But first—a report on the Tesar/Prince talk at UMass last week, “Learning phonotactic distributions”

(1) Pronunciation
['tesə']

(2) Subset problem
Child hears only data like [ta].
Which language should she learn: {[ta], [da]} or {[ta]}?

One suggestion for how to learn the subset language without negative evidence is an initial ranking of MARKEDNESS >> FAITHFULNESS: if there’s no evidence for demoting *VOICE, it stays high.

(3) When MARKEDNESS >> FAITHFULNESS is insufficient
Target language: {[ta], [na]}—we want to avoid learning {[ta], [na], [da]}

Initial state:
*VOICELESSNASAL, *VOICE, *NASAL >> IDENT[NASAL], IDENT[VOICE]

/ta/ → [ta], /na/ → [ta], /da/ → [ta]

If the child hears [na], she has two choices:

<table>
<thead>
<tr>
<th>/na/</th>
<th>*VOICELESSNASAL</th>
<th>*VOICE</th>
<th>*NASAL</th>
<th>IDENT[NASAL]</th>
<th>IDENT[VOICE]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[na]</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pa]</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td>!</td>
<td>!</td>
</tr>
<tr>
<td>[ta]</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td>!</td>
<td>!</td>
</tr>
<tr>
<td>[da]</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td>!</td>
<td>!</td>
</tr>
</tbody>
</table>

- In response to /na/ → *[ta], demote *NASAL and *VOICE below IDENT[NASAL]
  
  *VOICELESSNASAL >> IDENT[NASAL] >> *VOICE, *NASAL >> IDENT[VOICE]

/ta/ → [ta], /na/ → [na], /da/ → [ta]

Done.
• Or, in response to /na/ → *[ta], demote *NASAL and *VOICE below IDENT[VOICE]

  *VOICELESSNASAL >> IDENT[VOICE] >> *VOICE, *NASAL >> IDENT[NASAL]

  /ta/ → [ta], /na/ → [da], /da/ → [da]

Then, in response to /na/ → *[da], demote *NASAL below IDENT[NASAL]

  *VOICELESSNASAL >> IDENT[VOICE] >> *VOICE >> IDENT[NASAL] >> *NASAL

  /ta/ → [ta], /na/ → [na], /da/ → [da] superset language!

Done.

How do we get the learner to pick the demotion that will lead to a more restrictive grammar?

**(4) Biased constraint demotion**

Begin with no constraints ranked, and form strata from top to bottom.

**Faithfulness delay:** put off placing faithfulness constraints in a stratum as long as possible

Otherwise, place a constraint in a stratum as soon as possible (as soon as it does not prefer a loser, among the constraints left)

[na] example: only *VOICELESSNASAL prefers no losers, so place it in the top stratum

To rule out the remaining two losers, [ta] and [da], we need to place a faithfulness constraint. Which one?

**(5) r-measure (=restrictiveness measure)**

A ranking consistent with the data should be as restrictive as possible (i.e., not allow unobserved patterns like [da] to surface).

As a first approximation, this means that each faithfulness constraint should be dominated by as many markedness constraints as possible.

r-measure = sum of the number of markedness constraints dominating each faithfulness constraint.

\[
\text{r-measure}(*\text{VOICELESSNASAL} >> \text{IDENT[NASAL]} >> \text{VOICE,*NASAL} >> \text{IDENT[VOICE]}) = 4 \\
\text{r-measure}(*\text{VOICELESSNASAL} >> \text{IDENT[VOICE]} >> \text{VOICE} >> \text{IDENT[NASAL]} >> *\text{NASAL}) = 3
\]
“Freeing up” markedness constraints
Placing a constraint into a stratum frees up a not-yet-ranked markedness constraint if the markedness constraint could then be placed in a stratum.

If we place IDENT[NASAL] into a stratum...

*VOICELESSNASAL>>IDENT[NASAL]

it frees up *VOICE and *NASAL.

If we place IDENT[VOICE] into a stratum, however,...

*VOICELESSNASAL>>IDENT[VOICE]

it frees up only *VOICE.

So that’s how the learner gets the ranking

*VOICELESSNASAL>>IDENT[NASAL]>>*VOICE,*NASAL>>IDENT[VOICE]

Markedness cascades
Ranking a markedness constraints can in turn free up other markedness constraints. So two possible learning steps that free up the same number of markedness constraints immediately may not be equal down the line.

