Ling 211B: Topics in Phonological Theory
Topic: Phonological variation

Fall 2017

Brian W. Smith

LOGISTICS

Course website: on bcourses
E-mail: bwsmith@berkeley.edu
Office hours: Mo 1:00PM–2:00PM, We 11:00AM–12:00PM (and open door policy)
Time/place: TuTh 2:00PM–3:30PM @ Dwinelle 1303

We meet on the following dates:

Week	**W00**	**W01**	**W02**	**W03**	**W04**	**W05**	**W06**	**W07**	**W08**	**W09**
Tue	—	8/29	9/5	9/12	9/19	9/26	10/3	10/10	10/17	10/24
Thurs	8/24	8/31	9/7	9/14	9/21	9/28	10/5	10/12	10/19	10/26

Week	**W10**	**W11**	**W12**	**W13**	**W14**
Tue	10/31	11/7	11/14	11/21	11/27
Thurs	11/2	11/9	11/16	—	**11/29 ⟵**presentations

DESCRIPTION

Phonological data is subject to variation, both within and across speakers and lexical items. Relatively recently, phonologists have worked on developing theories for the treatment of variation, extending OT-like models to new cases of non-categorical data from corpora and experiments.[1] This seminar addresses 'free' variation and lexical variation in phonology (and a little bit in morphosyntax), with an emphasis on building and comparing grammatical models.

[1] At the 2016 Annual Meeting of Phonology, 58% of the talks presented variation data, 32% presented a grammar model of variation, 37% used corpus data, and 37% included a human experiment (AMP numbers courtesy of Kie Zuraw).

We'll focus on the advantages of constraint-based models of variation:

1. They make explicit (and implicit) connections to models used in statistics and sociolinguistics. We'll discuss, for example, how MaxEnt Grammar relates to Logistic Regression and VarbRul.
2. They make explicit predictions about data: both in terms of general patterns and in terms of quantifiable model fit. We'll discuss various ways to evaluate and compare models, including statistical tests and cross-validation (using holdout/test data).
3. They gracefully handle 'real' data, complete with exceptions and noise. We'll discuss and practice using raw corpus and experimental data to fit our models, and read many papers with both.
4. They're compatible with robust learning algorithms, with many software implementations. We'll run through a number of learning algorithms both by hand and using software.

COURSE GOALS

(1) characterize the range of factors that influence phonological variation
(2) compare models of phonological variation and their learning algorithms
(3) gain hands-on experience with related software

REQUIREMENTS

Enrolled?

- — Attend class and do the readings. Let me know if you can't make it to class.
- — Complete two or three practice data sets, most of which require using software tools (we'll go over these tools in class)
- — Present two or three papers during the semester. Papers that are marked as 'optional' are presentable, along with any other paper that's relevant
- — Write a final paper, which takes a set of variable data and analyzes it in at least two models that we discussed. Evaluate the models using quantitative model fit or hold-out data. If a particular model does better on the data, characterize the source of the difference. (You can use your own personal data for this, or existing data from the phonology/phonetics/sociolinguistics literature)
- — Present your final paper in class (15-20 minute presentation)

Sitting in?

- — Feel free to attend even if you haven't done the readings
- — Consider presenting a paper or related original research
- — Attempt the problem sets (these are the most useful part)

SCHEDULE

Weeks 0–1: Overview and review

Coetzee & Pater (2011) provide an overview of phonological variation and its models
OT review: ranking arguments, comparative tableaux, and learning algorithms

Weeks 1–3: Partially ordered constraints

Anttila (1997) defines a model of free variation using partial rankings (=POC)
Boersma (2000) summarizes Anttila's dissertation and raises some issues
Anttila et al. (2008) use POC to model variation in Singaporean English
Opt.: **Côté (2007)** uses Antila's model for another case of variation from French
Opt.: **Kiparsky (1994**) sketches an OT model of variation, which pre-dates POC

Weeks 4–5: Stochastic OT and the GLA

Boersma & Hayes (2001) propose a learning algorithm for Stochastic OT: the GLA
Pater (2008) shows a case that breaks the GLA
Albright & Hayes (2006) use the GLA to deal with induced 'junk' constraints
Opt.: **Boersma & Levelt (2000)** model acquisition of syllable structure using the GLA
Opt.: **Magri (2012)** provides a fix for the problem identified in Pater (2008)
Opt.: **Jarosz (2010)** expands on Boersma & Levelt, taking frequency into account
Opt.: **Jesney (2016)** tests predictions made by the GLA about learning rates

