# Class 2, 4/1/15:  Basics of Maxent and Iambic Pentameter

1. **Course bureaucracy**

- Read Russ's sketch of Hausa phonology
- Think about term paper projects — what data would you like to analyze?

MAXENT:  THE MATH AND THE BEHAVIOR

2. **What might we want a stochastic theory of constraint-based linguistics to do?**

- Output **probabilities** (if you like:  "degree of belief that this line is metrical")
- Distinguish between **stronger and weaker** constraints
- Allow **ganging**:  two bad things are worse than one of them alone[1]
- Allow **learning**:  exposure to data, given constraint set, leads to an accurate grammar.
  - ➢ Perhaps:  participants in a metrical tradition share their constraints, but adjust them to match what they hear in the tradition.

3. **Maxent, step one:  constraints and weights**

- The simplest conceivable way to give each constraint a strength is to give it a **number**, and that is what maxent does.
- The number called a **weight**.
- Weights are always non-negative and express a **penalty**.
  - ➢ Tricky to reward things!  If *TA has a negative weight then infinite [tatata…] beats all rivals.

4. **Maxent, step two:  Harmonic grammar**

- Every candidate gets a **harmony** score.
  - ➢ Badly named!  Harmony is a penalty.
- Simple **desiderata**:
  - ➢ More violations ought to be worse than fewer.
  - ➢ Violation of weightier constraints should be worse than violations of less-weighty ones.
- Simplest solution: **dot product** of violations and weights
  - ➢ i.e. multiply violations by weight for each constraints and then sum up

---

[1] How *much* worse?  This is a huge open question; maxent provides a very explicit answer but we don't know if it is the right answer.

> ➢ in Excel:  =SUMPRODUCT( weight-range, violation-range)
- Simple harmonic grammar:  lowest harmony candidate is the unique winner.[2]

## 5.  Next step

- For reasons given in Class 1, we need to make the theory probabilistic.

## 6.  Probabilities multiply

- If your penny has a probability of .5 coming up heads, and so does your nickel, then you get a probability of .25 that they will both come up heads on a particular occasion.

## 7.  Embodying the multiplicative law of probability with eHarmony[3]

- Scenario:
   > ➢ If you violate Constraint A then you can occur only 1/2 of the time.
   > ➢ If you violate Constraint B then you can occur only 1/3 of the time.
   > ➢ If you violate A and B you would occur 1/6 of the time.
- How to get this with harmony?

- Suppose that the "basic probability" = eHarmony = "probability all else being equal"

  eHarmony of a form = $1 / e^H = e^{-H}$

  where

   > ➢ H is the harmony of the form
   > ➢ *e* is the base of natural logarithms, about 2.718.

- Then simple math will get us what you want.

## 8.  Working it out

- Candidate 1 violates only A, so its harmony = $W_a$, the weight of constraint A.
   > ➢ eHarmony will be $e^{-Wa}$
- Candidate 2 violates only A, so its harmony = $W_b$, the weight of constraint A.
   > ➢ eHarmony will be $e^{-Wb}$
- Candidate 3 violate both A and B, so its harmony = $W_a + W_b$
   > ➢ eHarmony will be $e^{-Wa + -Wb}$
- But algebra tells us that this is
     $$e^{-Wa} * e^{-Wb}$$
  which multiplies the eHarmonies of Candidates 1 and 2.

---

[2] There is a variant:  make all the weights nonpositive and the highest harmony wins.

[3] This term, invented recently by Colin Wilson, has no official standing but I rather like it.

- We will see a real case shortly.

**9.  The last, slightly ugly part**

- Probabilities of all the candidates should sum to one.
- But this is not automatically true of eHarmonies themselves.
- So we **normalize** them
- Add up the eHarmonies of all candidates — call this Z
- Divide each candidate's eHarmony by Z to get probability.

**10.  Summing up:  the maxent procedure**

- For each candidate, compute Harmony as dot product of constraint weights and violations.
- Convert Harmony to eHarmony with formula $\mathbf{e^{-H}}$ (needed to get probability multiplication)
- Sum up eHarmony of all candidates to form Z.
- Probability of a candidate is the share of its eHarmony in Z.

**11.  Very simple case:  the taa meter**

- Assume every line is monosyllabic!

  *Ode.*
  *by.*
  *Bruce.*

  taa.
  taa.
  taa.
  ta.
  taa.

- It is preferred for that syllable to be heavy, so *taa* is a better line than *ta*.
- Look at spreadsheet.  Points that emerge:
  - ➢ Maxent is easy to do for small cases in Excel.
  - ➢ The Solver can be very convenient!  (Google "Excel Solver" to download).
  - ➢ We really do need to normalize — eHarmonies do not add up to one.

**12.  A more elaborate case:  the *Tadaa* meter (iambic monometer)**

- Every line is a two syllable iambic foot.
- The first syllable of the foot is preferentially light
- The second syllable is preferentially heavy.

