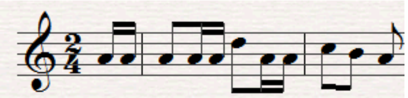
Bruce Hayes
UCLA

Swarthmore talk
April 8, 2009

The Textsetting Problem: the Intersection of Phonology, Music Cognition, and Computation

I. STATING THE PROBLEM

1. First verse of folk song; first line

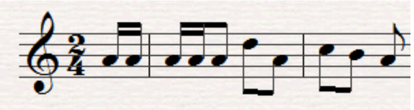
```
        x               x               x               x
x       x       x       x       x       x       x       x        =
x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
|   |   |       |   |   |       |   |   |       |       |
It  was late    in  the night   when the squire came    home
```

(“The Gypsy Laddie”, recorded in the Appalachian Mountains ca. 1917 by Cecil Sharp)

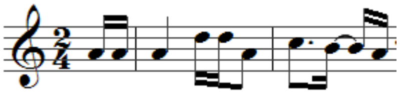
2. Note on grids

- Height of column = strength of beat
- Rows = theoretically isochronous levels of periodicity

3. A later verse, first line

“Oh saddle to me my milk-white steed”

```
        x               x               x               x
x       x       x       x       x       x       x       x        =
x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
|       |   |   |       |       |       |       |       |
Oh,     sad-dle to      me      my      milk-   white   steed
```

4. An ill-formed setting

```
        x               x               x               x
x       x       x       x       x       x       x       x        =
x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
|   |   |               |   |   |       |           |       |
Oh, sad-dle             to  me  my      milk-       white   steed
```

5. Shared intuitions

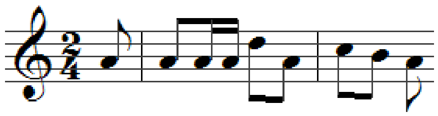
- Native speakers generally agree with one another on what settings should be preferred (Hayes and Kaun 1995—10 speakers, average of 2.2 settings per line).

6. Intuitions are sometimes gradient

- Example:

```
      x           x           x           x
x     x     x     x     x     x     x     x
x  x  x  x  x  x  x  x  x  x  x  x  x  x  x  x
|     |     |  |  |     |     |     |     |
Oh,   sad-  dle to me   my    milk- white steed
```

=



is perhaps not quite as good as the setting in (3), but surely not bad.

- Hayes-Kaun 1995 speakers show a modest preference for the type given in (3).
- This is typical, so we need to be able to predict such gradient intuitions as well.

7. The textsetting problem is a long established one

- Some references: Dell (1975, 2004), Stein and Gill (1980), Oehrle (1989), Halle and Lerdahl (1993), Halle (1999, 2004), Hayes and Kaun (1996), Hayes (in press), Keshet (2006 ms.)

8. Goals

- The analytical problem: find and state the principles that tacitly guide people when they set text in their language.
- We can and should do this explicitly—a machine implemented model that is trained from data and arrives at its own "intuitions" about textsetting.
    - ➢ a sort of micro-Turing test.
- Why address this problem?
    - ➢ We might learn more about musical and phonological structure.
    - ➢ We can test computational theories proposed as models of mental operations.

9. Overview of talk

- Theory of musical rhythm
- Phonological theory: phrasing, stress patterns
- Probabilistic, constraint-based grammars, and computational systems for learning them.

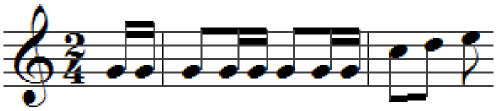
MUSICAL RHYTHM

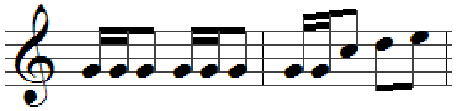
10. Metrical grids

- as above
- introducers: Lerdahl and Jackendoff (1983) *A Generative theory of Tonal Music* ; Liberman and Prince (1977) "On Stress and Linguistic Rhythm," *Linguistic Inquiry*

11. Purely-rhythmic principle (structural preferences)

- (These would hold true even in music without words.)
- From Lerdahl and Jackendoff (1983):
  - ➢ If a position is to be empty, then the weaker it is (few x's in grid), the better.
  - ➢ Accented elements (e.g. stressed syllables) should be placed in strong positions.
  - ➢ Strong elements are long. E.g. we mentally parse the notes below as on the left, not on the right:





PHONOLOGY

12. Word stress

- English is a language with basically phonemic (unpredictable) stress (cf. *thórough*/*Thoreáux*), and in general, the stressed syllables of words must fall in strong positions.
  - ➢ See (4), where mismatching *sáddle* produces a bad setting.
- Special strictness of word stress:
  - ➢ In poetry (Kiparsky 1975) and song (Hayes and Kaun 1995), it has been found that *stress* + *stressless* or *stressless* + *stress* tend to match the rhythm more strictly when the two syllables involved are in the same word.

