

Strictness functions in meter

1 Overview of the corpus and issues

- (1) The Kalevala: Finnic, 22,738 lines (67,129 words), anthologized by Lönnrot (1849)
- (2) Conventional description (Sadeniemi 1951, Kiparsky 1968, Leino 1986, 1994):
 - (a) Octosyllabic lines in couplets
 - (b) Obligatory alliteration within lines
 - (c) Trochaic: (S W) (S W) (S W) (S W)
 - (d) Primary stressed syllables must be heavy in S, light in W¹
 - (e) Exceptions frequent at the beginning, but taper off rapidly after the first foot
 - (f) Final monosyllables are prohibited

e.g. *kalanluinen kanteloinen*

| | | | | | | | |
|----|-----|-----|-----|-----|----|-----|-----|
| S | W | S | W | S | W | S | W |
| ka | lan | lui | nen | kan | te | loi | nen |

ei ollut osoajata

| | | | | | | | |
|----|----|-----|---|----|---|----|----|
| S | W | S | W | S | W | S | W |
| ei | ol | lut | o | so | a | ja | ta |

- (3) Kiparsky (1968) argues (controversially²) that Kalevala scansion relies on presurface forms. My corpus is designed to obviate this issue.³ After these exclusions and removing duplicates, 17,485 lines remain (77% of the original).
- (4) The conventional description misses a lot. The meter is gradiently strict along several scales:
 - (a) **Positional strictness:** Exceptions decay at particular rates across lines and couplets.
 - (b) **Increasing word size:** Boundaries are increasingly avoided towards the end of the line. Final monosyllable avoidance is only the most extreme manifestation of this tendency.
 - (c) **Stress:** Unstressed and secondary stressed syllables are also regulated, only more weakly. Degree of stress modulates degree of regulation.
 - (d) **Weight:** Syllable weight is richer than heavy vs. light. The more the duration of a syllable deviates from its metrical target, the more it is penalized.

¹Primary stressed \Leftrightarrow word-initial (excluding clitics). $C_0\check{V} \Leftrightarrow$ light. VCCV always VC.CV. No resyllabification.

²Cf. Manaster Ramer 1981, 1994, Devine and Stephens 1975, Lauerma 2001.

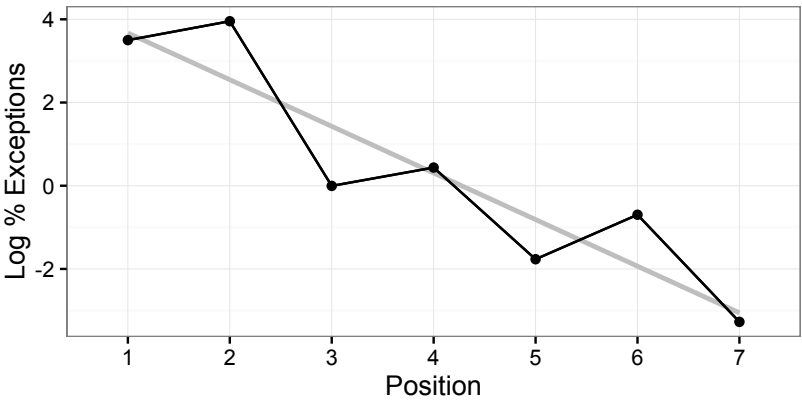
³First, I take only surface octosyllables, meaning that Kiparsky's rules of contraction and apocope couldn't have applied. Second, I exclude any line in which vowel or consonant gemination could have potentially applied.

2 Increasing strictness

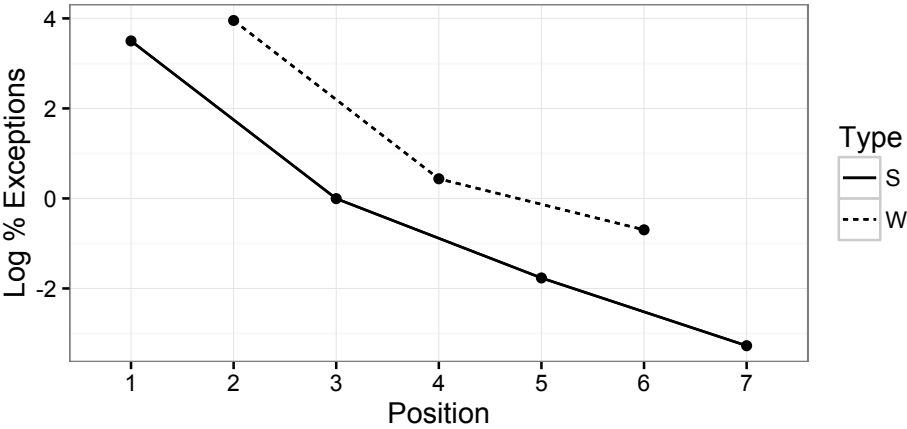
- (1) The meter “need not be obeyed at all in the first foot, may be in the second, should be in the third, and must be in the fourth” (Manaster Ramer n.d.).
- (2) Decay in exceptions. Only primary stressed syllables are diagnostic; non-initial syllables and clitics (subminimal words) are put aside for now.

| Foot | Position | Exceptions | Total Stressed Syllables | % of Stressed Syllables |
|------|----------|------------|--------------------------|-------------------------|
| 1 | 1 | 5,456 | 16,484 | 33.1% |
| | 2 | 836 | 1,603 | 52.2% |
| 2 | 3 | 95 | 9,539 | 1.0% |
| | 4 | 68 | 4,394 | 1.5% |
| 3 | 5 | 15 | 8,777 | 0.2% |
| | 6 | 27 | 5,423 | 0.5% |
| 4 | 7 | 1 | 2,637 | 0.0% |
| | 8 | 0 | 0 | — |

- (3) Decay is approximately logarithmic. Gray line is best-fitting regressor.



