Class 20, 6/6/13: Some themes of the course

1. Assignments etc.

- Due date for term papers: **Friday 6/14**

2. For today

- A once-over of the course content
- Course evaluations — be sure we have enough time.

3. Covered below

- The course actually covered quite a lot.
- No time to summarize all, but at least we can try to mention some recurring **themes**:
  - frameworks and their evaluation
  - biases
  - evaluation of data analyses
  - intuition as evidence
  - individual word identity
  - knobs

BASICS

4. The types of variation

- Class 1 gave the basic taxonomy:
  - **token** (same process, same input, different speaking occasions)
  - **type** (same process, different lexical inputs)
- We soon had reason to make this more nuanced — see below, "The role of individual word identity".

5. The Law of Frequency Matching

- Evidence in support given in Class 2.
- A very nice convenience for us: the same techniques that are applied to token variation can be applied to type variation.
- Interesting partly for being (partly) true, but especially for deviations, which perhaps reflect learning biases of UG — see below

FRAMEWORKS AND THEIR EVALUATION

6. How we can evaluate the frameworks

- Fit to the world's variable phenomena — what patterns can they derive?
- Simplicity and principled character
- Affiliation with an effective learning algorithm

7. Classical optional-rule theory

- The variationists added an interesting twist:
  - The maxent grammar sitting inside the rule! (Class 5)
  - They used logistic regression to output **apply** or **don't apply**, thus letting many factors probabilistically affect application rates.
  - See below on extra-grammatical factors in such rules
- Guy's classical t,d deletion work (Class 17); McPherson and Hayes take on a harder case and think his method won't generalize.

8. Constraint-based frameworks: a fundamental distinction

- Uncertainty about what the grammar is (Anttila, Stochastic OT, Noisy Harmonic Grammar)
- Grammar that itself generates uncertainty (Maxent)

9. Freely-ranked stratum theory

- Originated in 1990's research by Arto Anttila
- Idea of *crucially*-unranked strata (Kie's notation: jagged as opposed to dotted line)
- The line taken here on this framework: not powerful enough
  - you can't let one constraint "modulate" the frequency pattern created by other constraints.
- Also, the learnability situation is *bad*:
  - To this day (surprisingly, and frustratingly); no one has produced an algorithm that learns Anttila-grammars.

10. Stochastic OT

- **Learnability**: has convergence problems (Class 4).
  - Pater-problem (2008), fixed by Giorgio Magri (Class 4).
  - Descent-into-hell problem is hard for Stochastic OT — also a problem for other frameworks like maxent.
  - No objective function
  - No convergence proof except for nonstochastic applications under Magri's variant
  - Practical experience: descent into hell behavior is all too frequent
- **Accuracy and coarseness**: when the GLA works, you only get a rather approximate solution

- **Complexity**: To handle gradient scales, you need to split up constraints instead of using simple functions (Class 18; distance scales for Lyman's Law, other dissimilation)
- Generally hard to **diagnose** its behavior
  - but see Moreton and Smith's useful attempt to provide *rough* general predictions (Class 18)
- Cannot assigns independent probabilities for **multiple-site variation**, which is bad.
  - But does great on "in-synch" multi-site variation — if it exists (Class 15).
- Produces **sigmoid frequency curves** only occasionally and unreliably.
  - Where sigmoids are generated they are (I believe) symmetrical.

11. Maxent (Class 4 et seq.)

- **Computationally very strong**: trustable, accurate, fast
  - It uses an objective function, closely related to model evaluation (see below).
  - There is a learnability proof that covers stochastic cases.
- Predicts **ganging.**
  - Jury is out on this (Flemming)
  - Cases like Kawahara's Japanese (schematically: Class 4; nuances: Class 18) look encouraging.
- Assigns positive probability to **harmonically bounded candidates** (Class 4)
  - a problem? A virtue?
  - Related case: assigns no zeros in stringency-hierarchy cases (unlike Stochastic OT); class 18
- Assigns independent probabilities for **multiple-site variation**, which is good.
  - But cannot handle "in-synch" multi-site variation — if it exists (Class 15)
  - We might fix this with some addition of knobs.
- The **sigmoids** maxent generates are always symmetrical (logistic function).
  - Probably a good thing.
- **Gradient scales** can just be numbers; no need for big constraint families.
- The grammar framework itself cannot handle BH's descent-into-hell example (class 18)

