Class 2, 1/12/2023: Faithfulness, Factorial Typology, the Rich Base

1. Practical

- Can everyone access the course web site? I believe it is posted ok now.

2. Assignments

- Read: “How to solve phonology problems.” On web site.
- Start the first homework, on Ilokano. Download pdf and Word doc from website. It is due in a week, 1/19/23.

3. Finnish as a canonical OT-case

- Hopefully, complexity from simplicity via ranking

4. Algorithms (implemented in software) can help

- Many people have programmed Recursive Constraint Demotion (enacted last time), and it can check whether your constraints can select the right winner from your candidates.
- It also finds strata that define a set of feasible rankings.
- In some implementations, you can rerun the system eliminating each constraint in turn, testing for necessity.
- I went to my OTSoft program (https://linguistics.ucla.edu/people/hayes/otsoft/) and got some reassurance.

```
Stratum #1
  Align(W,Ft) Align(W,Ft)
  *Clash *Clash
  FootBin FootBin
Stratum #2
  Non-Fin Non-Fin
  Dep Dep
Stratum #3
  Lapse Lapse
Stratum #4
  *(L H) *(L H)
Stratum #5
  All feet L All feet L
```
• Note that this defines 12 rankings that work …

• Tableaux:

/L H L L L L/:  

\[
\begin{array}{cccc|ccc}
\text{Align}(W,Ft) & *\text{Clash} & \text{FootBin} & \text{Non-Fin} & \text{Dep} & \text{Lapse} & *(L H) & \text{All feet L} \\
(L H) (L L) L & | & | & | & 1 & 1 & 2 \\
(L H) L (L L) & | & | & | & 1 & 1 & 3! \\
(L H) L L L & | & | & | & 4! & 1 & \\
(L H) (L L) (L) & | & 1! & 1 & | & 1 & 6 \\
(L) (H L) (L L) & 1! & 1 & | & | & | & 4 \\
L (H L) (L L) & 1! & | & | & | & | & 4 \\
\end{array}
\]

etc.

• Necessity:

Status of Proposed Constraints: Necessary or Unnecessary

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Align(W,Ft)</td>
<td>Necessary</td>
</tr>
<tr>
<td>*Clash</td>
<td>Necessary</td>
</tr>
<tr>
<td>Non-Fin</td>
<td>Necessary</td>
</tr>
<tr>
<td>Lapse</td>
<td>Necessary</td>
</tr>
<tr>
<td>*(L H)</td>
<td>Necessary</td>
</tr>
<tr>
<td>All feet L</td>
<td>Necessary</td>
</tr>
<tr>
<td>FootBin</td>
<td>Necessary</td>
</tr>
<tr>
<td>Dep</td>
<td>Necessary</td>
</tr>
</tbody>
</table>

5. Another really helpful algorithm: FreD (Prince and Brasoveanu 2011)

• This finds all the ranking arguments and makes a Hasse diagram.
• This turns out to be a big computational task! For a large problem set an implemented version can run for half an hour.
• It seems that for nontrivial cases human efforts should not be trusted.

• OTSoft output for Finnish:

FootBin $\gg$ { *(L H), All feet L, Dep }
Dep $\gg$ { *(L H), All feet L }
*(L H) $\gg$ All feet L
Lapse $\gg$ { *(L H), All feet L }
Non-Fin $\gg$ { Lapse, *(L H), All feet L }
*Clash $\gg$ { Lapse, *(L H), All feet L }
Align(W,Ft) $\gg$ { Non-Fin, Lapse, *(L H), All feet L, Dep }
• We were curious how an alternative outcome could be derived:

\[ ['ravinˌtolat] \text{ is winner} \]

\[ ['ravintoˌlat] \text{ is winner} \]

• You can try comparing these Hasse diagrams (or wait for us to do the same thing with a MaxEnt grammar).

QUICK REVIEW OF FAITHFULNESS CONSTRAINTS

6. Historical note

• In the original OT work, Prince and Smolensky (1993) proposed a system of Faithfulness not well-articulated enough to handle phonology in general
• You will see references to it from time to time but it is not widely employed today.

7. Source for the standard constraints used today


8. In principle, there are many ways to do Faithfulness

• Faithfulness is based on resemblance.
• Resemblance can be formalized in many ways.
• Conventional SPE-representations (sequences of feature matrices) make it easy, but one might also pull in
  ➢ hierarchical structure (phrasing, syllables, feet, autosegmental tiers)
  ➢ actual phonetic form (spectra)
• More on this later.