Serene
Janine
Malone
Portrays
Sublime
Success.
Aha!

## 13. Some points to spot in the Tadaa simulation

- Multiplicative probability:  each constraint is violated 10% of the time, and violation of both happens 1% of the time.
- Again, eHarmonies do not add up to one.
- Not all frequencies can be fitted with this constraint set:  try frequencies 81, 9, 1, 9 and see what happens.
- Socrates:  what sort of frequencies would lead to the best weights being all zero (failed analysis)?

## 14. How are the best weights found?

- I set up the spreadsheet to handle the standard learning procedure for maxent.
- One cell (H8) holds the **objective function**:  make it as big as possible, and you will have the best analysis.

## 15. The standard objective function in maxent

- The basic principle is: *maximize the predicted probability of the data.*
- I.e. each datum has a probability assigned by the grammar; just multiply them all out!
- In a good grammar, the observed things should be assigned high probability, to the extent they are observed, and you should waste no probability on things unobserved.

## 16. Implementation of this

- Multiplying out a whole bunch of (often low) probabilities often produces a heroically low number, overwhelming the numerical capabilities of one's software.
- So we need a gimmick.
- Take the **log** of each probability — much more manageable!
- Whatever maximizes the predicted probability of the data also maximizes the predicted **log probability** of the data (since if X > Y, then log(x) > log(y))
- This is computationally very convenient
  - ➢ Instead of multiplying out probabilities to find the probability of the data (vastly small numbers),
  - ➢ we just add logs (quite tractable numbers)
- Excel has a log function, =LOG()

- John Goldsmith likes to abbreviate "log probability" as "**plog**", which practice I will follow.

## 17.  Return to the spreadsheet

- Look at calculation of plog
- Other than our not knowing how the Solver works, we've left the domain of magic.

## 18.  The last bit:  an aid to intuition

- Do you really believe in this "maximize plog" stuff?  You don't have to.
- We can also try **minimizing total squared error** instead.
- I've done this on the spreadsheet and it gets exactly the same outcome.

## 19.  Log odds

- The **odds** between two candidates is the ratio of their probabilities.
- Beautiful result:  if two candidates differ in just one constraint violation, the log of their odds will be the weight of that constraint — the most concrete answer to "what does the weight mean?"


IAMBIC PENTAMETER:  ORIENTAION

## 20.  One example:  Shakespeare's Sonnet 19

Devouring Time, blunt thou the lion's paws,
And make the earth devour her own sweet brood;
Pluck the keen teeth from the fierce tiger's jaws,
And burn the long-liv'd phoenix, in her blood;
Make glad and sorry seasons as thou fleets,
And do whate'er thou wilt, swift-footed Time,
To the wide world and all her fading sweets;
But I forbid thee one most heinous crime:
O! carve not with thy hours my love's fair brow,
Nor draw no lines there with thine antique pen;
Him in thy course untainted do allow
For beauty's pattern to succeeding men.
  Yet, do thy worst old Time: despite thy wrong,
  My love shall in my verse ever live young.

**21. Some lit stuff as preamble[4]**

- The earliest important English iambic pentameter poet was Geoffrey Chaucer, roughly end of 14th century—borrowing from Romance, particularly Italian models.
- Chaucer had the misfortune of writing just before the sound change *ə $\rightarrow \varnothing$ / ___ ]$_{word}$, which led to a period of centuries when his metrical intent was not understood.
- The iambic pentameter was then *re*discovered in the first half of the 1500's (Tudor poets: Wyatt, Surrey), and bloomed in the Elizabethan era.
- It had tremendous staying power, dominating English literature through the 19th century, and continues to be composed to this day.

**22. Meter**

- The iambic pentameter meter is often noted as

W S W S W S W S W S

where

S = strong rhythmic beat
W = weak rhythmic beat

- People often "recite the meter" using nonsense syllables like [də] and [ˈdʌː].
- We can consider S and W as shorthand for grid columns:

```
        x       x       x       x       x
    x   x   x   x   x   x   x   x   x   x
```

**23. Scansion**

- Scansion is the establishment of correspondence, not necessarily one-to-one, of the syllables of a line to the meter.
- We "scan" a line by notating this correspondence.

De vour ing Time, blunt thou the li on's paws,
```
  |  |   |   |    |    |   | | |    |
  W  S   W   S    W    S   W S W    S
```

- Occasionally there is more than one (discussed later) and one ordinarily provides the most probable one according to a good theory.

---

[4] This is me speaking as Joe Q. Public, reference sources needed.

**24. Communicating a scansion to other people**

- … is often done by using an artificial prosody that communicates the meter rather than the phonological structure—stress and lengthen the strong positions.
- Such a rendering is sometimes called the "**schoolboy**" scansion.

**25. Real-life readings follow a continuum:**

*"schoolboy"* ——————— *as if prose* ————— *metrically distorting*

TRADITIONAL LORE ON THE IAMBIC PENTAMETER METER

**26. Basics**

- An iambic pentameter consists of five iambic feet, each of which consists of a weak syllable followed by a strong.
- This is the general scheme, but it is adhered to only by approximation, with many unexpected weak syllables where you would expect strong, and vice versa.
- However, some deviations are so systematic as to be worth noting.

