Brown University Minicourse
May 21, 2009

B. Hayes
UCLA

Embedding Grammar in a Quantitative Framework:
Case Studies from Phonology and Metrics

Class 5: Maxent Metrics

1. Outline

- Metrics: the issues and why they have not been resolved
- Maxent grammars
- Data — Shakespeare and Milton
- Constraints for metrics
- Research questions
- Findings
- Course summary

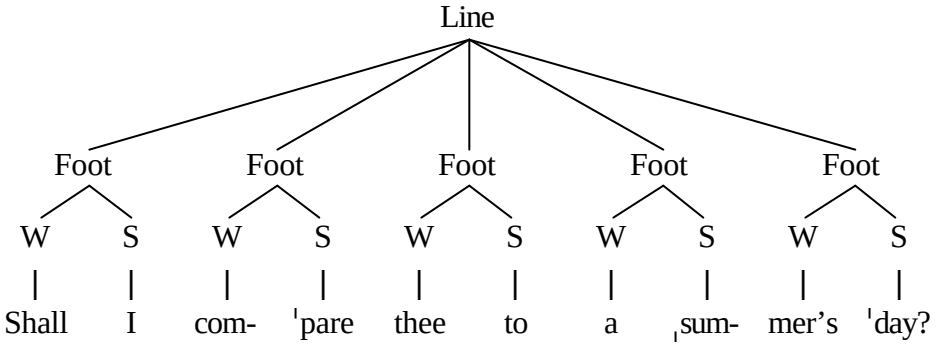
2. Coauthors

- Most of this talk represents current work being done with Anne Shisko (UCLA) and Colin Wilson (Johns Hopkins University).

BACKGROUND I: GENERATIVE METRICS

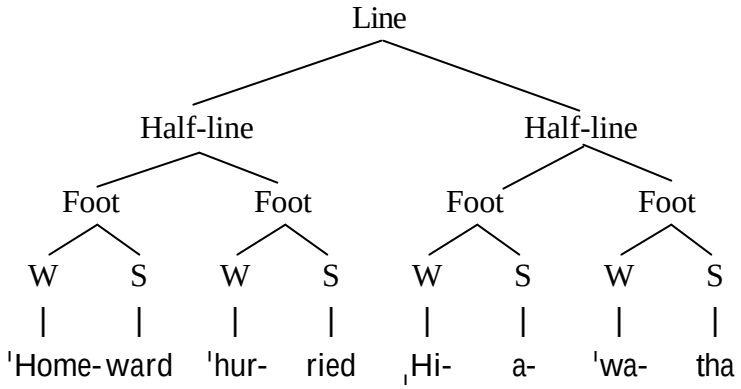
3. Object of study

- Phonological representations are constructed by poets to be arrangeable in correspondence with rhythmic representations (meters).
- Iambic pentameter:



Shakespeare, Sonnet 18

- Trochaic tetrameter:



Longfellow, "Song of Hiawatha"

- The **meters** (in languages like English) can be represented as constituent hierarchies, with Strong and Weak terminal positions. These are not too controversial, except perhaps at the highest levels.
- Harder: stating the well-formedness conditions on such correspondence
- Any theory that attempts to state these conditions explicitly as a grammar belongs to the field of **generative metrics** (Halle and Keyser 1966 et al., et seq.)

4. Proposed constraint types for English metrics

- **Matching stress** to W and S positions—many different ways to do this have been tried.
- **Matching the phonological bracketing** (phrases, word) to the bracketing of the meter.
- **More complex constraint types** (see below)

5. Metrics is not peripheral

- Kiparsky (1987:195–196): "... literature is universal .... There is no tribe on earth so wretched that it does not express its memories and desires in stories and poems." [1]
- In traditional societies, verse is ubiquitous, usually sung or changed, and shared by a whole community—see Dell and Elmelaoui's ethnographic background on Berber verse.
- The near-disappearance of conventional poetry from our culture (ca. time of Robert Frost) created a gap that quickly was filled by other, vernacular means, such as rap.
- I judge that in normal human societies, a large fraction of the input data available to the average child is sung or chanted.

[1] Kiparsky, Paul (1987). On theory and interpretation. In Nigel Fabb, Derek Attridge, Alan Durant, and Colin MacCabe (eds.), *The Linguistics of Writing: Arguments between Language and Literature*, Manchester: Manchester University Press, pp. 185–198.