Even worse, freeing up one markedness constraint that initiates a cascade may be better than freeing up two that don’t.

How far ahead should the learner look? Prince & Tesar decide to look only as far as when the next faithfulness constraint would have to be ranked.

Further issues (read the paper on ROA)
- Positional faithfulness vs. positional markedness
- Stringency relations between constraints

See also Colin Wilson’s 2001 NELS talk.

Anttila: grammars as partial orderings

Orders review
- total ordering: irreflexive, asymmetric, transitive, total
- partial ordering: irreflexive, asymmetric, transitive, need not be total
- stratified ordering: irreflexive, asymmetric, transitive, need not be total; if \( \neg(a>b) \) and \( \neg(b>a) \), then \( (x>a)\iff(x>b) \) and \( (a>x)\iff(b>x) \)
Meaning of partial ordering for Anttila

Anttila proposes that grammars are partial orderings (usually stratified, but perhaps not necessarily).

To generate an utterance, the speaker must translate the partial ordering into a total ordering that doesn’t contradict any of the pairwise rankings of the partial ordering.

It’s not totally clear how this should work when the grammar is not perfectly stratified.

Should each total ordering be equally probable?
Or should A>>B be as probable as B>>A?

Example: Finnish genitives

Some stems always take the ‘strong’ genitive –iden, some always take the ‘weak’ genitive –en/-jen, and some vary, but often with a preference one way or the other.

- Monosyllables always take the strong variant (INITIALSTRESS, *STRESSED LIGHT)
  /maa/ mái.den *má.jen

- Disyllabic stems ending in a light syllable always take the weak variant
  (INITIALSTRESS, *STRESS CLASH, *UNSTRESSED HEAVY)
  /kala/ *ká.loi.den, *ká.lòi.den ká.lo.jen

- Disyllabic and longer stems ending in a heavy syllable always take the strong variant
  (IDENTWEIGHT?? These cases aren’t really discussed.)
  /palttoo/ pált.toi.den *pált.to.jen

- Trisyllabic and longer stems ending in a light syllable vary.
  - Those ending in a high vowel prefer the weak variant
    /lemmikki/ ~lé.mik.kèi.den lém.mik.ki.en
  - Those ending in a low vowel prefer the strong variant
    /sairaala/ sái.raa.lòi.den ~sái.raa.lo.jen

  - Those ending in a mid vowel vary more freely (secondary stress is optional:
    *LAPSE must be freely ranked w.r.t. some anti-stress constraint)
    /fyysikko/ fýy.si.kòi.den fýy.sik.ko.jen

Weird quirk: these generalizations refer to underlying vowel height

*HEAVYHIGH >> *HEAVY MID >> *HEAVY LOW
*LIGHT LOW >> *LIGHT MID >> *LIGHT HIGH
*STRESSED HIGH >> *STRESSED MID >> *STRESSED LOW
*UNSTRESSED LOW >> *UNSTRESSED MID >> *UNSTRESSED HIGH

(do we really need all four scales?)
• In trisyllabic and longer stems, there’s also a tendency for a heavy antepenult to take the weak genitive and for a light antepenult to take the strong genitive.

“weight-clash” constraints: *H.H, *L.L

Weight-clash considerations conflict with vowel-height considerations (corpus data reported by Anttila, for 3-, 4- and 5-syllable words combined):

![Graph showing vowel height considerations for different types of stems]

Some categorical gaps in longer words:

/mí.nis.te.rèi.den mí.nis.te.ri.en

/már.ga.rii.nei.den már.ga.rii.ni.en

/már.ga.rii.nèi.den

/á.lek.sàn.te.rèi.den á.lek.sàn.te.ri.en

/kòor.di.nàa.tis.to.jen

/í.ta.li.àa.noi.den í.ta.li.àa.ni.en

Statistical issue: Anttila takes unattested forms (such as L.O.weak in 4- and 5-syllable words) as prohibited rather than as rare variants, but 0 may not be significantly different from 2 or 4...
(12) Proposed grammar
Nearly stratal, but not quite

(plus transitivity)

Thus variation in ranking is within each row, and among all the constraints of the last three rows, except that *LIGHTO>>*LIGHTI and *UNSTRESSEDA>>*UNSTRESSEDO>>*UNSTRESSEDI.

I’m not sure if Anttila really intends all of the indicated pairwise rankings. Does *LIGHTA really have to outrank *STRESSEDA, for instance?