- (4) Separating S and W positions reveals underlying monotonicity. Independent of decay, S positions appear to be uniformly stricter than W positions.



- (5) Two strategies for modeling the S vs. W discrepancy:
 - (a) Separate constraints, *LIGHT/S and *HEAVY/W (the former akin to STRESS-TO-WEIGHT, the latter to WEIGHT-TO-STRESS)
 - (b) One constraint, *MISMATCH, but exceptions are more frequent in W because most (67%) stressed syllables are heavy.
- (6) Here, I adopt the former strategy, since:
 - (a) These two constraints have always been treated as separate in generative metrics (e.g. Halle and Keyser 1971, Hanson and Kiparsky 1996, Hayes et al. 2012).
 - (b) Modeling the observed rate of exceptions is more straightforward than modeling the expectation-adjusted rate.
- (7) Summarizing so far:
 - (a) virtually free beginning (67% heavies in position 1)
 - (b) virtually categorical ending (1 exception in position 7)
 - (c) interpolation function is logarithmic
 - (d) distance metric could be the position or the foot (with two constraints, it's moot)

3 Scalar functions in maxent grammar

- (1) Distance-based decay (and perhaps all scalar mapping) in phonology, metrics, etc. seems generally to be exponential, e.g.
 - (a) harmony (esp. Zymet to appear, also Kimper 2011)
 - (b) morphology (McPherson and Hayes to appear)
 - (c) text-setting (McPherson and Ryan in prep.)
 - (d) ganging cumulativity (Zuraw 2012)
 - (e) cf. scalar perception generally (e.g. the Weber-Fechner law⁴)
- (2) In maxent HG,⁵ decays are necessarily exponential: Because harmony is exponentiated, any linear differential in harmony translates to an exponential differential in p .
 - (a) $p(\text{cand}_0) = e^{H_0} / \sum_i e^{H_i}$, where H_0 is cand_0 's nonpositive sum of weighted violations and i ranges over all candidates.
 - If one cares only about relative probabilities, the denominator can be omitted, leaving $Lk(\text{cand}_0) = e^{H_0}$ (Hayes and Wilson 2008).

⁴The percept tends to be proportional to the logarithm of the stimulus.

⁵On maxent for metrics, see e.g. Hayes et al. 2012, Ryan 2011, and Hayes 2016.

- (b) Logistic models also have this property, since $p(\text{cand}_0) = 1/(1 + e^{H_0})$ (where H_0 is the candidate’s negated log-odds = harmony).

(3) A simple illustration:

a.

| | p | H | CONSTRAINT <i>weight</i> = 1 |
|----|-----|-----|---------------------------------|
| a. | .73 | 0 | |
| b. | .27 | −1 | * |

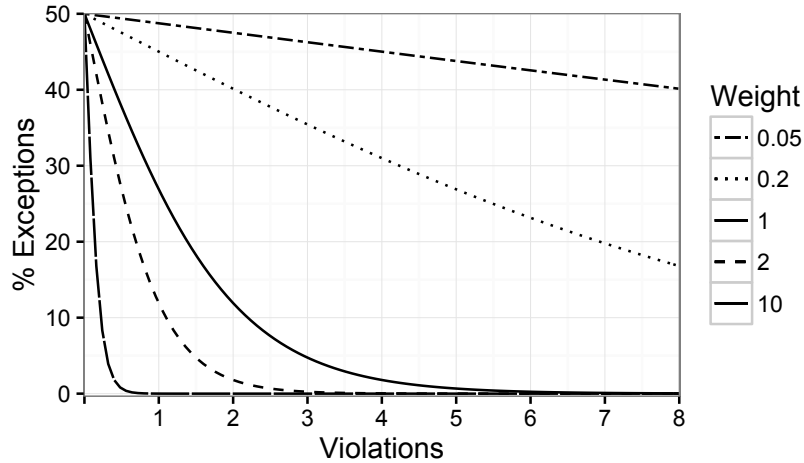
b.

| | p | H | CONSTRAINT <i>weight</i> = 1 |
|----|-----|-----|---------------------------------|
| a. | .88 | 0 | |
| b. | .12 | −2 | ** |

c.

| | p | H | CONSTRAINT <i>weight</i> = 1 |
|----|-----|-----|---------------------------------|
| a. | .95 | 0 | |
| b. | .05 | −3 | *** |

(4) More generally, for $n \in [1, 8]$ violations with CONSTRAINT at various weights:



(5) The weight of the constraint modulates (a) the “left intercept” in the space of the line, (b) the right intercept, and (c) the shape of the decay between the two, but all three (a–c) are mutually constrained, since there’s only one free parameter.