9. McCarthy and Prince’s strategy: SPE representations + atomism

• I believe their idea was to find the ways that two representations could differ — allocating constraints to the smallest possible differences.
10. Indices

To make the differences explicit, we put an index on every segment (we will see shortly why this is necessary).

- **IDENT** = differ in one feature value
  
  /p₁ a₂ k₃/, candidate [b₁ a₂ k₃] violates IDENT(voice)
  /p₁ a₂ k₃/, candidate [m₁ a₂ k₃] violates IDENT(voice) and other constraints

- **MAX** = an underlying segment of some natural class (specified with features) is missing in the surface form.
  
  /p₁ a₂ k₃/, candidate [p₁ a₂ ] violates ?? (multiple answers)

- **DEP** = a surface segment of some natural class (specified with features) is missing in the underlying form.
  
  /a₂ k₃/, candidate [?₁ a₂ k₂] violates multiple DEP constraints.

- **LINEARITY**, violated when the linear order of any pair of segments is switched. Count the violations here:
  
  /p₁ a₂ k₃/, candidate [p₁ k₃ a₂] has one violation
  /p₁ a₂ k₃/, candidate [k₃ a₂ p₁] (how many violations?)

  Socrates: What about this candidate: [p₃ a₂ k₁]

- **CONTIGUITY**, violated when two segments are adjacent in the input but not the output, or vice versa (IO, OI)

- **ANCHOR**, violated when a segment is adjacent to an edge in the input but not the output.

11. Not a standard Faithfulness constraint

- **IDENT(p)** “Don’t change anything about [p] so it isn’t [p] any more.”
  
  ➢ You’re welcome to deviate from the “standard” theory — but in the interest of scientific coherence, you need to label such deviations.
  ➢ This raises issues of what is your “scientific community”, your “community standards”—inevitable in writing up research results.
12. Extensions of Faithfulness

- Between a base form and a form derived from it (Benua 1997\(^1\) and much later work)

  \(\text{heal} \quad [\text{hi}\tilde{l}] \quad \text{healing} \quad [\text{hi}\tilde{\text{i}}\text{n}]\)  
  compare: \(\text{Darjeeling} \quad [\text{da}r\text{'d}j\tilde{\text{e}}\text{l}\text{n}]\)

  You might try this for the homework.

- Between a reduplicated morpheme and its base (McCarthy and Prince 1994)\(^2\)

DEFINING FACTORIAL TYPOLOGY

13. Assume

- A class of universal inputs
- Some version of GEN
- A set of constraints

then, each ranking of the constraints defines a set of outputs
(many rankings define the same output set, however)

The set of distinct sets of outputs constitutes the **factorial typology** of the system
[called “factorial” because \(n\) constraints permit \(n!\) rankings]

14. The appeal of factorial typology

- Proposals can be tested against typological data.
- We can consider them for defects of both undergeneration (instantly fatal) and overgeneration (hard to assess)

15. Some work in factorial typology


---


16. How to compute a factorial typology

- Chose an empirical domain (we can’t do all of phonology at once).
- Choose constraints: these embody a proposal in phonological theory.
- Choose inputs: they must somehow suffice to illustrate the full range of possible phenomena.
- Choose candidates (ditto)
- Calculate the typology, by hand or with software.

17. Language example: Turkish epenthesis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>İdil</td>
<td>idili</td>
<td>idilim</td>
<td>idile</td>
<td>idiller</td>
<td>idilden</td>
<td>‘idyll’</td>
</tr>
<tr>
<td>vezin</td>
<td>vezni</td>
<td>veznim</td>
<td>vezne</td>
<td>vezinler</td>
<td>vezinden</td>
<td>‘meter’</td>
</tr>
<tr>
<td>tʃenk</td>
<td>tʃenki</td>
<td>tʃenkim</td>
<td>tʃenke</td>
<td>tʃenkler</td>
<td>tʃenkten</td>
<td>‘hand’</td>
</tr>
</tbody>
</table>

- “Epenthesize before a consonant that wouldn’t be syllabifiable.”
- The idea is that [idil] is /idl/.
- It has to be Epenthesis, not Syncope
  - alternating vowel is always /i/ (or its partners in vowel harmony)
  - /vezin/ doesn’t alternate.
- /tʃenk/ has good sonority and makes a good syllable in Turkish

18. Egyptian Arabic

- This language also has /vezn/ for ‘meter’, but in this language the epenthetic vowel is final: [vezni]

19. Palestinian Arabic

- Like Turkish, but all CC sequences are repaired, not just sonority-violating ones:
  /tʃxt/ → [tʃxt] ‘bed’

20. Sample inputs for factorial typology

- /ipl/ Exemplifies a bad-sonority coda
- /ilp/ Exemplifies a coda with good sonority profile

21. Reasonable outputs

/ipl/ i.pl /ilp/ i.lip
i.pi i.l.pi
i.pl i.pi.i
i.pi.i
• Do we have them all?
• If not, what would we include to expand the typology?