**27. Extrametrical syllables**

- It is common for the poet to allow an extra stressless syllable (how stressless?) at the end of a line.
- Sonnet 20 is unusual in exercising this option for all 14 lines:

A woman's face with nature's own hand painted,
Hast thou, the master mistress of my passion;
A woman's gentle heart, but not acquainted
With shifting change, as is false women's fashion:
An eye more bright than theirs, less false in rolling,
Gilding the object whereupon it gazeth;
A man in hue all 'hues' in his controlling,
Which steals men's eyes and women's souls amazeth.
And for a woman wert thou first created;
Till Nature, as she wrought thee, fell a-doting,
And by addition me of thee defeated,
By adding one thing to my purpose nothing.
  But since she prick'd thee out for women's pleasure,
  Mine be thy love and thy love's use their treasure.

**28. Extrametrical syllables in dramatic verse**

In the looser system Shakespeare used for dramas, extrametricals can occur provided they are

- after S

- at a strong "break", in a sense to be defined—intuitively, a place that deserves comma or stronger punctuation.
- Weakly stressed, in some sense that must be determined, and which varies from poet to poet.

[Find the extrametricals in these lines (Kiparsky 1977, 231-232):]

Quite overcanopied with luscious woodbine
That is the madman. The lover, all as frantic
Must curtsy at this censure. Oh, boys, this story
Oh, what a war of looks was then between them!
Sick-thoughted Venus makes amain unto him

[Make up a line in which an extrametrical occurs after W.]

### 29. Supernumerary syllables for phonological reasons

| | |
|---|---:|
| All these I better in one general best. | Sonnets, above |
| Which for memorial still with thee shall stay. | Sonnet 74 |
| Poor forlorn Proteus, passionate Proteus | (T. G. V. 1.2.124) |
| To emblaze the Honor that thy Master got | (2H6 4.10.76) |
| Take heed of perjury, thou art on thy deathbed | (Oth. 5.2.51) |
| Who? Silvia? / Aye, Silvia, for / your sake | (T. G. V. 4.2.25) |
| Weeds among weeds, or flowers with flowers gathered | (Son. 124) |

### 30. Paraphonology

- Kiparsky (1977) hypothesizes a **paraphonology**[5] (his "prosodic rules"), which can derive phonological representations used for poetry.

Formalizing, we can state the two most important paraphonological rules as follows:

**Prevocalic Desyllabification**

[+syllabic] $\rightarrow$ [–syllabic] / ___ [+syllabic]

Where the output segment is legal in English.[6]

**Posttonic Schwa Drop**

$\schwa \rightarrow \varnothing$ / [+syllabic] ___

---

[5] More often, he uses the term **prosodic rules**, which I feel is too ambiguous.

[6][6] Easy to do this if you're using OT.

**31. Another familiar paraphonological rule**

**Ambisyllabic /v/ Drop**

$$v \rightarrow \varnothing \ / \ V \underline{\quad} \begin{bmatrix} V \\ -stress \end{bmatrix}$$

Ordering:  feeds Posttonic Schwa Drop

And trouble deaf heaven with my bootless cries,
From sullen earth, sings hymns at heaven's gate,

This gives rise to still-familiar "poetry pronunciations":  o'er, e'er, e'en, se'ennight

**32. What is paraphonology?**

- Optional rules in the phonology of Shakespeare's day?
- The case for conventionalized paraphonology seems clearer for contemporary French, with schwa epenthesis, where the divergence in pronunciation between normal speech and singing/verse is very clear.

**33. Inversion**

- At the beginning of the line, it is common for the stress to be "inverted"; that is, we get a stressed syllable in the initial weak position, and a stressless in the following strong position.
- Less often, the same pattern is observed in the third or fourth foot, when preceded by a large break.
- This line has two inversions:

Richer than wealth, prouder than garments' cost,

[Find all of the inversions in this Sonnet.]

XCI

Some glory in their birth, some in their skill,
Some in their wealth, some in their body's force,
Some in their garments though new-fangled ill;
Some in their hawks and hounds, some in their horse;
And every humour hath his adjunct pleasure,
Wherein it finds a joy above the rest:
But these particulars are not my measure,
All these I better in one general best.
Thy love is better than high birth to me,
Richer than wealth, prouder than garments' cost,

Of more delight than hawks and horses be;
And having thee, of all men's pride I boast:
  Wretched in this alone, that thou mayst take
  All this away, and me most wretched make.

- Every once in a while, inversions can take place in any foot:

  Never, never, never, never, never     *King Lear*

**34. Summarizing the conventional wisdom theory**

- five iambic feet
- matching is good, but not great
- inversion at particular locations
- extrametricals at particular locations
- extra syllables can be trimmed back by phonological processes