13. Stress in phrases

- English has rules determining the stress pattern when words are combined into phrases.
  - ➢ Example: verb + particle, like *went on*, has rising stress; hence

```
          x                   x                   x                   x
x         x         x         x         x         x         x         x
x    x    x    x    x    x    x    x    x    x    x    x    x    x    x    x
          |         |         |         |    |    |         |    |    |
          He        went      on        till he   came      to   his  den
```

is slightly preferred to

```
          x                   x                   x                   x
x         x         x         x         x         x         x         x
x    x    x    x    x    x    x    x    x    x    x    x    x    x    x    x
|         |         |         |         |         |         |    |    |
He        went      on        till      he        came      to   his  den
```

despite the defect of leaving the first two positions empty.

14. Phrasing

- Stressed syllables at the *ends* of phrases strongly prefer to be in strong rhythmic positions (Kiparsky 1977, Hayes and Kaun 1995)
- Line endings must coincides with phrase endings — "run-ons" are disfavored (See Hayes and MacEachern (1996) "Are there lines in folk poetry?")

CONSTRAINT-BASED GRAMMARS

15. How to turn lists of constraints into explicit grammars?

- This is a major topic research in linguistics and related fields.
- One approach with a strong track record is Optimality Theory (Prince and Smolensky 1993 et seq.), the basis for much work in phonology.
- I here use a slightly different constraint-based approach, namely **maxent grammars** (Goldwater and Johnson 2003, Wilson 2006) Why?
    - ➢ We need to capture gradient intuitions (see (3) vs. (6) above).
    - ➢ Current Optimality-theoretic approaches don't converge (GLA: Pater 2008) or haven't been proven to converge.
    - ➢ The math of maxent has been completely worked out and is fully trustable.

16. Overall approach

- We find *every logically possible setting*
    - ➢ With the grids used here, this is never more than about 14,000, so with a bit of fairly elementary computer use we can check them all.
    - ➢ Checking all possibilities: essentially the "GEN" function of Optimality Theory.
- We set up a batch of constraints, and assess the number of constraint violations of each setting.
- Every constraint has a **weight**, a non-negative number that intuitively expresses its strength.
    - ➢ The higher the weight, the worse a setting that violates it is likely to be sound.
- From this, a standard formula (below) predicts for each setting a **probability**—claimed to match up with its degree of well-formedness.

17. Some sample probabilities

- I haven't yet explained how these are obtained, but these illustrate the ability of the system to match intuition at a rough level.

  He rode through woods and copses, too

- Top 6, plus two samples from the lunatic fringe. Probability is given in the right column.

		x				x				x				x		
x		x		x		x		x		x		x		x		
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
He		rode		through		woods		and		cop-		ses		too		0.817
	He	rode		through		woods		and		cop-		ses		too		0.090
He		rode		through		woods			and	cop-		ses		too		0.027
He		rode			through	woods		and		cop-		ses		too		0.027
He		rode		through		woods		and		cop-	ses			too		0.020
		he		rode	through	woods		and		cop-	ses			too		5×10^{-7}
	He	rode	through	woods	and	cops-				ses				too		$1.2\ 10^{-12}$

18. How the math of maxent works

- For each candidate, Compute the **harmony**, [1] which in notation is:

$$h(x) = \sum_{i=1}^{N} w_i C_i(x)$$

where w_i is the weight of the ith constraint,
$C_i(x)$ is the number of times that x violates the ith constraint, and
$\sum_{i=1}^{N}$ denotes summation over all constraints (C_1, C_2, … C_N).

- Compute the "**Maxent value**":

  Given a phonological representation x and its score $h(x)$ under a grammar, the *maxent value* of x, denoted $P^*(x)$, is:

$$P^*(x) = \exp(-h(x))$$

---

[1] The concept of *harmony* is developed in Smolensky (1986) and subsequent work (Smolensky and Legendre 2006).

- Compute the **probability**

  Given a phonological representation x and its maxent value $P^*(x)$, the *probability* of x, denoted $P(x)$, is:

$$P(x) = P^*(x) / Z \qquad \text{where } Z = \sum_{y \in \Omega} P^*(y)$$

  That is, its share of maxent values among all candidates.

19. Where do the weights come from?

- This is a long standing problem.
- The approach taken here assumes that they are *learned*—you attend to data from the musical idiom around you, and this gives you the information you need.
- The relevant algorithm (e.g. Della Pietra et al. 1997) attempts to **maximize the predicted probability of the observed data**, a standard criterion in computer science.
- For an attempted clear layman's explanation of how the algorithm works, see Hayes and Wilson (2008).
- I would be happy to share with you the maxent software (work of Colin Wilson/Benjamin George) I used to do the simulations; bhayes@humnet.ucla.edu.