0.05 \Rightarrow a meter that is almost completely flexible throughout (cf. free verse)

10 \Rightarrow a meter that is (virtually) completely strict throughout

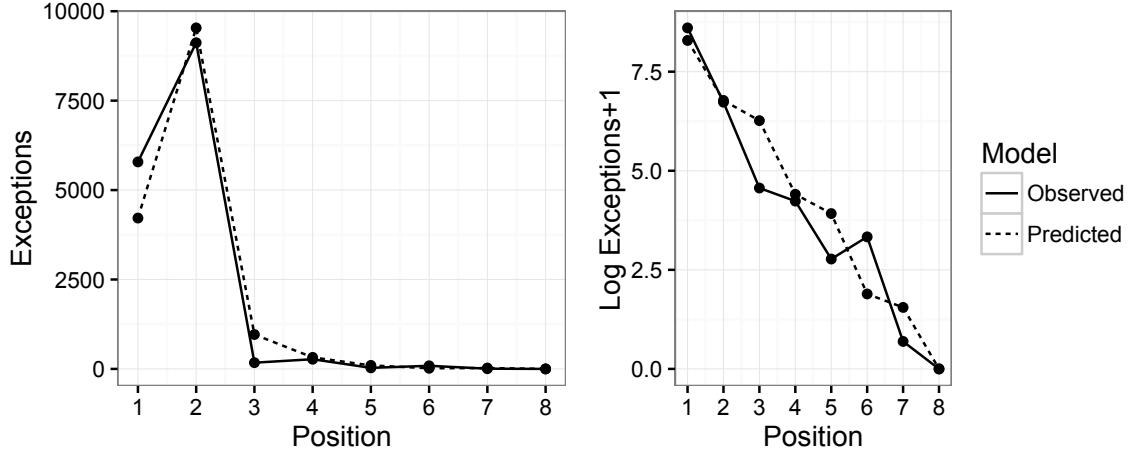
0.2 \Rightarrow a meter that is rather flexible throughout, but with a lower “right intercept”

(6) One impossibility in this model: linear interpolation from free to fully strict (e.g. a version of the Kalevala with the same intercepts but linear decay).

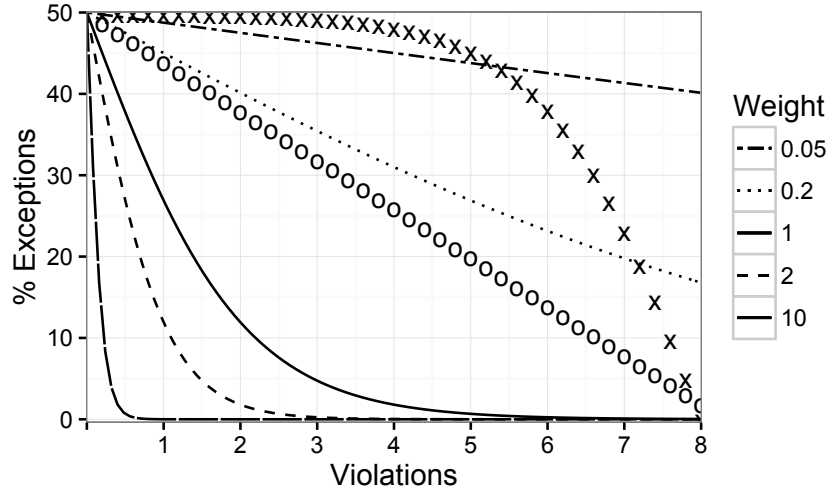
(7) One solution (to be revised): scaled mapping. Multiply every mapping violation by the position of the violation.⁶

⁶This is formally equivalent to cloning the constraint (*LIGHT/ S_{foot_1} , etc.), but with fixed proportions between weights. I prefer to think of multiplication as being part of the violation function, since multiply constraints imply multiple free parameters.

- (8) With optimal weights of 1.156 and 1.334 for scaled *LIGHT/S and *HEAVY/W, respectively:



- (9) Typological prediction: If we find a sharp change in strictness, it should occur at the beginning of the constituent, not the end (as with e.g. the line in x's).

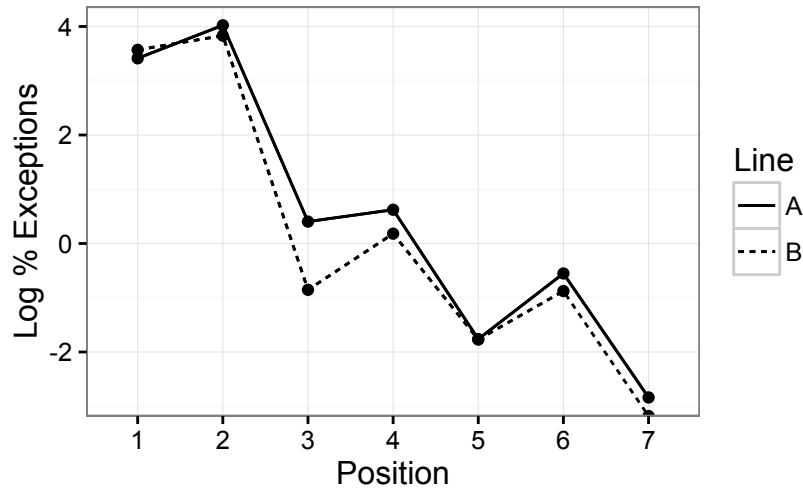


Moreover, linear-looking decay is only possible if the right-intercept is fairly high (cf. o's).

- (10) This model is compatible with various left-edge licenses common in meter, e.g.
- (a) In English, inversion is found only IP-initially, esp. line-initially
 - (b) In the Kalevala, resolution is almost exclusive to the first foot
 - (c) In the Greek trimeter, resolution is most frequent in the first position of each metron (of four positions), esp. in the line-initial metron
- (11) Caveat: The line is not always the only relevant constituent. Metra and hemistichs can sometimes cause (partial) “resets.”
- (12) E.g. a sharp difference between hemistichs is found in the Sanskrit *śloka*, where the first half of the line (*pāda*) is quite loose and the second half quite strict, with no transition.

