Variation in phonology
Graduate seminar (Ling 219)

Course website: on eCommons
E-mail: brwismit@ucsc.edu
Office hours: TBD
Time/place: 10:40am-12:15pm / Monday, Wednesday / Stevenson Lib 102 (The Cave)

We meet on the following dates:

Week	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**
Monday	4/3	4/10	4/17	4/24	5/1	5/8	5/15	5/22	~~5/29~~	6/5
Wednesday	4/5	4/12	4/19	4/26	5/3	5/11	5/17	5/24	5/31	6/7

5/29 = memorial day

Description and goals. Phonological data is subject to variation, both within and across lexical items and speakers. 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 its interaction with phonology, with an emphasis on building and comparing grammatical models. There are two goals: (1) survey the literature on phonological variation; (2) gain hands-on experience with models of variation.

We'll especially 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 practice using software tools such as the MaxEnt Grammar Tool, Praat, the UCLA phonotactic learner, maybe Excel, and maybe R. We'll also discuss some of these learning algorithms in detail, and even execute a few by hand.

[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). Recently = the last decade. Among the 40 readings in the reading list: 6 are pre-2006; 17 are 2006-2010, and 17 are post-2010.

Course outline. Two guest lectures are (tentatively) planned. Stephanie Shih (UC Merced) will discuss computational model selection (AIC/BIC and random forests), and Claire Moore-Cantwell (UConn) will present original work on modeling the interaction of frequency and variation in a probabilistic grammar.

Part 1: Empirical: a typology of phonological variation

- Intraspeaker vs. interspeaker, free vs. lexical, type vs. token
- Gradient phonotactic knowledge, gradient grammaticality, gradient productivity
- Role of frequency and formality
- Role of phonology in syntactic, morphological, and sociolinguistic variation

Part 2: Theories of variation (with software implementations)

- Distributions over constraints rankings
  - Partial ranking/Free ranking (Software: T-Order Generator and OT-Soft)
  - Stochastic OT and the Gradual Learning Algorithm (Software: Praat)
- Weighting instead of ranking
  - Noisy Harmonic Grammar (Software: Praat)
  - Maximum Entropy Grammar (Software: MaxEnt Grammar Tool)
- Connections outside of phonology
  - Logistic/linear regression from statistics
  - VarbRul from sociolinguistics
- Explorations of the various learning algorithms used in these theories

Part 3: Model evaluation and selection

- Levels of model selection: constraint sets, priors, parameters
- Methods of model selection and measures of fit
  - statistical approaches (likelihood-ratio test, AIC)
  - cross-validation

Part 4: More on learning algorithms

- Gradual vs. batch learning (online vs. offline)
- Some results of adding biases as priors, sampling, and decay to learning

Expectations:

• Read the readings and present two to four papers during the quarter (depending on how many are enrolled). Ideal world: about one student presentation per 1.5 class meetings

• Do the written assignments. There will be three or four very straightforward assignments that require using new software. I'll give you data, and you build a model.

• Write a term paper, which presents and evaluates a model of non-categorical data. The variation can be phonological, morphological, syntactic, or some combination of the three. Ideally, you'll analyze variation that's conditioned by phonological factors, but this isn't strictly necessary.

Readings by part. We'll choose a subset of these as we go, and add some if they're relevant. E-mail me if you'd like to discuss or present a paper not on this list.

Part 1: A typology of phonological variation

Overview

Coetzee and Pater (2011)
Guy (1980) : English t/d deletion
Bayles et al. (2016) : French schwa

Frequency effects

Bybee (2000) : English (t/d deletion)
Tily and Kuperman (2012) : Dutch

Gradient phonotactic knowledge (and probability matching)

Hayes et al. (2009) : Hungarian
Kager and Pater (2011) : Dutch
Becker et al. (2011) : Turkish

Sociolinguistic variation

Mendoza-Denton et al. (2003) : overview

Syntactic variation (conditioned by phonology)

Shih and Zuraw (2017/ms) : Tagalog
Benor and Levy (2006) : English

Morphological variation (conditioned by phonology)

Plag (1999/excerpt)[2] : English
Smith (2016/ms) : English

Part 2: Theories

Partial/Free Ranking

Anttila (1997) : Finnish
Itô and Mester (1997) : Japanese
Côté (2007) : French (schwa)
Anttila and Andrus (2006) : manual for T-order generator

[2] NB: Rumor is that Ingo Plag will be around campus! Consider meeting with him if you're interested in this line of work.

Stochastic OT (and the Gradual Learning Algorithm)

Smith and Moreton (2011)	: good overview
Boersma and Hayes (2001)	: the GLA
Zuraw (2010)	: Tagalog, relies on the GLA
Pater (2008)	: data that breaks the GLA

Other approaches to variation and gradience in OT

Coetzee (2006)	: ranking losers (English t/d deletion)
Kaplan (2011)	: suppressing violations
Keller (2005)	: linear OT

Harmonic Grammar (and Noisy Harmonic Grammar)

Pater (2016)	: accessible overview
Kawahara (2006)	: Japanese
Potts et al. (2010)	: Lango
Jesney and Tessier (2011)	: HG-GLA and biases
Coetzee and Kawahara (2013)	: English (t/d deletion), Japanese
Zuraw and Hayes (2017)	**:** French

Maximum Entropy Grammar

Goldwater and Johnson (2003)	
Jäger and Rosenbach (2006)	: compares with StOT
Jesney (2007/handout)	: compares with Noisy HG
Hayes and Wilson (2008)	: Shona, Wargamay
Pater et al. (2012)	: hidden structure learning in MaxEnt
Smith and Pater (2017/ms)	: French (schwa)
McPherson and Hayes (2016)	: Tommo So

Connections to statistics and sociolinguistics

Johnson (2009)	: discussion of R implementation of VarbRul

<u>Part 3: Model evaluation and selection</u>

Cross-validation exemplified by Hilpert (2007)
Model comparison exemplified by Hayes and Shisko (2012)
AIC-based model selection exemplified by Kuperman and Bresnan (2012)

<u>Part 4: Learning algorithms, cont'd</u>

TBD, based on interest/output of Part 2