22. **Should we include [i.lpi] among the outputs?**

• Proposal: to keep the problem under control, you can assume some constraints to be undominated (i.e. “we are considering only the class of languages that don’t allow sonority-reversed codas”).
• Often, this leaves enough languages under consideration to make the problem still worthwhile.
• Thus, let’s avoid, for /ilp/: i.lpi, il, ip, lip, pli, ilu (p becomes u).
  ➢ Undominated: *HORRIBLE ONSET, MAX(C), LINEARITY, IDENT(syl)

23. **Proposed constraints**

<table>
<thead>
<tr>
<th>Name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>DEP(i)</strong></td>
<td>Penalizes insertion of vowels; quality ignored here.</td>
</tr>
<tr>
<td>2. <strong>BAD SONORITY CODA</strong></td>
<td>Should be suitably formalized; violated by final [pl].</td>
</tr>
<tr>
<td>3. <strong>CODA CODA</strong></td>
<td>Violated by [ilp] as well.</td>
</tr>
<tr>
<td>4. <strong>CODA</strong></td>
<td>Inviolable in Hawaiian, Bantu languages, etc.</td>
</tr>
<tr>
<td>5. <strong>BRANCHING ONSET</strong></td>
<td>i.e. *[σ CC</td>
</tr>
<tr>
<td>6. <strong>CONTIGUITY</strong></td>
<td>one violation for each pair of segments adjacent input but not output — penalizes internal epenthesis</td>
</tr>
<tr>
<td>7. <strong>ANCHOR</strong></td>
<td>one violation for each segment adjacent to a particular word edge in the input but not in the output — penalizes external epenthesis</td>
</tr>
</tbody>
</table>

24. **Some outputs we need never consider**

/ipl/ → [ip.il]

These bad outputs are **harmonically bounded** and (in classical OT) can never win.
[ Socrates: take a look at the tableau rows. ]

25. **Harmonic bounding**

Candidate A **harmonically bounds** candidate B if A has a strict subset of B’s violations.

26. **A software-derived factorial typology of epenthesis**

Method employed: keep adding inputs, run Constraint Demotion over and over, regain all combos that have a feasible ranking.
<table>
<thead>
<tr>
<th>Output #1</th>
<th>Output #2</th>
<th>Output #3</th>
<th>Output #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ipl/:</td>
<td>[ipl]</td>
<td>[i.pl]</td>
<td>[ip.li]</td>
</tr>
<tr>
<td>/ilp/</td>
<td>[ilp]</td>
<td>[ilp]</td>
<td>[ilp]</td>
</tr>
<tr>
<td>faithful outcomes; Persian (\textit{va}zn, \textit{t}æxt)</td>
<td>Turkish</td>
<td>Palestinian Arabic</td>
<td>I know no case; = Egyptian-Turkish blend</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output #5</th>
<th>Output #6</th>
<th>Output #7</th>
<th>Output #8</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ipl/</td>
<td>[i.pli]</td>
<td>[i.pli]</td>
<td>[i.pli]</td>
</tr>
<tr>
<td>/ilp/</td>
<td>[i.pli]</td>
<td>[i.pli]</td>
<td>[i.pli]</td>
</tr>
<tr>
<td>Egyptian Arabic</td>
<td>French, at least in some speaking styles</td>
<td>I know no case; = Egyptian-French blend</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output #9</th>
<th>Output #10</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ipl/</td>
<td>[i.pli]</td>
</tr>
<tr>
<td>/ilp/</td>
<td>[i.pli]</td>
</tr>
<tr>
<td>Japanese loan adaptation: \textit{tax} \rightarrow \textit{[takusu]}</td>
<td></td>
</tr>
</tbody>
</table>

27. Basics of Factorial Typology interpretation

- **Undergeneration**: real case that the constraint system can’t get under any ranking; indubitably bad
- **Overgeneration**: hypothetical case that the linguist feels could never occur. How bad?

- Linguists differ greatly in how seriously to take the argument “gosh, I’ve never seen a case of this kind”. A forceful statement of skepticism of this kind of thinking may be found in:
  - Perhaps some day we will use some form of statistical analysis to authenticate our gaps.