4 Increasing strictness beyond the line: IPs and couplets

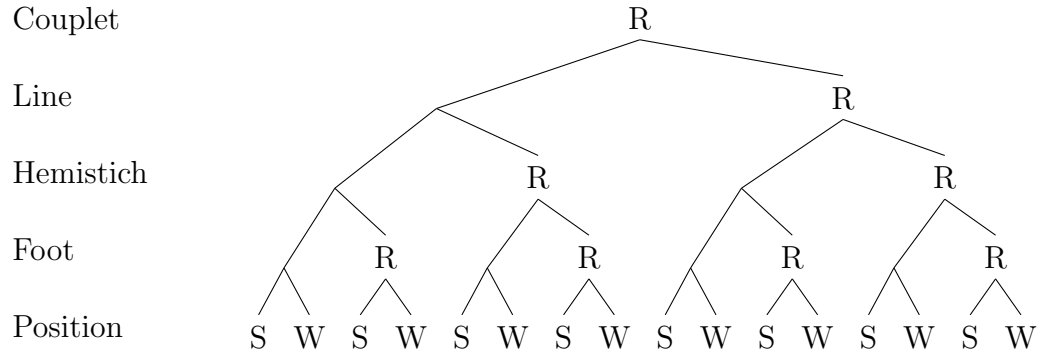
- (1) As a working operationalization, take punctuation to indicate an IP break.
- (2) Within the line (excluding the first foot), exceptions are 2.5 times more likely IP-initially than IP-internally (1.6% initial vs. 0.6% internal, $p < .001$).
- (3) This factor doesn't override (or explain) positional strictness; it just modulates it.
- (4) Next, consider the couplet. Excluding the first foot, exceptions are 1.9 times more likely in the A-line than in the B-line (4.1% A-line, 2.2% B-line, $p < .0001$).



- (5) Final strictness thus applies across various overlapping levels of constituency:
 - (a) Lines are looser at the beginning
 - (b) IPs are looser at the beginning
 - (c) Couplets are looser in the A-line
- (6) To better assess each syllable's penalty, the model should consider its position in the line, IP, and couplet.
- (7) In terms of constraints with multipliers (to be rejected):
 - (a) *LIGHT/S/line: multiply by position in the line
 - (b) *LIGHT/S/IP: multiply by position in the IP
 - (c) *LIGHT/S/couplet: multiply by position in the couplet

5 Replacing multipliers with structure

- (1) Multiplying by position is the sort of counting operation linguists abhor.
- (2) Fortunately, we can do without, while preserving the typological predictions above.
- (3) Richer binary metrical constituency for the Kalevala, with right branches labeled R:



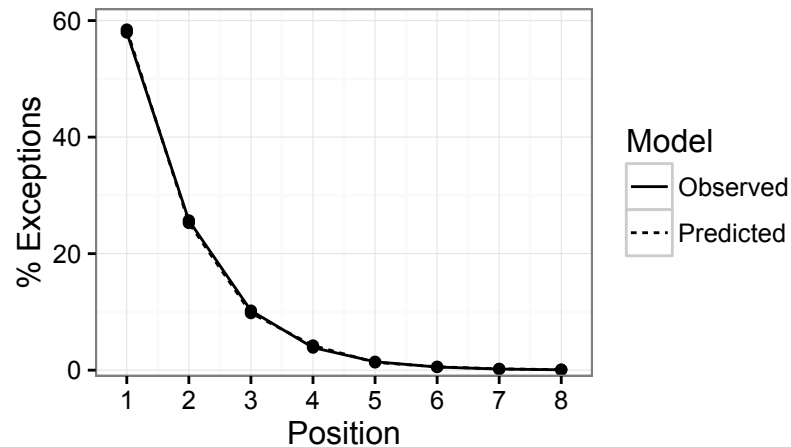
- (4) Mapping constraints indexed to the right branch of every level:

*MISMATCH/line-R
 *MISMATCH/hemistich-R
 *MISMATCH/foot-R

- (5) E.g. *MISMATCH/line-R: Assign a penalty to each violation of the meter (heavy in S or light in W) under line-R.
- (6) Why not replace R with S (heads)? This would only work if meters were universally analyzed as right-headed above the position, as strictness is universally right-oriented.
- (7) To handle non-metrical constituency, including IPs, I assume that such constraints would also come indexed to non-metrical prosodic structure.
- (8) Thus, counting horizontally is replaced by “counting” vertically (running up the tree), but no constraints are gradient; it’s just summing over categorical violations.
- (9) Decays are still predicted to be logarithmic only.
 - (a) Sharp decays are possible only at the beginning of a constituents.
 - (b) Linear-looking decays are possible only with a middle-range right-intercept.