(13) Predicted variation patterns—one example

/L CI/ possible variants: LÍ L I
competing constraints are *HEAVYI, *STRESSED I, *L L

Since these three constraints are freely rankable, we expect to see 1/3 LÍ (actually 37%), 2/3 L CI (actually 63%).

In general, the corpus numbers are a very good match (see table on p. 23).

(14) Grain
How fine-grained the predictions can be depends on the number of conflicting, freely ranked constraints in a given case.

There are never more than 5 in Anttila’s analysis, so the smallest percentage greater than 0 that can be predicted is 20%.
(15) *Course of learning*
In Anttila’s model, learning should proceed from greater variation to more variation.

If the initial state has two big strata, one for markedness and one for faithfulness, and pairwise rankings are instantiated as necessary, then the progression is from absence to presence of marked structures.

Note that the Finnish case is almost entirely about conflicting markedness constraints, so even with such an initial state, the progression will be mainly from more to less variation.

**Boersma/Hayes/MacEachern stochastic constraint ranking**

(16) *Review: stochastic constraints*
(Boersma 1998,¹ Hayes & MacEachern 1998²)
For Boersma, the ranking of constraints is neither fixed nor freely variable, but probabilistic.

- Each constraint in an individual’s grammar has a *ranking value*, given in arbitrary units.
- In each utterance, the speaker generates *selection points* (“disharmony” for Boersma) for each constraint
  
  \[ selectionPoint = rankingValue + rankingSpreading \times z \]

- *rankingSpreading* (“noise” in OTSoft) is some constant, typically 2
- \( z \) is a Gaussian random variable with mean 0 and standard deviation 1.
  Gaussian random variables are what give us “bell curves” (normal distributions):

---

Example of selection points for 10 runs of 2-constraint grammar (Boersma p. 331)

<table>
<thead>
<tr>
<th>trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>50.5</td>
<td>51.2</td>
<td>50.2</td>
<td>49.1</td>
<td>52.9</td>
<td>52.9</td>
<td>52.7</td>
<td>53.8</td>
<td>55.4</td>
<td>54.3</td>
</tr>
<tr>
<td>(ranking value = 50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>50.8</td>
<td>48.3</td>
<td>50.7</td>
<td>51.2</td>
<td>48.9</td>
<td>48.8</td>
<td>48.2</td>
<td>50.3</td>
<td>48.1</td>
<td>48.7</td>
</tr>
<tr>
<td>(ranking value = 50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Constraints are ranked, for the utterance in question, according to their selection points. We can call this the “effective ranking”.

(17) We can think of each constraint as associated with a probability density function centered on the ranking value:

(18) Gradient well-formedness judgments
For Boersma and Hayes, a pronunciation’s well-formedness is a function of its probability of being produced by the grammar.

Actually, they propose a function for the difference between two well-formedness judgments.
This is because, although frequencies of variants must sum to 1, the sum of the well-formedness judgments of variants is not constant (they might all sound pretty bad or good because of something else about the word).

The difference in well-formedness between two pronunciations A and B, on a seven-point scale, is predicted to be

$$\text{wf}(A) - \text{wf}(B) = \log \left( \frac{P(A) + P(B)}{P(A)} - 1 \right)$$

\[ \log 0.2 \]
Why not just $7(P(A) - P(B))$?

The log function means that you need a very big difference in frequency to get a big difference in well-formedness (“it would be a rash learner that concluded that there is anything seriously wrong with a form that appears in a sixth of all cases”—Boersma & Hayes, appendix B).

Say $A$ occurs 5/6 of the time, and $B$ occurs 1/6 of the time.

$7*(P(A) - P(B)) = 4.67$ (big difference)

$\log(1/P(A) - 1) / \log(0.2) = 1$ (not-so-big difference)

(19) **Grain**

As fine as you want—doesn’t depend on the number of constraints

The probability of CONSTRAINT1 >> CONSTRAINT2 can never be 0% or 100%, but it can be arbitrarily close to 0% or 100%.

(20) **Errors**

Is there a difference between an error and a very rare production in the stochastic model?

If errors are thought to originate outside the grammar (improper assembly or choice of underlying form, intrusion of gesture in the motor component, etc.), then errors do have a special status for the speaker.

But how does the learner detect that? This is a problem in just about any model, except one in which very rare variants can’t be generated by the grammar—in that case, perhaps the learner can infer that anything rare is an error.

If the learner uses the GLA, detecting errors isn’t that important, because they won’t have a big effect on learning anyway.