6. Early history of generative metrics—w.r.t. English iambic pentameter

- Halle and Keyser (1966, 1971) got the field started by stating the problem clearly, stating what would constitute a explicit solution, and proposing an initial (far too simple) account.
- Kiparsky (1975, 1977) established the intricacy of the problem:
  - It's not just stress-matching, but a complex interaction of stress and bracketing (word level, phrase-level/foot level, etc.) that establishes metricality.
  - Analysis of **metrical idiolects**: Shakespeare's practice is not Milton's, etc.
- Hayes (1983, 1989) proposed various empirical improvements on Kiparsky's rules, relating them to the then-emerging theories of metrical stress and phonological phrasing, and establishing **patterns of gradience** based on the rank of phonological phrase types.
- Youmans (1989), Hammond (1991) found interesting further constraints.
- And of course other languages and traditions are studied…

7. The iambic pentameter problem strikes me as being stalled

- Different scholars draw different conclusions about what was right about the research of the 1980's, or have moved off in directions that do not depend on those results.
- Why (do I think) this is so?

8. Difficulties

- The **counterexample problem**: constraints proposed as inviolable face counterexamples (see, e.g., Tarlinskaja 1996).
  - Our own study (below) readily finds counterexamples to most constraints that have been proposed.
- The proper reply is to move to a **gradient theory**: lines are assessed gradiently rather than being outright metrical or unmetrical.
  - This matches most people's metrical intuitions:

The lion dying thrusteth forth his paw	*perfect*
Pluck the keen teeth from the fierce tiger's jaws	*difficult*
Ode to the West Wind by Percy Bysshe Shelley	*horrible*

  - This idea was proposed, and roughly implemented, by Youmans (1989), Hayes (1989), others.
- This has produced an **explicitness problem**: no one in the field knew any principled way to formulate gradient metrical grammars.
- In addition: we have a severe **assessment problem**: how rare should the lines violated by a constraint be to make the rarity meaningful?

APPLICATION OF MAXENT TO METRICS

9. Two kinds of maxent, two kinds of metrics

- The Hungarian vowel harmony work of Class 2 was carried out with "conditional random fields"-type maxent, useful for analyzing choices (given stem, what suffix?)
- The English phonotactics work of Class 3 was carried out with "unconditional random fields" maxent, useful for assessing gradient membership in the class of well-formed objects.

10. Metrics has both

- There are choice problems, and well-formedness problems

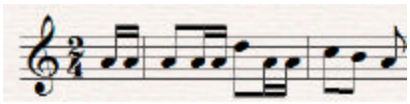
11. Metrics as a choice problem: textsetting

- Given:
  - a rhythmic pattern (like the rhythm of a song already in progress)
  - a linguistic text (like the text of a new verse the singer is improvising)

  what is the best way, or ways, to set the syllables to the rhythm?

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
```

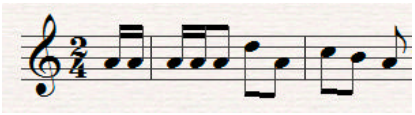


("The Gypsy Laddie", recorded in the Appalachian Mountains ca. 1917 by Cecil Sharp)

- Grids à la Lerdahl and Jackendoff (1983)
  - Height of column = strength of beat
  - Rows = theoretically isochronous levels of periodicity

A later verse, first line:
A. "Oh saddle to me my milk-white steed"

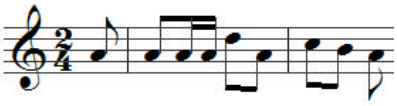
```
      x       x       x       x
x     x   x   x   x   x   x   x
x x x x x x x x x x x x x x x x x
|     | | |     |     |     |
Oh, sad- dle to me my milk- white steed
```



or perhaps:

B.

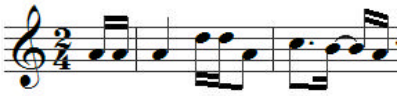
```
        x           x           x           x
    x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
  Oh,  sad-  dle  to  me  my  milk-  white  steed
```



but not:

C.