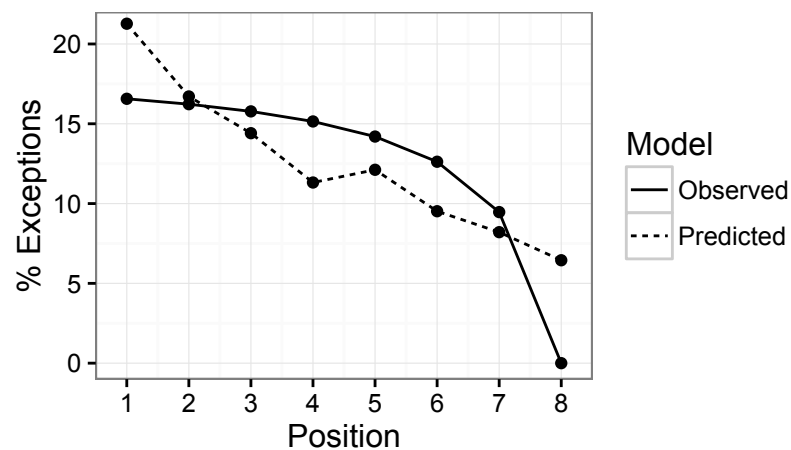
(10) The grammar can match a perfectly logarithmic decay perfectly, e.g.

| | | Spos | Wpos | Rfoot | Rhemi |
|----------|--------|------|------|-------|-------|
| 10000000 | 2689.0 | 1.0 | 0.0 | 0.0 | 0.0 |
| 01000000 | 1192.0 | 0.0 | 1.0 | 0.0 | 0.0 |
| 00100000 | 474.0 | 1.0 | 0.0 | 1.0 | 0.0 |
| 00010000 | 180.0 | 0.0 | 1.0 | 1.0 | 0.0 |
| 00001000 | 67.0 | 1.0 | 0.0 | 0.0 | 1.0 |
| 00000100 | 25.0 | 0.0 | 1.0 | 0.0 | 1.0 |
| 00000010 | 9.0 | 1.0 | 0.0 | 1.0 | 1.0 |
| 00000001 | 3.0 | 0.0 | 1.0 | 1.0 | 1.0 |

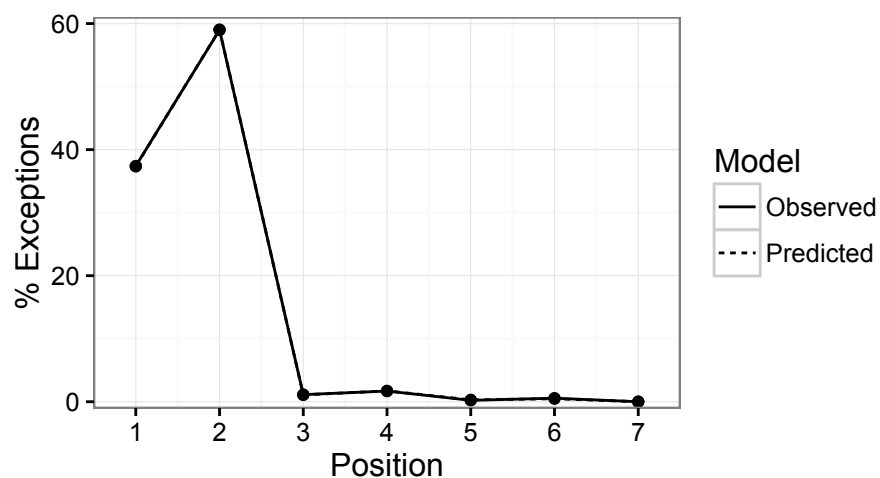


(11) But it fails horribly for an inverted logarithmic decay (accelerating into the end), e.g.

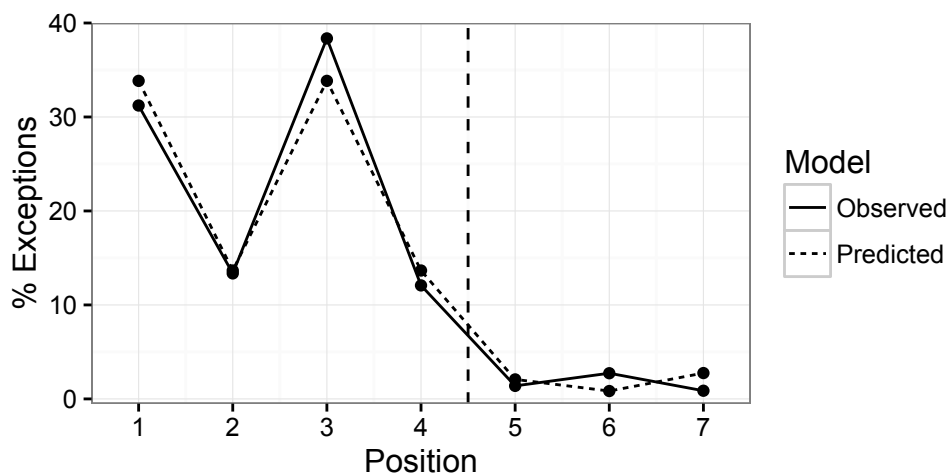
| | | Spos | Wpos | Rfoot | Rhemi |
|----------|-------|------|------|-------|-------|
| 10000000 | 875.0 | 1.0 | 0.0 | 0.0 | 0.0 |
| 01000000 | 857.0 | 0.0 | 1.0 | 0.0 | 0.0 |
| 00100000 | 833.0 | 1.0 | 0.0 | 1.0 | 0.0 |
| 00010000 | 800.0 | 0.0 | 1.0 | 1.0 | 0.0 |
| 00001000 | 750.0 | 1.0 | 0.0 | 0.0 | 1.0 |
| 00000100 | 667.0 | 0.0 | 1.0 | 0.0 | 1.0 |
| 00000010 | 500.0 | 1.0 | 0.0 | 1.0 | 1.0 |
| 00000001 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |



(12) The Kalevala is tightly matched.