(21) **Free vs. lexical variation**

Both Anttila and Boersma/Hayes/MacEachern model free variation—any input can be pronounced several ways.

But what about lexical variation, where some words behave one way and others behave another way, and despite some consistent patterns, absolute phonological predictability is lacking?

Here’s a grammar I proposed for just such a case, Tagalog nasal substitution (each word is unpredictable, but stems with a voiceless initial consonant undergo it more, as do stems with a fronter initial consonant):
The top two constraints force faithful use of lexical information. The variably ranked constraints on the bottom (whose rankings is learned by the GLA) determine the pronunciation of newly coined words, and their acceptability ratings.

Such a grammar could also be expressed in Anttila’s model, except that ranking preference among the lower constraints wouldn’t be expressible.

(22) Floating constraints
the Boersma/Hayes model can’t capture grammars like A>>B, with C unranked (C>>A>>B, A>>C>>B, A>>B>>C equally probable).

Anttila doesn’t have any examples like this, but he doesn’t seem to rule them out.

A
B
C

‘Floating’ constraints like this were proposed by Nagy & Reynolds, Reynolds, and my MA thesis.

Boersma has some reanalyses of the Nagy/Reynolds cases (too terse to understand without having seen the original papers) at http://www.fon.hum.uva.nl/paul/gla/.

The cases from my MA thesis are easy to reanalyze. They’re all of the form IDENT-IO>>IDENT-BR, with a floating phonotactic constraint.

\[
\begin{align*}
*\theta >> \text{IDENT-IO, IDENT-BR} \\
\text{IDENT-BR} >> *\theta >> \text{IDENT-IO} & /\text{mag+RED+θeŋkju}/ \rightarrow [\text{mag-te-θeŋkju}] \\
\text{IDENT-IO} >> *\theta >> \text{IDENT-BR} & /\text{mag+RED+θeŋkju}/ \rightarrow [\text{mag-te-θeŋkju}] \\
\text{Ident-IO, Ident-BR} >> *\theta /\text{mag+RED+θeŋkju}/ \rightarrow [\text{mag-θe-θeŋkju}]
\end{align*}
\]

p. 10
So actually, any ranking yields an attested form.

If all three constraints are variably ranked (Anttila) or have the same ranking value (Boersma & Hayes), we expect to see about 50\% [mag-te-θενκ̄ju], 17\% [mag-te-θενκ̄ju], and 33\% [mag-θε-θενκ̄ju].

If the phonotactic truly floats freely (Anttila-style), the three should be equally probable.

If the floating constraint is Gaussian but has fatter tails than the ‘fixed’ constraints, [mag-te-θενκ̄ju] should be the most probable.

I don’t have good data to get frequencies from, but a quick web search suggests IDENT-IO >> PHONO >> IDENT-BR is more frequent for PHONO = *COMPLEXONS.

from www.google.com-- includes prefixes (nag-, mag-, pag-) and variant spellings

<table>
<thead>
<tr>
<th>P &gt;&gt; IDENT-IO</th>
<th>%</th>
<th>IDENT-IO &gt;&gt; P &gt;&gt; IDENT-BR</th>
<th>IDENT-IO, IDENT-BR &gt;&gt; P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ga-garaduate</td>
<td>1</td>
<td>ga-graduate</td>
<td>17 gra-graduate</td>
</tr>
<tr>
<td>ta-tarabaho</td>
<td>3</td>
<td>ta-trabaho</td>
<td>791 tra-trabaho</td>
</tr>
<tr>
<td>pa-paraktis</td>
<td>2</td>
<td>pa-praktis</td>
<td>21 pra-praktis</td>
</tr>
<tr>
<td>pi-priprito</td>
<td>1</td>
<td>pi-priprito</td>
<td>6 pri-priprito</td>
</tr>
<tr>
<td>to-toroso</td>
<td>0</td>
<td>to-toroso</td>
<td>18 tro-toroso</td>
</tr>
</tbody>
</table>

This is consistent with PHONO having a ranking value somewhere in between IDENT-IO’s and IDENT-BR’s in a Boersmian grammar.
I don’t think it’s capturable in an Anttilan grammar, unless with different constraints.

Next time: Computability of OT and variants

For next time
- Reading: Tesar & Smolensky ch. 7
- Assignment: think of variants of OT whose computational implications you’d like to discuss and let me know. If you’re enrolled for credit, decide when you’d like to lead discussion of a paper.