```
        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
```



12. A possible analytic approach

- Same constraints as below, using maxent in its "choice" mode—pick from the several thousand possible alignments of text and grid.
- This seem to work pretty well
  - (Hayes, Bruce (2008) "The Textsetting Problem: the Intersection of Phonology, Music Cognition, and Computation". Handout; course web site.)

13. Other references

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.
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.

14. Meter as a well-formedness problem

- Iambic pentameter counts syllables, and the alignment to the grid is very straightforward (other than a little bit of phonology, changing syllable count)
- Thus there is no compelling reason to do metrics as a matter of choices.
- Instead it's quite similar to phonotactics, though with larger objects.
- We suggest that metricality can be analyzed as probability, exactly as in the phonotactic discussion earlier.

15. Tacit conjoined constraint in metrics

- Constraints **overlap** in their effects, especially complex ones.
- Example:
  - Constraint X: "A line is bad when conditions A and B are present"
  - Suppose we also have

Constraint Y: "A line is bad when A is present"
Constraint Z: "A line is bad when B is present"

- Then it is possible that constraint X, while racking up impressive statistics, is *redundant*.
- The metrics literature frequently fails to consider this possibility.
- Maxent can help find this—constraint weights are additive, so any conjoined constraint will have to prove its worth by adding explanatory value (measured by predicted probability of the corpus.

16. Comparing grammars of different sizes

- When you add a constraint to a grammar, and predicted probability improves, this *may or may not* be meaningful.
- You've added a degree of freedom, so a minor improvement might not be worth the cost.

17. A new method of assessing the worth of added constraints

- Colin Wilson has developed a new version of the learner used in Hayes and Wilson (2008).
- It not only knows the gradient (direction up the slope, used for weighting), but also *computes the predicted probability* of the data corpus.
- Once you know this, you can carry out a **likelihood ratio test**:

$$2 * \log\left(\frac{\text{data prob. with constraint}}{\text{data prob. without constraint}}\right)$$

  - This is distributed as a chi-square with one degree of freedom and yields the probability of the hypothesis "improvement could be true by accident."

- Constraints with zero weights always flunk this test, and other constraints sometimes do—usually when their weights are low.
- We're interested in which constraints survive, and how big their weights are.

DATA

18. Corpora

- William Shakespeare, *The Sonnets* (2154 pentameter lines)
- John Milton, *Paradise Lost*, Books VIII and IX (2293 lines)

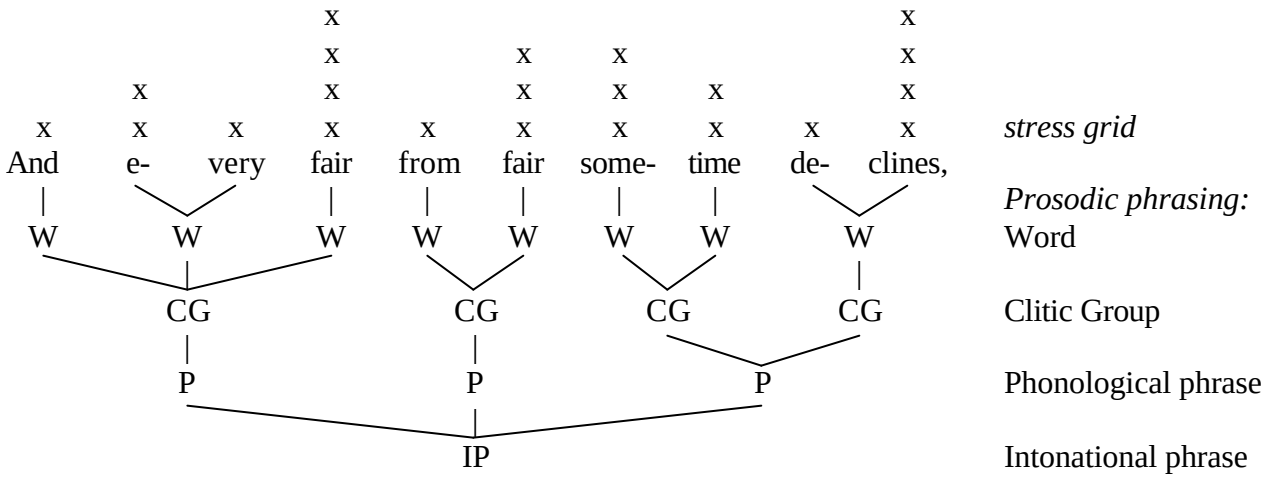
These are very well studied, having been core data for most of history of metrical study.

19. Annotation

- To apply constraints automatically, you need to encode each line for its metrically-relevant phonological properties.

- For every syllable of the line:
  - Its **level of stress**, on a 1-4 scale
  - The **rank** of the highest level of phrasal category for which it is the rightmost syllable, on a 1-5 scale.
- We followed literature in phonological theory (e.g. Hayes 1995, Selkirk 1982, 1986, Nespor and Vogel 1986, Hayes 1989) in assigning default stressings and phrasings.

20. Example annotation of a line for stress and phrasing



Compact numerical notation for the above:

	And	e	very	fair	from	fair	some-	time	de	clines,
stress	1	2	1	4	1	3	3	2	1	4
juncture	2	1	2	4	2	4	2	3	1	5

21. Independent coders (B. Hayes/A. Shisko) get reasonably good agreement

- Milton data:

Stress			Phrasing		
same	20479	0.890	same	21457	0.933
1 off	2092	0.091	1 off	1406	0.061
2 off	377	0.016	2 off	120	0.005
3 off	55	0.002	3 off	11	0.000
total	*23003*		4 off	9	0.000
			total	*23003*	

THE THEORY OF METRICS

22. Goal

- Treat the data with all the best resources of the literature on metrics.
- We do not necessarily expect everything will be helpful…

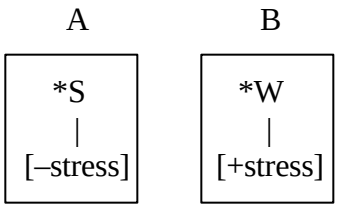
23. Expository Scheme

- We give general principles, then specific constraints that implement the general ideas in particular ways, then completely formalized constraints that can be read by our software.

THEORY OF METRICS I: STRESS MATCHING

24. "Fill S positions with stress, W positions with stressless"

- This is traditional wisdom; see also Kiparsky (1989), Hansen (1992), Hanson and Kiparsky (1996).
- Constraints:

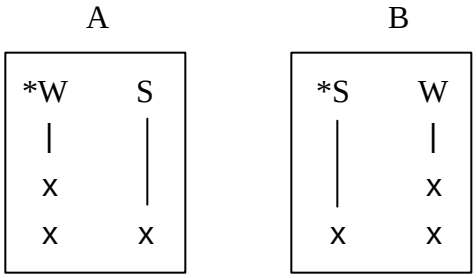


- The computer sees:

```
A:  *[+Strong,–Accent]
B:  *[–Strong,+Accent]
```

  - where [Strong] distinguishes the S and W positions of the meter.
  - where [Accent] is our feature that distinguishes stressed (any degree) from stressless

25. "Match ups in meter with ups in stress, and downs with downs"

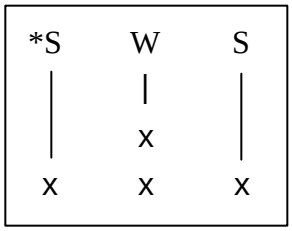


- Otto Jespersen conceived this constraint in 1900.
- later pursued by Kiparsky (1977), Hayes (1983, 1989)

- The computer sees:
  ```
  A:  *[–Strong,+Fall]
  B:  *[+Strong,+Rise]
  ```
  - where [Rise] and [Fall] are features of convenience computed from the numerical stress levels.

26. The Stress Maximum Principle

- If a single stressed syllable violates both A and B above at the same time, this counts as a violation of Halle and Keyser's (1966, 1971) **Stress Maximum Principle**



- In OT this would be **local constraint conjunction**; see Smolensky (1995) et seq.
- The SMP took a lot of flak when introduced because poets often violate it.
- However, it might still be valid in helping to assign gradient well-formedness.
- The computer sees: `*[+Strong,+Rise][+Fall]`
- We also included versions requiring all three syllables to be in the same Phonological Phrase, or same Intonational Phrase—this is part of Halle and Keyser's original prosposal
- The computer sees:
  ```
  *[+Strong,+Rise,–J4][+Fall,–J4]
  *[+Strong,+Rise,–J5][+Fall,–J5]
  ```
  where [–J4] = "in same Phonological Phrase with the next syllable"
  [–J5] = "in same Intonational Phrase with the next syllable"

27. Match ups/downs with great strictness when the sequence is within a simplex word

- Magnuson and Ryder (1970), Kiparsky (1975)
- A pair from Kiparsky. Shakespeare wrote:

	x	x			x	x			
Pluck	the	keen	teeth	from	the	fierce	ti-	ger's	jaws
W	S	W	S	W	S	W	S	W	S

Kiparsky revised:

	x	x			x	x			
*Pluck	im-	mense	teeth	from	en-	raged	ti-	ger's	jaws
W	S	W	S	W	S	W	S	W	S

28. Simple versions of this principle

A. Falling stress:

	*W	S	
	|	|	
	x	|	
[w…	x	x	…]w

B. Rising stress:

	*S	W	
	|	|	
	x	x	
[w…	x	x	…]w

- The computer sees:

```
A. *[–Strong,+Accent,-J2]
B. *[+Strong,+Rise,–J2]
```

  - where [–J2] means "this syllable belongs in the same simplex word as the next syllable."

29. More elaborate versions

- "**Inversion**": A special licence is granted when the mismatch comes at the beginning of some kind of phonological phrase.
- Example from the *Sonnets*:
  [IP Richer than wealth,]IP [IP prouder than garments' cost]IP
- We can state this as a requirement that a falling mismatch be preceded by a phrase break.

If:

	*W	S	
	|	|	
	x	|	
[w	x	x	…]w

Then:

the context must be / [IP ___ (Intonational-phrase-initial)

This has various flavors:

- The two syllables are specified as being either
  - in the same word (as above, called a **lexical inversion**), or
  - any *stressed* + *stressless* sequence.
- The phrase type specified can be IP, PP, CG (order of increasing strictness).
- The computer sees:

```
*[^+Strong,+J5][–Strong,+Accent,–J2]
*[^+Strong,+J4][–Strong,+Accent,–J2]
*[^+Strong,+J3][–Strong,+Accent,–J2]
```

