Class 9, 2/7/23: Type Variation II; Phonotactic Analysis

1. Current assignments
   - BH hand back Hebrew homework
     - On web site
     - No summary required
   - New homework on Phonotactics is due Thurs. Feb. 16.

2. Today
   - Go over Hebrew homework
   - Finish type variation
   - Orientation on phonotactics
   - Orientation to the phonotactics homework

   TYPE VARIATION II

3. Zuraw's study: Frequency of Tagalog Nasal Substitution varies in the lexicon according to the stem-initial consonant

   ![Graph showing frequency of nasal substitution for various stem-initial consonants]

4. Native speakers are tacitly aware of this pattern
   - Again Zuraw, a “wug” test (following Berko 1958). Preference for the nasally-mutated form (difference between both options, each rated on 1-10 scale)
5. The Law of Frequency Matching

- Hayes, Zuraw et al. (2009) go for broke in their rhetoric:

  *Speakers of languages with variable lexical patterns respond stochastically when tested on such patterns. Their responses aggregately match the lexical frequencies.*


- Sociolinguistic study demonstrates frequency matching by children during real-life phonological acquisition (Labov 1994, Ch. 20).

6. The Law in (much) broader perspective

- Frequency-matching is known to be a common ability in animals (Gallistel 1990, ch. 11); and in humans for nonlinguistic tasks (Hasher and Zacks 1984).

   **MODELING TYPE VARIATION**


- Words are memorized—even inflected ones—as they are heard.

- Psycholinguistic work has strongly supported a **huge capacity for word memorization** in humans (contra early generative phonology, which emphasized data compression)
  
  
- See Baayen’s web site for further work
➢ Basic argument: recognition-ease or speed of fully-inflected forms is dependent on the frequency of the form itself, not its morphological base or paradigm as a whole.

• Back to Zuraw: claims that there is hard analytic work as well as memorization: a stochastic grammar is created from the data — treating them as if they were free variation data.
• I.e.: memorize, but be ready to project.
• If you have a listed form, you generally use it: USE LISTED

8. Zuraw describes a near-optimal human

• Memorization is just great for producing irregulars accurately.
• Children’s memorization capacity is strong but not unlimited.
• So grammar-based back-up is sensible too.
• … and a grammar is essential for production and understanding¹ of novel forms.

9. An alternative: constraint cloning theory

• When you hit a ranking contradiction, make a copy of the relevant Faithfulness constraint, indexing it to the words that are lexically allowed to be more marked.
• Hence the grammar encodes the exceptionality directly.
• … and is non-stochastic
• New forms must be projected — somehow — from the populations of existing forms that violate the various Faithfulness constraints.
• References:

10. Exceptions to the Law of Frequency Matching

• Here, there is a formula for research:
  ➢ Establish nonveridical learning — deviations from Frequency Matching
  ➢ Find the causes — why learn nonveridically?
  ➢ Ideally, form a learning model that learns nonveridically just like people.

11. The sources of nonveridical learning

• Simplicity bias — prefer simpler, more general constraints
  ➢ Hayes and White (2015, *LI*)

¹ When I first heard [mıˈdɛfɪə] for midwifery, I was extremely surprised but knew exactly what was meant, since it is the output of Trisyllabic Shortening (cf. divine ~ divinity).
Jennifer Kuo (2020) UCLA M.A. xxx fill in

- Naturalness bias — prefer outcomes that obey Markedness principles
- Paradigm uniformity bias — prefer outcomes that reduce alternation
- Other
  - See Becker et al, readings on Initial Syllable Faithfulness

12. You can do learning simulations with MaxEnt that incorporate bias

- Wilson (2006) is the pioneering work; see also work of White and Kuo
- The math is simple — a penalty in the objective function for deviating from natural weights.

13. The key interest in studying nonveridical learning

- Are we getting at UG? How else do we explain the observed effects?

**A SUBSET OF THE LITERATURE ON NON-VERIDICAL LEARNING**

14. The ur-reference, I believe

- His UG principle is the **P-map** (Steriade, Zuraw, more later on): avoid alternation when it is phonetically salient.
- ki ~ tʃi is less phonetically salient than ke ~ tʃe
- Artificial grammar experiment: train on ke ~ tʃe, generalizes to ki ~ tʃi, but not the other way around.

**PHONOTACTICS**

15. This is an ancient topic.

- B. L. Whorf published a formula for the English monosyllable in 1940 in *Technology Review*, the MIT alumni magazine:
16. Phonotactics in OT

- Classical Rich Base theory (Prince and Smolensky 1993)
  - The ranking of the universal constraint inventory defines a (non-probabilistic) filter through which the forms of the Rich Base must pass.
  - This is a tough theory to probabilize, since stochastic repair predicts free variation! /pɔɪk/ → [pɔɪk] or [paɪk] (or whatever)
- Maxent-over-GEN theory
  - Simply let the grammar assign a probability to every form in GEN
  - This theory may also have problems: how to get special/general relations in allophony? See:
    - We’ll try it anyway.

17. Can gradient phonotactics be made more rigorous?

- Advantages, perhaps, of maxent-over-GEN model
  - good frequency matching ability (if the constraints are good)
  - ability to disentangle effects of overlapping constraints (see below)
  - statistical testing of hypotheses
18. How to employ maxent-over-GEN: two ways

  - It uses a finite-state machine, following principles developed by Jason Eisner, to explore a vast space of whole-word candidates in feasible time.
- Simple and contingent: Find a subset of the phonotactics, hoping it is somewhat isolated from the rest of the system.
  - E.g., medial clusters, V … V sequences
  - With some patience, this can be done entirely with conventional office software.

DEMO: VOWEL PHONOTACTICS OF TURKISH IN MAXENT-OVER-GEN

19. Turkish vowels

\[
\begin{array}{cccc}
\text{i} & \text{y} & \text{u} & \text{ɯ} \\
\text{e} & \text{ø} & \text{o} & \text{a}
\end{array}
\]

20. The famous rules of Turkish vowel harmony

- Backness Harmony: a suffix must agree with the preceding vowel in backness.
- Rounding Harmony:
  - A high voweled suffix must agree with the preceding vowel in rounding.
  - A low voweled suffix is unrounded

21. Examples

`rope` `girl` `face` `stamp` `hand` `stalk` `village` `end`

ip−in kız−in jüz−ün pul−ün el−in sap−ın køj−yn son−un (gen.sg.)
ip−ler kız−ler jyz−ler pul−ler el−ler sap−ler køj−ler son−ler (nom.pl.)

- There are many suffixes like -In, many like -Ar

22. Does Vowel Harmony govern stems?

- Not clear, because numerous exceptions have been introduced in loanwords.
- This ref.:
  - Says "no".
- With stochastic phonology and significance testing, we can check more carefully.

23. Turn to Homework Assignment