12. Noisy harmonic grammar

- Main discussion: Class 16
- Ok but not great on **accuracy** and **speed**.
- Proven **trustable** only for non-stochastic applications (Boersma/Pater)
- Generally behaves like maxent
  - E.g., it predicts ganging
- No **harmonically-bounded winners** — *if* weights are kept positive.
- Screws up on "out-of-synch" **multiple-site variation** (Class 15)
  - fixable if we change the theory: cell-specific noise
  - Cell-specific noise make NHG more like maxent: harmonically bounded winners, no deriving zero with stringency hiearchies
  - Unaltered classical NHG does great with "in-synch" multi-site variation — if it exists (Class 15)

- Gradient scales can just be numbers; no constraint-explosion needed.
- Sigmoids:
  - Often produces weird asymmetrical sigmoid curves[1] (Class 18)
  - where symmetrical, the shape is the cumulative normal distribution; almost identical to the logistic function
  - The possible problem of asymmetrical sigmoids is fixable (should it be fixed?) if we change the theory: post-multiplicative noise

13. Maxent's doppelganger: logistic regression (Classes 5 and 6)

- Binomial logistic regression *is* maxent — *if*:
  - Two candidates per input (e.g., vowel harmony)
  - Negative weights disallowed
- Hooray, logistic regression allows significance testing! Classes 5 and 6
  - See section below on model evaluation.
- Multinomial logistic regression is not so close to maxent:
  - Needs parallel candidate arrays for all inputs.
- We did a "How to" for beginners: Seuss exercise and R script

BIASES

14. Biases connect to:

- Learnability
- Deviations from the Law of Frequency Matching

15. Conservativity bias

- Keep the weights low in the face of impoverished evidence (Classes 5, 6)
- Derives **Martinian leakage**: studies on "leaky grammars" (Class 7)
- Derives **Ryanian variationogenesis:**
  - What the militant frequency-matcher thinks is evidence for detailed frequency-matching actually comes for free, training on the most common cases.
  - Work of Kevin Ryan on Tagalog (Class 7)

16. Complexity bias

- Moreton and Pater's literature survey (readings)
- They think complexity bias is the most important and doubt other forms.

17. Naturalness bias

- Factorable (Class 8) into:

[1] When symmetry holds, the function is the cumulative normal distribution.

- **Faithfulness bias**: avoid salient alternation
  —Skoruppa/Peperkamp experiment
- **Markedness bias**: avoid articulation/perception difficulties[2]
  —this has met with less success; e.g. French harmony/disharmony experiment (Class 8)

18. Confine learning to a finite specified constraint set

- … as in classical OT
- Becker and Nevins work on initial-syllable faithfulness (Class 19)

MODEL EVALUATION

19. Themes

- Once our grammars are stochastic, we're in the same position as other scientists who produce quantitative models; and general scientific standards for model evaluation apply.
- We want to find the "sweet spot" between overfitting and underfitting.

20. Methods of model evaluation: does a constraint belong in our grammar? (Classes 5, 13, 14)

- **Wald test**, with doubts about it in the literature
- **likelihood ratio test**, with a better reputation
  - Do-able *en masse* for a model in R (Class 13)
  - Do-able *ad hoc* in Excel (Class 14)
  - Comparison of results with Wald test (Class 14) — pretty close in the case we covered.
- **Akaike Information Criterion; Bayesian Information Criterion** (Class 13)
- **cross-validation** (Class 13)
- **random forests** (Stephanie Shih, Class 13B)
  - potential tradeoff of trustability vs. grammar-interpretability
  - application to Loucherbem (Class 19)

21. The challenge of asking, "How are we doing?" (Class 14)

- I.e., how is the model doing overall?
- Methods:
  - Scattergram
  - Histogram
  - Correlation?
  - Summed error?
  - d′ (d-prime)
- The deceptiveness of **overlapping dots** and what one might do about it.