28. Sometimes the missing case walks in the door

- **Majority-rule vowel harmony**: “Harmonize a suffix in backness with which ever kind of vowel (back or front) occurs more often in the stem.” (Linda Lombardi; for a remedy see Baković 1999.)

\[\text{Unless Margit is right…}\]


• But Margit Bowler’s work suggests this happens in Warlpiri.
• I believe that “cases walking in the door” are fairly frequent — so many languages, so little analysis.

29. Typological work

• You can mine libraries (Noah Elkins, 2021 UCLA MA thesis; my 1995 book *Metrical Stress Theory*
• UCLA has a fantastic descriptive collection (stroll one minute north to YRL, fourth floor, roughly PK-PL).
• Interlibrary Loan has gotten more efficient (on my desk, a book from Harvard)
• And the internet has more resources

30. Interpretation of factorial typology: T-orders

• We examine the factorial typology, and compile all true statement of the following form:

  In all grammars in which Output A\textsubscript{m} is derived from Input A, Input B must yield Output B\textsubscript{n}.

• Arto Anttila, in various papers\textsuperscript{6} emphasizes this kind of implication (i.e. a complete set of them) as a way of diagnosing the structure of a factorial typology.
• For a brief intro see http://www.stanford.edu/~anttila/research/torders/t-order-manual.pdf

31. The T-order for the epenthesis typology (complete)

<table>
<thead>
<tr>
<th>If this input</th>
<th>has this output</th>
<th>then this input</th>
<th>has this output</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ipl/</td>
<td>[ipl]</td>
<td>/ilp/</td>
<td>[ilp]</td>
</tr>
<tr>
<td>/ipl/</td>
<td>[i.pi.li]</td>
<td>/ilp/</td>
<td>[i.li.pi]</td>
</tr>
</tbody>
</table>

THE RICH BASE, PHONOTACTICS AND ALLOPHONY

32. Phonotactics

• = phonological legality, grammaticality
• Chomsky and Halle (1965, *Journal of Linguistics*):
  
  \[
  \begin{array}{l}
  \text{[b\text{\textsuperscript{\textregistered}k}] is possible and exists} \\
  \text{[bl\text{\textsuperscript{\textregistered}k}] is possible and doesn’t exist} \\
  *\text{[bn\text{\textsuperscript{\textregistered}k}] is impossible}
  \end{array}
  \]

• Phonotactic well-formedness is gradient (’?[p\text{\textsuperscript{\textregistered}k}]), but (since we are working with classical OT) we will idealize for the moment to a grammatical/ungrammatical distinction.

\textsuperscript{6} e.g.
• Phonotactics is learned in the absence of negative evidence: from what we do hear, we figure out what would should never expect to hear.

33. Why do we have phonotactic knowledge?

• One theory is that it helps guide speech perception: we prefer phonotactically well-formed interpretations of the waveform.
• A famous paper:
  ➢ A synthesized [r]-[l] F3 continuum is biased to be heard as [r] after [t], as [l] after [s].

34. The standard OT account of phonotactics: the Rich Base (Prince and Smolensky 1993)

• Assume that the set of possible inputs is *every conceivable phonological representation*.
• An adequate grammar converts any unpronounceable input into a pronounceable one — thus expressing the phonotactics.

35. Pseudo-derivations in Rich Base theory

• For bad things: Markedness outranks at least one Faithfulness constraint whose violation could repair an input that is bad.
• This assumes fictional “derivations”, like English /qæt/ → [kæt]
  ➢ (BH recite the Parable of the Space Aliens.)
• Such grammar is often (harmlessly, I think) indeterminate — we don’t know what repair “would be” used.\(^8\)
• Socrates: what other Faithfulness constraints could be violated in repairing /qæt/?

36. The fundamental principle for phonotactics in Rich Base theory

• The higher-ranked the Faithfulness constraints, the more things emerge from the grammar.
  ➢ Intuition: without Faithfulness, we could say only [ta] (or perhaps nothing).
• This basic point can be elaborated into a typology of phonotactics, based on ranking
  ➢ illegality
  ➢ phonemicness
  ➢ allophony
  ➢ contextually-limited contrast (two ways to do it)

\(^7\) They attribute the idea to David Stampe, who thought of it for his theory of Natural Phonology, a rule-based precursor to OT (universal rule set!).

\(^8\) Though we can make an educated guess (work of Steriade, later): default is repair to the phonetically closest legal entity.
WORKING THROUGH RICHNESS OF THE BASE WITH MINI-TABLEAUX

37. Case I: Illegality
   - As above, for /qæt/

38. Case II: Phonemicness
   - A feature will be phonemic if the faithfulness constraint for it dominates the conflicting markedness constraint.

\[
\text{IDENT([voice])} \quad \text{(Faithfulness)} \\
*[-\text{sonorant}, +\text{voice}] \quad \text{(Markedness; cf. Hawaiian)}
\]

39. Case III: Allophony
   - This is: two conflicting Markedness constraints, special one with context, general one ranked second, both above Faithfulness.