Weeks 5-6: (Noisy) Harmonic Grammar

Pater (2016) provides an overview of Harmonic Grammar
Walker (2017) provides further comparison of HG and Local Conjunction
Pater & Boersma (2016) show how the GLA can be adapted for Noisy HG
Opt.: **Jesney & Hsu (2016)** use Harmonic Grammar to model loanword adaptation
Opt.: **Kawahara (2006)** uses HG to model cumulativity in Japanese

Week 6: Maximum Entropy Grammar (and model comparison)

Goldwater & Johnson (2003) show how MaxEnt can be used to Anttila's (1997) data
Hayes, Wilson, & Shisko (2012) use MaxEnt for metrics, exemplify model comparison
Opt.: **Shih (2017)** uses conjoined constraints in MaxEnt as interaction terms

Week 7: Ganging and cumulativity in MaxEnt vs. StOT

Smith & Pater (2017/ms) compare MaxEnt, StOT, and HG with respect to ganging
Zuraw & Hayes (to appear) compare MaxEnt, StOT, and HG
Opt.: **Irvine & Dredze (2017/ms)** do the same, but for syntactic variation in Czech
Opt.: **Jaeger & Rosenbach (2008)** compare MaxEnt and StOT

Weeks 8–9: Other models (and local optionality)

Benor & Levy (2006) use logistic regression for English binomial ordering
Kaplan (2012) proposes markedness suppression, using data with local optionality
Opt.: **Hilpert (2007)** uses logistic regression, and also illustrates the use of holdout data

Opt.: **Coetzee (2006)** proposes a model in which variation comes from ranking losers
Opt.: **Kaplan (2016)** reanalyzes the data from his 2012 paper using POC
Opt.: **Johnson (2009)** shows how Varbrul relates to logistic regression
Opt.: **Shih & Zuraw (to appear)** use logistic regression for Tagalog word order
Opt.: **Hayes (to appear)** considers different formulations of noise in Noisy HG

If we're ahead of schedule: Frequency and 'performance'

Tily and Kuperman (2012) illustrate frequency effects in Dutch epenthesis
Coetzee & Kawahara (2013) model frequency effects using constraint scaling in HG
Opt.: **Smith & Moore-Cantwell (to appear)** model frequency effects in MaxEnt

Weeks 10–11. Phonotactics and naturalness

Hayes & Wilson (2008) propose a MaxEnt model of phonotactics
Hayes & White (2013) test machine-learned phonotactics with real speakers
Martin (2011) shows how phonotactics can affect other parts of the grammar
Opt.: **Smith (ms)** similarly argues that speakers extend phonotactics to new suffixes
Opt.: **Kager & Pater (2012)** identify a phonotactic speakers know but isn't machine learned

Weeks 12–13: The law of frequency matching and lexical variation

Ernestus & Baayen (2003) show frequency matching for final devoicing in Dutch
Hayes et al. (2009) model frequency matching in Hungarian in MaxEnt
Zuraw (2010) models frequency matching in Tagalog in Stochastic OT
Becker et al. (2011) model frequency matching (and non-matching) in Turkish
Opt.: **Becker and Gouskova (2016)** use multiple phonotactic grammars for Russian yers
Opt.: **Pater et al. (2012)** present a learning model for lexical variation in MaxEnt
Opt.: **Saffran et al. (1996)** show that infants can use transitional probabilities to learn word boundaries

Weeks 14: "Naturalness" (complexity and phonetic substance) and artificial grammar learning

Pycha et al. (2003) consider the roles of simplicity and phonetic grounding in the learning of phonological patterns
Moreton and Pertsova (2016) consider the role of different types of learning ('explicit' and 'implicit') in artificial grammar learning
Opt.: **Moreton et al. (2017)** compare the effects of complexity on pattern learning across different domains (linguistic and visual)
Opt.: **Wilson (2006)** shows that artificial grammar learning is sometimes influenced by phonetic substance