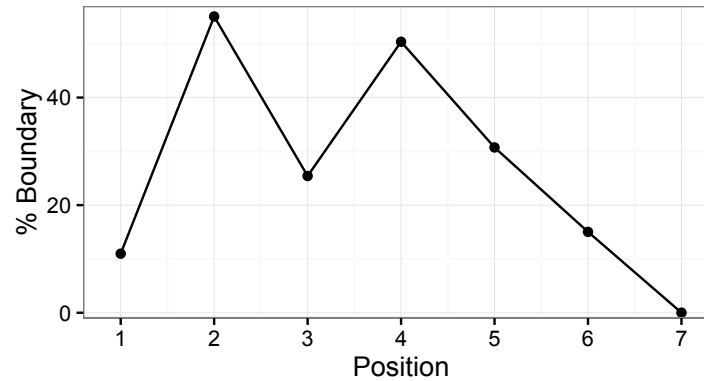


(13) Cf. octosyllabic lines (“normal dimeter”) in the R̥g-Veda ($n = 11,235$; Oldenberg 1888, Arnold 1905). Fairly loose first half-line (pre-dashed divider), fairly strict thereafter.

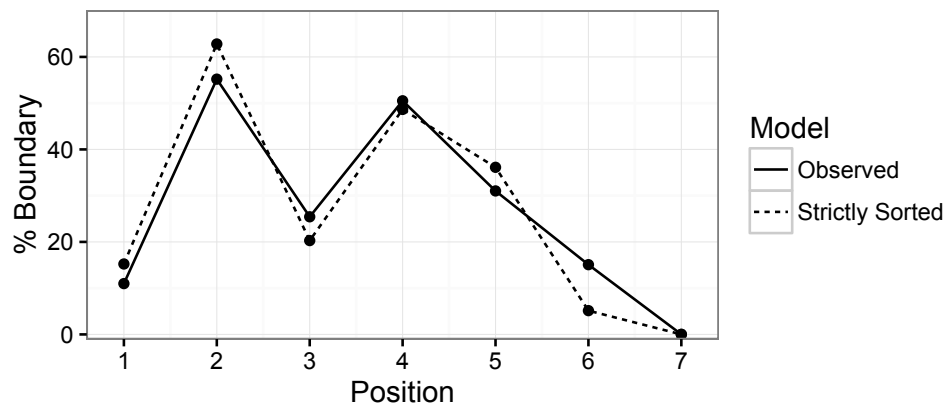


6 Word boundary distribution

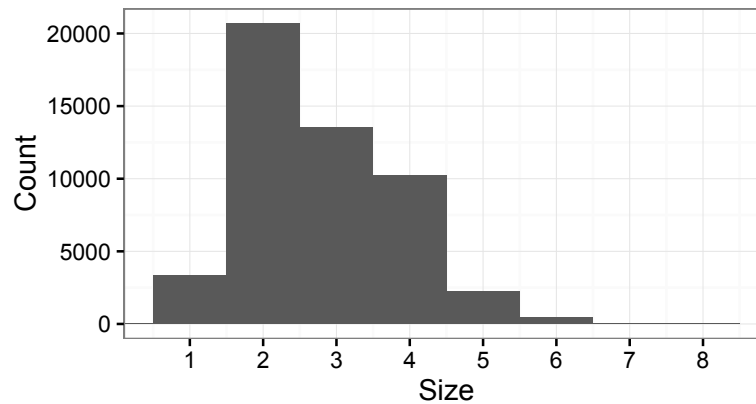
- (1) Traditional rule: A final monosyllable is prohibited. Cf. overall boundary distribution:



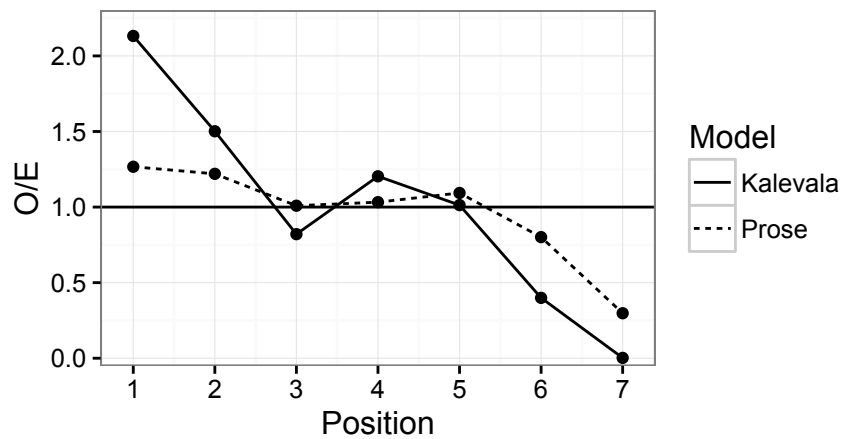
- (2) The decay doesn't look logarithmic, but it couldn't possibly. Cf. a model in which the words in each line are sorted from shortest to longest completely strictly.



