparov, Mikhail L. 1971. "Quantitative methods in Russian metrics: achievements and prospects." *Metre, rhythm. stanza, rhyme. Russian poetics in translation* (1971).

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

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

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

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

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

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

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

Hayes, Wilson and Shisko

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

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

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

Kolmogorov, Andrej, and Aleksandr Proxorov. 1968. K osnovam russkoj klassiceskoj metriki [On the principles of Russian classical metrics]. *Sodruzestvo nauk i tajny tvorcestra*. Ed. B. S. Mejlax. Moscow: Iskusstov, pp. 397-432. [Not seen: citation and brief summary in Gasparov 1971.]

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

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

Prince, Alan and Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Technical report, Rutgers University Center for Cognitive Science. [Published 2004: Blackwell]

Prince, Alan. 1981. 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.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Selkirk, Elisabeth O. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge: MIT Press.

Sprott, S. Ernest. 1953. *Milton's art of prosody*. Oxford: Blackwell.

Tarlinskaja, Marina and Lilya M. Teterina. 1974. Verse—Prose—Meter. *Linguistics* 129: 63-86.

Tarlinskaja, Marina. 1976. *English verse: theory and history*. The Hague: Mouton.

Wilson, Colin and Ben George (2009) Maxent Grammar Tool. Software. Available www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool.

Youmans, Gilbert. 1982. Hamlet's testimony on Kiparsky's theory of meter. *Neophilologus* 66: 490-503.

Youmans, Gilbert. 1983. Generative tests for generative meter. *Language* 59: 67-92.

Youmans, Gilbert. 1989. Milton's meter. In Paul Kiparsky and Gilbert Youmans, eds., *Rhythm and meter*, 341-379. Orlando, FL: Academic Press.