```
*[^+Strong,+J5][–Strong,+Fall]
*[^+Strong,+J4][–Strong,+Fall]
*[^+Strong,+J3][–Strong,+Fall].
```

- ^ means "* unless"
- [+J4] means "syllable ends a Phonological Phrase
- [+J5] means, "syllable ends an Intonational Phrase."

30. The lexical inversion constraints are thought to be weaker for Milton

- His occasional lexical inversions *not* following a break have long been noticed:

Universal reproach, far worse to bear	*Paradise Lost* 6.34
To the Garden of Bliss, thy seat prepared	8.299
By the waters of life where're they sat	11.79
w s w s w s w s w s	

31. The Kiparskyan "mystery constraint"

- Source: Kiparsky (1977), revised by Hayes (1983)[2]
- Mismatches such as (30) are, strangely, usually preceded by a *stressless* syllable—

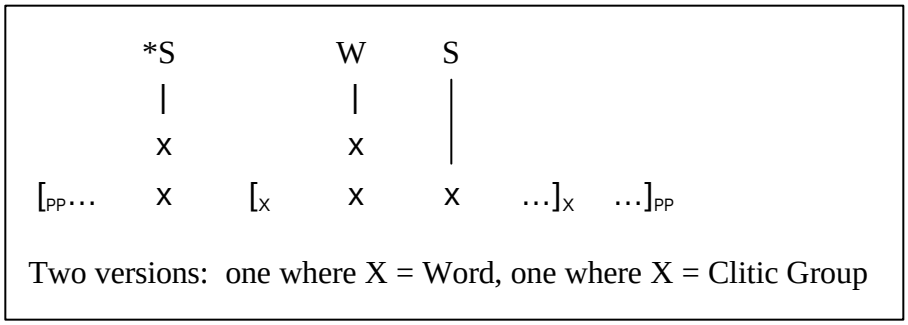
OK:	By the waters of life where're they sat	
BAD:	*By *still* waters of life where're they sat	
	w s w s w s w s w s	

why require an S position to be filled with stressless??

- The same holds for Shakespeare, when he commits non-lexical inversion:

OK:	If *I* dream not, thou art Emilia	*Comedy of Errors* 5.1.346
BAD:	If *Will* dream not, thou art Emilia	
	w s w s w s w s	

- Proposed constraint schema:



Two versions: one where X = Word, one where X = Clitic Group

[2] Our findings demonstrate that Hayes's revision was correct.

- The encompassing $[_{PP}$ …$]_{PP}$ brackets are needed since otherwise the configuration is just an ordinary post-break inversion.

The computer sees:

```
*[+Strong,+Accent,−J4][+Fall,−J3]
*[+Strong,+Accent,−J4][+Fall,−J2]
```

32. Special strictness for phrase-final positions

- This is the major finding of Kiparsky (1977).
- A rising stress sequence *if phrase-final* should not mismatch the meter.
- I.e., a location of (claimed) special strictness.

*S	W	
|	|	
|	x	
<x>	x	$]_X$

where $]_X$ is the right edge of a phrasal category (IP, PP, CG)…

- Hayes (1989) suggests a **gradient** system: the higher-ranking $]_X$ is, the worse the violation.

33. "Dialect variation" in the phrase-final constraint

- Kiparsky (1977) claims two flavors.
  - Shakespeare: violation only if the <x> position is fully stressless.
  - Milton: any sort of rise violates it.
- Thus, Shakespeare likes to write lines like these, which are unusual in Milton:

> But, like a sad slave, stay and think of naught
> w s w s w s w s w s
> Resembling strong youth in his middle age
> And see the brave day sunk in hideous night

- Shakespeare himself avoids lines like these (common in Wyatt, Frost):

> For good is the life, ending faithfully — Wyatt, "The Longe Love"
> w s w s w s w s w s
>
> One by one he subdued his father's trees — Robert Frost, "Birches"
> By riding them down over and over again

34. Implementation

- So we tried six versions, varying whether <x> in (32) is specified as stressless, and giving three ranks to $]_X$.
- The computer sees:

```
*[+Strong,–Accent,+Rise,–J5][+J5]
*[+Strong,–Accent,+Rise,–J4][+J4]
*[+Strong,–Accent,+Rise,–J3][+J3]
*[+Strong,+Rise,–J5][+J5]
*[+Strong,+Rise,–J4][+J4]
*[+Strong,+Rise,–J3][+J3]
```

35. The last foot is special

- See, e.g. Youmans (1989).

> *Falling stress in last two positions
> *Non-rising stress in the last two positions

where the first is expected to be a stricter constraint.

The computer sees:

```
*[–Rise][][+lb]
*[+Fall][][+lb]
```

where [ ] is any position and [+lb] denotes line terminus.

36. Summary of stress matching

- Raw (24)
- Relative (25)
- Relative, conjoined (26)
- Within a word (27)-(30)
- At end of phrase (32)-(34)
- In final foot (35)

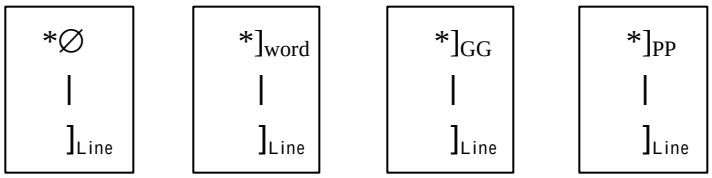
THEORY OF METRICS II: MATCH BRACKETING

37. Other, obvious traditions

- Serbo-Croatian epic decasyllable (Jakobson 1952) has **bridges**: no Clitic Group break may occur in the middle of the last foot of either half line.
- Japanese children's verse has stong **caesuras**: word boundary needed at Line Breaks, Half Line breaks, and favored at quarter-line breaks.

38. English: Line Demarcation

- A line should begin and end with a big phrasing break.
- We use a family of ALIGN (McCarthy and Prince, 1993) constraints, recording the strength of a break:



- Sample violations, going from left to right:

  I caught this morning morning's minion king-
  dom of daylight's dauphin… G. H. Hopkins, "The Windhover"

  My soule doth tell my body that he may]Word
  Triumph in love, flesh stays no farther reason. Sonnet 151

  Those hours that with gentle work did frame]CG
  The lovely gaze where every eye doth dwell Sonnet 5

  [Last type is numerous, over 10% of all lines.]

- The computer sees:

```
*[–J2][+lb]
*[–J3][+lb]
*[–J4][+lb]
*[–J5][+lb]
```

39. Line cohesion

- It's bad to put a big break in the middle of a line.
- We go with two versions:
  - PP-break or higher

    *[Line …]PP[PP …]Line
  - IP-break or higher.

    *[Line …]IP[IP …]Line

- The computer sees:

```
*[+J4][]
*[+J5][]
```

40. Foot Cohesion

- We go with two versions:
  - PP-break or higher

*[Foot …]PP[PP …]Foot

  - IP-break or higher.

*[Foot …]IP[IP …]Foot

- Computer sees:
  `*[–Strong,+J4]`
  `*[–Strong,+J5]`
- We added special versions for initial and final feet (1+9, 9+1 line division)

*[Foot …]IP[IP …]Foot]Line
*[Line[Foot …]IP[IP …]Foot

- Computer sees:
  `*[+J5][][+lb]`
  `*[+lb][+J5]`

41. Poetic variation

- Milton is noted as being permissive here—he writes in "verse paragraphs" like this one discussing his blindness.

> Then feed on thoughts, that voluntarie move
> Harmonious numbers; as the wakeful Bird
> Sings darkling, and in shadiest Covert hid
> Tunes her nocturnal Note. Thus with the Year
> Seasons return, but not to me returns
> Day, or the sweet approach of Ev'n or Morn …

- So we expect these constraints to be weaker for Milton than Shakespeare.

CONSTRAINTS III: OTHER

42. Phonology

- A more realistic analysis assumes that the poet can only make use of the phonological resources of the language—these should never be given a spurious metrical explanation.
- So we include constraints of English phonology:
  - *[word [–stress][–stress] (*[pʔtʔ'tæmʔr])

- *[+stress][+stress] within a single word (not allowed in our representational system; see Kiparsky (1977), Hayes (1983)
- Bans on unaccented monosyllabic Clitic Groups, Phonological Phrases, Intonational Phrases.
- *[$_{word}$ … [–stress][–stress][–stress] …]$_{word}$

- The computer sees:

```
*[^–Rise,–J2][–Rise,–Fall,–J2]
*[–Rise,–Fall][+Accent]
*[+J3][–Accent,+J3]
*[+J4][–Accent,+J4]
*[+J5][–Accent,+J5]
*[–Accent,–J2][–Accent,–J2][–Accent]
```

HYPOTHESES

43. Some constraints will disappear as redundant

- Where to look: complex constraints whose effect could be the result of the combined weights of their simpler "ingredients", also in the grammar
- The Stress Maximum Constraint (overlaps with various other constraints).
- Kiparsky's phrase-final constraints (32) are a conjunction of (25) (match stress) and (40) (foot cohesion).[3]

44. Gradience and juncture continua

- Constraints with juncture continua (W-CG-PP-IP): if valid, they should reach significance for *all* juncture values, following the general approach to gradience given in Hayes (1989).

45. Shakespeare and Milton are different

- We expected substantially laxer line-cohesion and line-demarcation for Milton, the author of "verse paragraphs".
- Different phrase-final filters (33): Kiparsky claims that Milton uses an "any sort of rise" constraint, and that Shakespeare uses a "don't rise out of stressless" constraint.
- Kiparsky claims Milton, unlike Shakespeare, has a certain tolerance for falling lexical inversions (30).

[3] Thanks to Kristine Yu for pointing this out to us.

RESULTS

46. Results of weight assignment for the Shakespeare and Milton corpora

- Conventions:
  - If a weight came out zero, we show a blank.
  - If a p-value came out > .05, we show a blank.

	Shakespeare			Milton		
	weight	**p-value**	**viols**	**weight**	**p-value**	**viols**
(a) Fill S positions with stress, W positions with stressless ((24))						
*[–Strong,+Accent]	**2.21**	< .001	2,219	**2.19**	< .001	2,324
*[+Strong,–Accent]			5,017			5,254
(b) Match ups in meter with ups in stress, and downs with downs ((25))						
*[+Strong,+Rise]	**0.52**	< .001	776	**1.03**	< .001	573
*[–Strong,+Fall]			1,000			1,284
(c) Stress Maximum Principle ((26))						
*[+Strong,+Rise][+Fall]			208			150
*[+Strong,+Rise,–J5][+Fall,–J5]			170			110
*[+Strong,+Rise,–J4][+Fall,–J4]			92			66
(d) Match within simplex word ((27))						
*[+Strong,+Rise,–J2]	**3.30**	< .001	4	**1.06**	< .001	24
*[–Strong,+Accent,–J2]			256			519
(e) Mismatch requires preceding break ((29))						
*[^+Strong,+J5][–Strong,+Accent,–J2]	0.06		29			39
*[^+Strong,+J4][–Strong,+Accent,–J2]	0.18		6			17
*[^+Strong,+J3][–Strong,+Accent,–J2]	**5.16**	0.001	0	**1.00**	0.046	3
*[^+Strong,+J5][–Strong,+Fall]	**0.78**	< .001	299	**1.18**	< .001	203
*[^+Strong,+J4][–Strong,+Fall]			190			142
*[^+Strong,+J3][–Strong,+Fall]			121			93
*[][–Strong,+Accent,–J2]			61			122
*[][–Strong,+Fall]	**0.40**	< .001	419	**0.31**	< .001	441

	Shakespeare			Milton		
	weight	**p-value**	**viols**	**weight**	**p-value**	**viols**
(f) Special strictness for phrase-final positions ((32))						
*[+Strong,–Accent,+Rise,–J5][+J5]	**1.42**	0.025	2			5
*[+Strong,–Accent,+Rise,–J4][+J4]			114			44
*[+Strong,–Accent,+Rise,–J3][+J3]			486			328
*[+Strong,+Rise,–J5][+J5]	0.30		17	**1.64**	< .001	10
*[+Strong,+Rise,–J4][+J4]			180			72
*[+Strong,+Rise,–J3][+J3]			506			342
(g) The Kiparskyan "mystery constraint" ((31))						
*[+Strong,+Accent,–J4][+Fall,–J3]	**1.98**	< .001	16	**1.47**	< .001	20
*[+Strong,+Accent,–J4][+Fall,–J2]			5			14
(h) The last foot is special ((35))						
*[–Rise][][+lb]	**1.22**	< .001	442	**1.97**	< .001	266
*[+Fall][][+lb]	**0.83**	< .001	20	**0.16**		21
(i) Line demarcation ((37))						
*[–J2][+lb]	2.52		0	3.46		0
*[–J3][+lb]	**2.50**	< .001	21	**2.27**	< .001	122
*[–J4][+lb]	**0.19**	0.033	261			1,442
*[–J5][+lb]	**2.47**	< .001	550	**0.56**	< .001	2,421
(j) Line cohesion ((39))						
*[+J4][]	**0.20**	< .001	5,905	**0.34**	< .001	7,362
*[+J5][]	**0.81**	< .001	1,613			3,905
(k) Foot cohesion ((40))						
*[–Strong,+J4]	**0.61**	< .001	2,208	**0.53**	< .001	2,746
*[–Strong,+J5]	**0.26**	< .001	428			1,331
(l) Foot cohesion — first and last foot ((41))						
*[+J5][][+lb]	**2.59**	< .001	8	**3.01**	< .001	16
*[+lb][+J5]	**0.57**	< .001	88	**2.57**	< .001	37
*[+J2][][+lb]			3,075			3,235

47. Overview

- With exceptions here and there, the constraints proposed by metrists are generally receiving positive weights and passing the significance test.
- The gaps are (mostly) gaps in *constraint families*, which have at least one representative.
- The failures are often in cases where the constraints are in a family, and only a subset of them emerge as significant.

48. Looking at individual lines

- We don't actually have probabilities (recall: sum weights time violations, reverse sign, take *e* to the result, find share of all candidates).
- But we do have sums of weights times violations, which suffice for comparison—the higher, the worse the line is predicted to be. This should suffice here.

Perfect (no violations):

Shakespeare

And make me travel forth without my cloak,
But not to tell of good or evil luck,
For slander's mark was ever yet the fair;

Milton

And short retirement urges sweet return.
But fondly overcome with female charm.
The high injunction not to taste that fruit,

Medium (weights x violations about 5.0):

Shakespeare