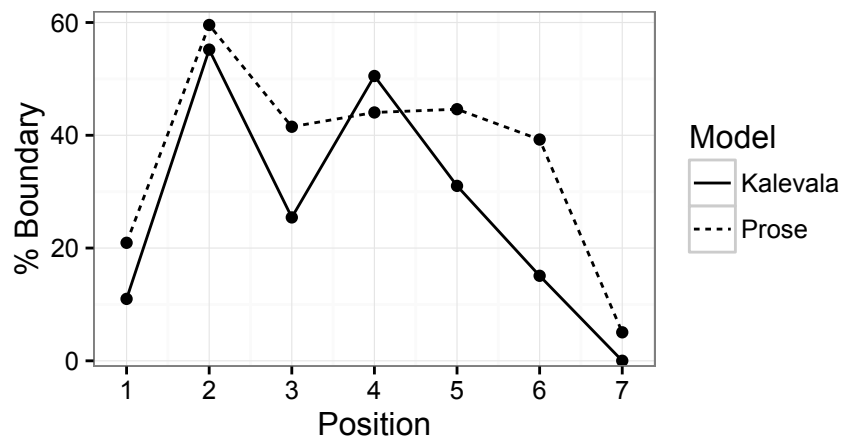
- (3) E.g. a break after position 1 is possible only with an initial monosyllable, but 85% of lines lack monosyllables altogether. Words are predominately 2–4 syllables:



- (4) The Kalevala is stricter than prose (here, 8,289 octosyllabic inter-punctuation phrases extracted from the Bible translation, *Vuoden 1776 Raamattu*). E values were obtained by randomly shuffling words within each line. Verse has much greater tilt.

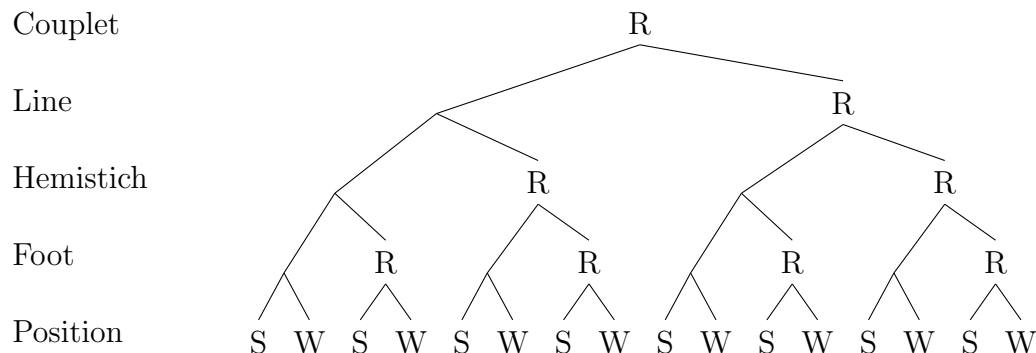


- (5) Cf. Kalevala vs. prose raw:



- (6) Thus, the prohibition on final monosyllables is only the most extreme manifestation of a more general tendency to avoid boundaries more stringently towards the end of the line.
- (7) Of course, one could also say “prefer longer words” instead of “avoid boundaries.” Avoiding boundaries is also tantamount to avoiding primary stresses.

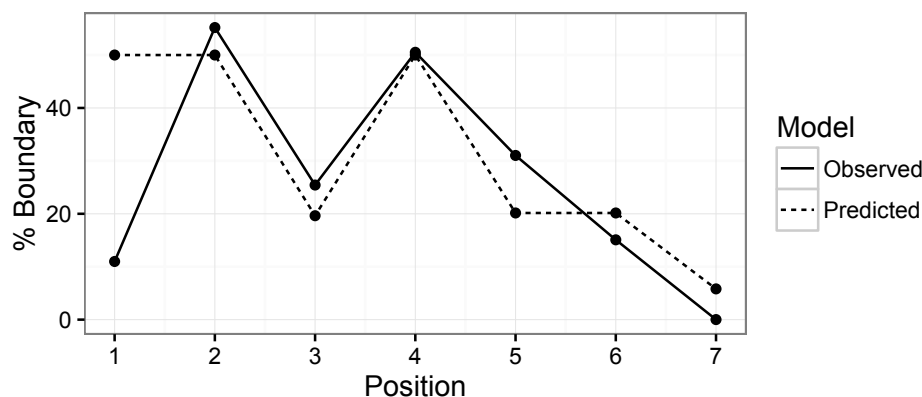
(8) As with positional stringency, we can index a constraint schema to R branches.



(9) Here, *BOUND: Penalize a boundary (within⁷ some R-domain), e.g.

- *BOUND/hemistich-R: Penalize each boundary within hemistich-R.
- *BOUND/foot-R: Penalize each boundary within foot-R.

(10) A maxent model with just these two constraints:



(11) Some issues with this approach:

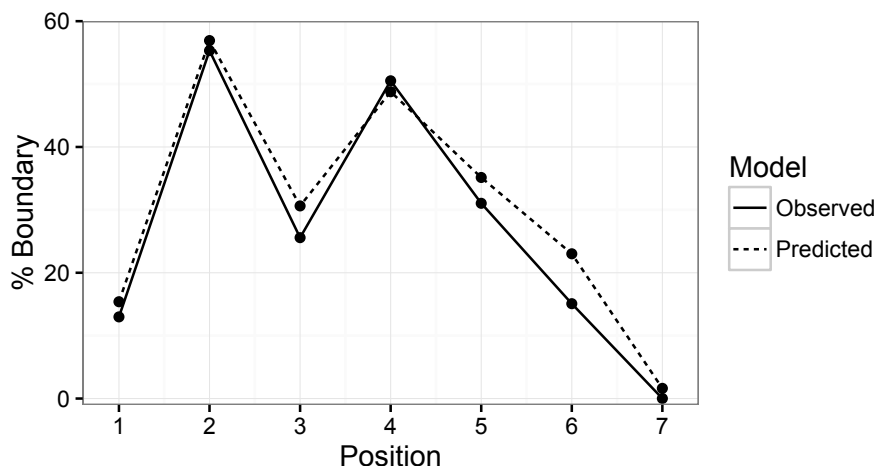
- Positions 1–2 are forced to be 50% because nothing regulates them.
- Boundaries are overgenerated at the end. The decay is rather anti-logarithmic — accelerating into the end — so the model’s right intercept is predictably too high (§3.9).
- Impressionistically, the first hemistich is free. But the model doesn’t know that, and compromises at the end to get a slightly better fit at the beginning.