THE SPECIFICS OF THE PRESENT ANALYSIS AND SIMULATION

20. Data corpus

- Hayes and Kaun (1996): 10 consultants each chanted the text of 670 lines of traditional English folk song, in rhythm.
- Goal is to model the share of the vote that each setting got—this can serve as an approximation for gradient intuition.

21. Linguistic annotation of the lines

- Hayes and Kaun independently transcribed the data:
  - **Stress values** for each syllable (as in Chomsky and Halle 1968)
  - **Phonological phrasing**, using rules from Hayes (1989)'s synthesis of earlier literature (Selkirk 1980, Nespor and Vogel 1982)
- They achieved reasonably good intersubjective agreement.

22. Grid — with labels of convenience for columns

```
      x           x           x           x
x     x     x     x     x     x     x     x
x  x  x  x  x  x  x  x  x  x  x  x  x  x  x  x
M  W  S  W  M  W  S  W  M  W  S  W  M  W  S  W
```

where W = Weak, M = Medium, S = Strong

23. Constraints used

- No time to do these in detail, but a quick outline.
- I would like to try trimming, adding; i.e. this is preliminary.
- The numbers are the weights that were learned for each constraint in the simulation.

- The ups and downs of stress must match the rhythm:

5.00	Regulate SW	"regulated" = stronger stress, or overt syllable vs. null
0.00	Regulate MW	(turned out to be useless)
1.11	Regulate SM	
0.32	*Phrase-Final Rise	special phonological context
1.19	*Word-Internal Mismatch of Stress	special phonological context
0.89	*Stress In M	
2.53	*Stress In W	

- Use of null vs. overt syllable must reinforce the rhythm:

5.00	Fill Strong	S positions can't be empty
2.40	Fill Medium	M positions can't be empty
2.40	Don't Fill W	W positions can't be filled

- Prefer to demarcate the lines with long pauses, by making their terminal positions empty:

2.59	Don't Fill 1
4.60	Don't Fill 16

- The durations of syllables as set in song must match their natural phonetic durations:

2.46	Non-Word Final Syllables are Short

- Inherent connections between metrical strength and duration:

1.37	Strong Is Long	Penalize gradiently when the S positions don't initiate a syllable linked to multiple positions.

- Avoid rhythmic obscurity:

4.75	Avoid Lapse	*3 empties in a row

- Other:

  1.40 WEAK RESOLUTION stressless in S wants to be short — dunno why…

24. The simulation

425	Lines in the corpus (I removed lines found only in some stanza types)
8.4	Average # valid “votes” per line / 9 subjects
2.2	Average # of distinct settings among the votes

- Goal: find weights that predict the distribution of votes as accurately as possible
- I also did “cross-training” runs: train on one half, test on other; this yielded similar results.
- I used maxent software created by Colin Wilson.

25. Results I: sample output

This was shown above in (17).

26. Statistical report of results

- For the entire set of candidates, the correlation r of predicted probability vs. “vote share” is r = 0.883.
- This is only a rough measure, since most values for both voting and prediction are at or close to zero.

27. Results II: Data and predictions in bins

Predicted probability

% volunteered by consultants

	0 - .1	*.1 - .2*	*.2 - .3*	*.3 - .4*	*.4 - .5*	*.5 - .6*	*.6 - .7*	*.7 - .8*	*.8 - .9*	*.9 - 1*
0 - .1	48462	191	41	10	7	3	1			
.1 - .2	259	34	19	4	3	3	2	1	1	
.2 - .3	67	13	10	4	2	2	5		1	1
.3 - .4	26	12	11	1	4	2	4	3	3	
.4 - .5	12	13	6	3	6	3	2	4	4	
.5 - .6	6	6	8	4	8	3	7	3	7	
.6 - .7	3	1	5	5	3	6	17	6	14	1
.7 - .8	4	5	2	4	4	6	12	6	18	1
.8 - .9	2	4		4	3	12	20	13	33	5
.9 - 1		2	1	2	4	9	28	24	27	12

28. Improvements possible?

- The constraints could be improved, I think.
- Keshet (2006), working non-gradiently, has discovered some new and interesting rules, but I’ve not had time yet to implement them.

29. Differences between consultants

- Hypothesis: the set of constraints embodies the general theory, part of the competence of all participants.
- Individual idiosyncrasies must be due to consultant-specific weighting.
- We can detect this by training the weights on the data specific to each consultant.
- Example: RH vs. DS's weights for two constraints, which often conflict.