Making a couplement of proud compare
Look in your glass, and there appears a face
Whilst many nymphs that vow'd chaste life to keep

Milton

Had unbenighted shone, while the low Sun
So saying, her rash hand in evil hour
For this we may thank Adam; but his thanks

Extreme lines (weights x violations over 14.0)

Shakespeare

Whoe'er keeps me, let my heart be his guard;
Savage, extreme, rude, cruel, not to trust;
O! that you were your self; but, love you are [/ no longer yours]

Milton

Nor I on my part single, in mee all
Mee mee only just object of his ire.
Against God only, I against God and thee,

FATE OF OUR HYPOTHESES

49. "Some constraints will disappear as redundant"

- All versions of the Stress Maximum Constraint **bit the dust** (46c).
- Kiparsky's phrase-final constraints (32) **survive**: it is evidently worse to violate them than merely to both mismatch stress and end a phrase in mid foot (46f).

50. "Gradience and juncture continua"

- These did play out, with significance across the continuum of juncture, for a couple cases: line demarcation (46i), plus line cohesion and foot cohesion in Shakespeare (46j,k,l).
- However, no gradient effects was found for the metrical constraints: licensing of inversion (46e), nor phrase-final strictness (46f).
- We have doubts about whether we distinguished [3juncture] very well from [4juncture]; relying too much on a mechanical criterion. Perhaps had we simply used our juncture intuitions…

51. Shakespeare and Milton are different

- Milton shows considerably lower weights on the constraints regulating line-cohesion and line-demarcation (46i,j)—a sanity check, showing the model matches traditional wisdom in this area ("verse paragraphs").
- Phrase-final constraints: Kiparsky claims (33) that
  - Milton uses a constraint that must be applied no matter what stress level of first syllable)
  - Shakespeare uses a constraint that applies only if first syllable is stressless.
  - This is entirely confirmed by our modeling; see (46f).
- Kiparsky claims (see (30)) that Milton, unlike Shakespeare has a certain tolerance for falling lexical inversions. This is reflected in our system; see the big weight difference (5.16 to 1.00) in (46e).

DISCUSSION

52. Overall

- The casual traditional methods of generative metrists — read the corpus and keep track of what is missing — turned out to work fairly well when tested using our method.
- But we find it reassuring that the results stand up to testing—even with a corpus that is not all that long (2K lines per poet).
- Some hypotheses, however, find no support, particularly those that have alternative explanations under the Observed/Expected principle.

53. Further work

- We suspect we have not found all the relevant constraints and would like to keep exploring.
  - The capacity of the Hayes/Wilson system to do open-ended constraint search is worth trying out here.
- We would love to find a way to explain the Kiparskian Mystery Constraint (31), a thorn in our side that got stuck in deeper by coming out significant (46f).
- On the whole, we are *not* eager to apply the method right away to Shelley, Tennyson, Spenser, Frost, … it's too much work!
- But other metrical traditions—especially quantitative—might be more amenable to automated scansion and prosodic annotation, and might be easy to investigate rigorously following our method.
- We'd especially like to do a tradition where the poets are alive and willing to submit to queries about their well-formedness intuitions.

THIS COURSE: GENERAL REMARKS AND CONCLUSION

54. Maxent as a tool

- It doesn't do any thinking! That is up to us.
- But if our thoughts can come up with a constraint-based scheme for how a given domain works, the extreme accuracy and tacit "rationality" of maxent seems to be very helpful in cashing out our scheme as an explicit, testable theory.
  - Some of the patches added to their system by Albright and Hayes proved unnecessary under a maxent model of constraint interaction.
  - Maxent readily exposed the metrical constraints whose content was predictable from other well-motivated constraints.
- Maxent also doesn't cramp our style, since it's up to us as theorists to work out the constraints, or the device that generates them.

55. The need for a better account of bias

- We saw the need for two biases in learning.
  - naturalness (UG) bias (Hungarian study, Class 2)
  - generality bias (small scale exceptionless constraint need downgrading)
- As yet, we know of no principled way to bias.
- Moreover, the simple method of Wilson (2006)—Gaussian prior with μ set at 0—emerges as unreliable—other weights will squiggle around, eel-like, and undermine our purpose.

56. The need for overlearning

- We don't yet know of a single algorithm that both forces use of lexical entries and finds all the generalizations across the lexicon.
- Maxent cheats, slighting one or the other depending on sigmas.

57. Elsewhere

- Linguistics in all areas has gradience.
- Linguistics in all areas has tactics (static well-formedness) and choices, so each kind of application of maxent is in principle useful in these areas.

—FINIS—