(12) But perhaps the *BOUND-R schema is fine, and it’s the approach in (10) that’s flawed.

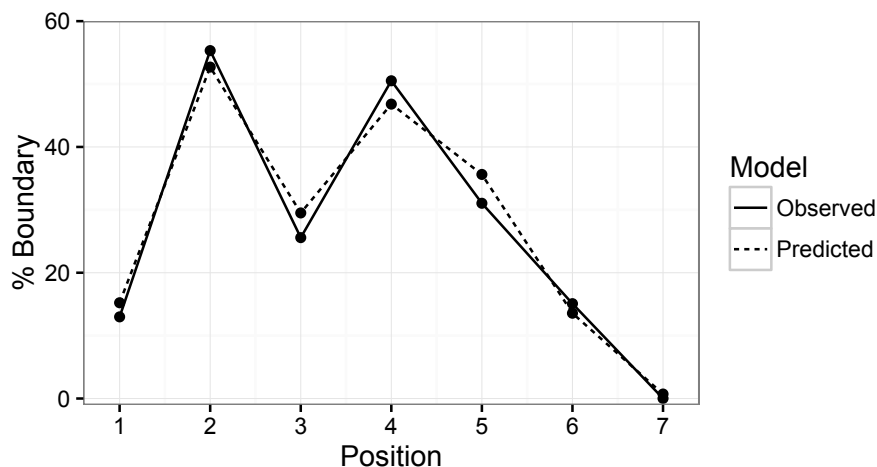
(13) We can envision the goal of the poetic grammar to take the natural language base and filter out verse-worthy lines. To this end, I combine verse and prose octosyllables into a single corpus to simulate the “base.”

⁷I take “within” to exclude boundaries at either periphery of the domain.

- (14) A logistic model can then predict whether each line is verse (1) or prose (0) using metrical constraints. The better the grammar, the better its separation of verse from prose.
- (15) Intercept-only model. The model ignores the form of the line, and blindly guesses that each line is verse with $p = .68$.
- The solid line is the real (target) distribution, as above.
 - The dotted line is the distribution of boundaries among lines the model predicted to be verse. In this case, it's essentially the average of verse and prose.



- (16) After adding *BOUND/hemistich-R and *BOUND/foot-R, the fit significantly improves.



References

- Arnold, Edward Vernon. 1905. *Vedic metre in its historical development*. Cambridge, U.K.: Cambridge University Press.
- Devine, Andrew M., and Laurence Stephens. 1975. The abstractness of metrical patterns: Generative metrics and traditional explicit metrics. *Poetics* 16.411–430.

- Halle, Morris, and Samuel Jay Keyser. 1971. *English stress: Its form, its growth, and its role in verse*. New York: Harper and Row.
- Hanson, Kristin, and Paul Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72.287–335.
- Hayes, Bruce. 2016. Stochastic constraint-based grammars for Hausa verse and song. Invited presentation, 13th Old World Conference in Phonology, Bupapest.
- Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.
- Hayes, Bruce, Colin Wilson, and Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88.691–731.
- Kimper, Wendell A. 2011. *Competing triggers: Transparency and opacity in vowel harmony*. Doctoral Dissertation, University of Massachusetts-Amherst.
- Kiparsky, Paul. 1968. Metrics and morphophonemics in the Kalevala. *Studies presented to Roman Jakobson by his students*, ed. by Charles Gribble. Cambridge, MA: Slavica.
- Lauerma, Petri. 2001. Larin Parasken metriikasta. *Virittäjä* 1.
- Leino, Pentti. 1986. *Language and metre. Metrics and the metrical system of Finnish*. Number 31 Studia Fennica. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Leino, Pentti. 1994. The Kalevala Metre and its development. *Songs beyond the Kalevala: Transformations of oral poetry*, ed. by Anna-Leena Siikala and Sinikka Vakimo, 56–74. Helsinki: Suomalaisen Kirjallisuuden Seura (SKS).
- Lönnrot, Elias. 1849. *Kalevala taikka vanhoja Karjalan runoja Suomen kansan muinoisista ajoista*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Manaster Ramer, Alexis. 1981. *How abstruse is phonology?*. Doctoral Dissertation, University of Chicago.
- Manaster Ramer, Alexis. 1994. Stefan George and phonological theory. *Phonology* 11.317–323.
- McPherson, Laura, and Bruce Hayes. 2012. Relating application frequency to morphological structure: The case of Tommo So vowel harmony. Handout from the Manchester Phonology Meeting, 24 May 2012.
- Oldenberg, Hermann. 1888. *Die Hymnen des R̥gveda I: Metrische und textgeschichtliche Prolegomena*. Berlin: Hertz. Reprinted 1982 by Steiner, Wiesbaden.
- Ryan, Kevin M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28.413–454.
- Sadeniemi, Matti. 1951. *Die Metrik des Kalevala-Verses*. Helsinki: Folklore Fellows Communications.
- Zuraw, Kie. 2012. Quantitative patterns of constraint interaction. Invited presentation, West Coast Conference on Formal Linguistics (WCCFL), UC Santa Cruz.
- Zymet, Jesse. 2015. Distance-based decay in long-distance phonological processes: Probabilistic models for Malagasy, Latin, English, Hungarian, and Sanskrit. MS., submitted.