	Non-Word Final Syllables are Short	Strong is Long
RH	1.472	3.418
DS	2.480	0.879

"The remarkable day that I was wed"

Consultant DS's setting satisfies Non-Word Final Syllables are Short:

```
          x                   x                   x                   x
x         x         x         x         x         x         x         x
x    x    x    x    x    x    x    x    x    x    x    x    x    x    x    x
|    |    |    |    |         |         |         |         |         |
The  re-  mar- ka-  ble       day       that      I         was       wed
```

Consultant RH's setting satisfies Strong is Long:

```
          x                   x                   x                   x
x         x         x         x         x         x         x         x
x    x    x    x    x    x    x    x    x    x    x    x    x    x    x    x
|    |    |         |    |    |         |         |         |         |
The  re-  mar-      ka-  ble  day       that      I         was       wed
```

30. DS and RH's own grammars predict these settings as favorites

Probabilities:

	RH's choice	DS's choice
RH's grammar	0.689	0.065
DS's grammar	0.251	0.819

31. Upshot

- The maxent approach not only characterizes the data as a whole fairly well, but gives us a means of characterizing individual differences in style.

32. Caveat: do RH and DS really have different grammars?

- Maybe, but my guess is that they are construing the experimental situation differently:

- Each commands a variety of idioms.
- They accessed different ones in performing the experimental task.

33. Summary: Situating the approach

- The textsetting problem has traits seen elsewhere in cognitive science.
    - An identifiable **structural basis**, with a need for theoretical ideas taken from generative linguistics and formal music cognition.
    - Extensive **gradience** of native speaker intuitions and behavior, long a barrier to the use of structural approaches.
    - An **"apples and oranges"** problem, in which we have to weight the relative importance of constraints that have quite different teleologies.
- I think the right approach to such problems is a kind of "statistical generativism" (e.g., Boersma and Hayes (2001), Yang (2002)
    - Traditional structural constraints are used, but
    - …embedded in a quantitative system that predicts gradience, and
    - … fine-tunes the grammar in response to learning data

- This kind of research implies we need corpora, experiments, easy-to-apply computational models. This is more work but I think the work can be fun and gets us more accurate and insightful results.

References

Dell, François (1975). Concordances rythmiques entre la musique et les paroles dans le chant: l'accent de l'e muet dans la chanson française. In Le soucie des apparences, Marc Dominicy (ed.), 121-136. Brussels: Editions de l'Université de Bruxelles.

Dell, François (2004) Singing as counting syllables: text-to-tune alignment in traditional French songs. Ms.

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

Halle, John and Fred Lerdahl (1993) "A generative textsetting model," Current Musicology 55:3-23.

Halle, John (1999) A Grammar of Improvised Textsetting. Ph.D. dissertation, Columbia University.

Halle, John (2003) Constituency matching in metrical texts. Submitted for publication in the proceedings of the conference Words and Music, University of Missouri-Columbia, March 14, 2003.

Halle, Morris and S. Jay Keyser (1966) "Chaucer and the theory 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. New York: Harper and Row.

Hanson, Kristin (1990) Resolution in Modern Meters, Ph.D. dissertation, University of California, Berkeley.

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, pp. 201-260.

Hayes, Bruce (in press) “Textsetting as constraint conflict,” to appear in Aroui, Jean-Louis and Andy Arleo, eds. (forthcoming) Towards a Typology of Poetic Forms. Amsterdam, Elsevier.

Hayes, Bruce and Margaret MacEachern (1998) “Quatrain form in English folk verse,” Language 74, 473-507.

Hayes, Bruce and Abigail Kaun (1996) “The role of phonological phrasing in sung and chanted verse,” The Linguistic Review 13, 243-303.

Keshet, Erza (ms., 2006) “Relatively Optimal Textsetting,” http://web.mit.edu/ekeshet/www/

Kiparsky, Paul (1975) “Stress, Syntax, and Meter,” Language 51:576-616

Kiparsky, Paul (1977) “The rhythmic structure of English verse,” Linguistic Inquiry 8.189-247.

Lerdahl, Fred and Ray Jackendoff (1983) A Generative Theory of Tonal Music. Cambridge, MA: MIT Press.

Marr David (1983) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W. H. Freeman.

Oehrle, Richard (1989). Temporal structures in verse design. In Paul Kiparsky and Gilbert Youmans, eds., Rhythm and Meter, Academic Press, Orlando, FL, pp. 87-119.

Stein, David, and David Gil (1980) Prosodic structures and prosodic markers. Theoretical Linguistics 7. 173-240

Temperley, David (2001) The Cognition of Basic Musical Structures. MIT Press, Cambridge.

Temperley, David (2006) Music and Probability, MIT Press, Cambridge.

Yang, Charles (2002) Knowledge and learning in natural language. Oxford University Press

Bruce Hayes
http://www.linguistics.ucla.edu/people/hayes/
bhayes@humnet.ucla.edu