Special markedness \(\downarrow\)
General Markedness \(\downarrow\)
Faithfulness

\[
\text{VTV} \quad *[-\text{son}, -\text{voice}] / V \quad \_ V \quad \text{(cf. Korean, Yidi, various Australian)} \\
*D \quad *[-\text{sonorant}, +\text{voice}] \quad \text{(Hawaiian)} \\
\text{IDENT IDENT([voice]) in obstruents}
\]

   - Socrates: How does this ranking distribute [t] and [d]?

40. Case IV: contextually limited contrast by the Markedness Sandwich
   - Assume:

\[
\text{IDENT([voice] in obstruents / \_ [+sonorant])} \\
*[-\text{sonorant}, +\text{voice}] \\
\text{IDENT([voice] in obstruents)}
\]

   - Describe the resulting distribution of [t] and [d], justifying your answer with little tableaux.
   - General recipe: \text{SPECIAL FAITHFULNESS} >> \text{MARKEDNESS} >> \text{GENERAL FAITHFULNESS}

41. Socrates
   - What is the factorial typology of the three constraints we’re working with?
42. Case V: contextually limited contrast by the Faithfulness Sandwich

- Here, what’s different is that we put the context into the Markedness constraint, not Faithfulness.
- Assume:
  
  \[
  \begin{align*}
  &*[\text{-}\text{sonorant}, +\text{voice}] \text{ in codas} & \text{special markedness} \\
  &\text{IDENT(}[\text{VOICE}] & \text{Faithfulness} \\
  &*[\text{-}\text{sonorant}, +\text{voice}] & \text{general markedness}
  \end{align*}
  \]

- Describe the resulting distribution of [t] and [d], justifying your answer with little tableaux.

RICHNESS OF THE BASE, ALLOPHONES, AND CONCRETE URS

43. Often, many URs are compatible with a given SR

Socrates: suggest some URs for [kʰ̆ẗ] ‘cat’

44. One way to resolve: Lexicon Optimization (hypothesis of Prince and Smolensky 1993)

- All else being equal, learned morphemes are represented in the lexicon in whatever way minimizes their constraint violations (necessarily, of Faithfulness).
- The “all else being equal” is a pretty broad loophole…
- [ Socrates: what would lead you to list something different from phonetic form? ]

45. Good old allophones

- Beginners to this day are taught to collect sounds that are phonetically similar and in complementary distribution (e.g. for /t/: [tʰ], [t], [t̚], [t]).
  
  Socrates: dredge up your memory of where these allophones of /t/ are distributed.
- Standard OT provides no basis for this actually happening. Why?
  
  Old theory: constraints on underlying forms included phoneme inventory.
  
  But if the constraints all apply at the surface there is no such thing, only a phone inventory.
- Nonstandard OT? Perhaps we relax Lexicon Optimization, permitting underlying-surface disparities to minimize the number of categories in UR.

46. Should phonemes be unified?

- For experiments suggesting that native speakers sometimes hear allophones as “the same sound”, see
  
  
Likewise, experiments teaching people how to read: what letter do they pick for an allophone untaught to them? Gudschinsky, Popovich, and Popovich (1970)\textsuperscript{9}, who got a untutored speaker to spell \textla{jk} as \textla{kk}, \textla{[i]} being the contextual allophone of /kl/.

Anecdote: my son, when little, trained to string-reverse words by sound: \textit{cat} \textla{kʰæt} → \textla{tʰæk}, \textit{fill} \textla{fɪl} → \textla{lf}.

\section*{47. Can phonemes be unified under standard OT?}

Some may be unified at a pre-symbolic, categorization level—creation of categories by the warping of perceptual space. References:

\begin{itemize}
  \item UCLA dissertations of Ying Lin, Kristine Yu, on our web site
\end{itemize}

More interesting are allophones that are not in a cloud, but discrete, e.g. \textla{r} and \textla{χ} as allophones (putative) of “/r/” in Portuguese.

OT opts for “invariance”:\textsuperscript{10}

\begin{itemize}
  \item For every phoneme, there is a set of feature values shared by all allophones of that phoneme and by none of allophones of any other phoneme.
\end{itemize}

To my knowledge, no one has published anything on invariance-violating phonemes in OT …