[2] Or: don't violate constraints that are innate and that your genome says to rank high

- Seeking improvement through inspection of **outliers**

INTUITION AS EVIDENCE

22. Task

- Find a trustable way to relate intuition-judgments (e.g. Likert scale) to scores obtained by learning algorithms/corpus.
- This task is in its early stages.
- Discussed in Class 9.
- Leading idea: gaphood is gradient — *maybe* some form is ok, but speaker not sure.

23. The variety of intuition-gathering tasks

- e.g. binary choice, Likert scale, magnitude estimation, pairwise comparison
- how they stack up against each other: Class 10
- A task I recommend you not pursue further: allocation of probability to null parse (*-able* exercise)

24. Theories of gaphood in phonology and morphology

- without an explicit GAP entity — Class 11
- and with (Null parse) — class 12; example from Lofstedt on Swedish

25. The puzzle of how to allocate probability — Class 12

- **Global** choice, amongst all possibilities
- **Local** choice, as a yes/no decision for each candidate
- These are not the same formally, though we don't know how they differ empirically

26. Goodness of forms and diachrony

- Martin's work on reshaping a lexicon — Class 11

THE ROLE OF INDIVIDUAL WORD IDENTITY

Here is a continuum of views on the role of individual word identity in variation.

27. Classical generative phonology

- optional rules; obviously no word identity effects except outright exceptionhood

28. Frequency indexation

- The phonology doesn't know what the word is, but knows its frequency.
  - Work by Coetzee/Kawahara, Hay, surveyed in Class 17.
  - This is a **frequency knob**; see below for style knob.

29. Multiple representations

- All the surface allomorphs of a morpheme are candidates for UR status, and the competition is never resolved; probabilities among UR's remain at the end of acquisition.
  - Pater et al.'s system of finding (often multiple) UR's; Class 19
- Hence every morpheme gets its own individual say — modulated by the grammar — concerning the frequency of its surface forms.

30. Episodism (Classes 19 and 20)

- Even multiple allomorphs are too abstract; rather, you just superimpose all experience as a cloud of structured tokens.
- Typically unspecified mechanisms of generalization permit the theory to account for productive behavior:
  - wug testing, both artificial and real-life
  - phonotactic intuitions
  - structurally and socially-conditioned variation
- Eliminativist episodism
  - Same as above, but the stored tokens are structureless waveforms.

RELATING VARIABLE PHONOLOGY TO STYLE AND SOCIETY: THE QUESTION OF KNOBS

31. Variation and society

- Sociolinguistics covers aspects of society: class, age, gender, ethnicity
- All of them matter to phonology, and there are semi-lawful patterns of the sort we would expect in social science
  - lower-middle class tendency to hypercorrection
  - new sound changes led by young women
  - etc. See Class 1.
- We've pondered the way these can be put into the grammar

32. The degree-of-knowledge distinction (Class 2, Class 19)

- Kawahara and Sano's careful Japanese-geminate-devoicing example.
- **Rich theory**: everyone knows how everyone talks, so stuff like PALATALIZE IF YOU'RE FEMALE goes into the grammar.
  - Is this Labov's view?

- **Impoverished theory**: grammar has a limited number of style knobs — perhaps just one; people have a rich theory of societal structure.
    - The view of the traditional, modularity-oriented generative grammarian.

33. If there are knobs, where do they attach to the grammar?

- Empirical question is how much the phonological processes are in lockstep?
- We saw *some* degree of lockstep in Class 2; but sufficient non-lockstep to make one pessimistic about one-knob phonology (failures of simulations in Class 2; as in special behavior of r-dropping)
- We may be headed for a fairly rich theory, in which multiple constraints are separately connected to multiple knobs.