Syllabus: Quantitative models in phonology

Time	Wednesday 10-1 pm
Place	2-320
Office Hours	Tu 2:30 – 4:30 & by appointment (office: 3-407)
E-mail/ Phone	jongho@snu.ac.kr / 880-6169

Prerequisite

Some basic knowledge of Optimality Theory is necessary.

Description & Goals

The typical data patterns employed in traditional generative phonology are systematic with no variation. Even when exceptions or variant output forms are attested, they have often been ignored or much simplified in formal analysis. However, the variable application of phonological processes (and gradient grammaticality intuitions) have recently attracted much attention with the understanding that the exceptions and variants occur in a systematic way in many cases. Moreover, the application rate of such variable processes is often influenced by phonological factors similar to those which condition the occurrence of corresponding categorical processes. As the need to explain the variation in phonology increases, probabilistic versions of Optimality Theory (OT), including a Maximum entropy (maxent) grammar, and their related learning models have been developed.

In this course, in order to explore characteristic patterns of variable phonological processes, probabilistic grammatical theories and their learning models, we survey recent studies of the variation in phonology. In so doing, we will focus on the following topics:

- an overview of statistical methods and their relation to maxent
- an overview of probabilistic OT theories
- frequency matching in variation
- variation in Korean phonology
- phonotatics and phonotactic learning

The goals of this course include not only to address the issues involved in the above topics but also to learn how to explain variation in your own data using the software, shown below:

- ✓ OTSoft (http://www.linguistics.ucla.edu/people/hayes/otsoft/)
- ✓ Praat (http://www.fon.hum.uva.nl/praat/)
- ✓ Maxent Grammar Tool
  (http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/)
- ✓ UCLA phonotactic learner
  (http://www.linguistics.ucla.edu/people/hayes/Phonotactics/Index.htm)

(I'll probably not discuss the following in class.)
- ✓ Harmonic Grammar in R (http://blogs.umass.edu/hgr/)
- ✓ OT-Help (http://people.umass.edu/othelp/)

In addition, I will sometimes use the free statistical package called R in class. It is good to install it on your computer:

- ✓ R (https://www.r-project.org/), RStudio (https://www.rstudio.com/)

This course is an extension of the second half of my phonology II class (fall 2014), and modelled in part on Zuraw (2012) and Hayes & Zuraw (2013).

Readings

Statistics (optional)	Johnson, Keith (2008) Quantitative methods in linguistics. Malden, MA: Blackwell. Baayen, R. H. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. (draft: http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf)
Additional required readings	If a paper discussed in class can be downloaded from the Web, I will just inform you of the Web address. Otherwise, its electronic copy will be stored in the etl class site or sent to you by email.

Requirements

- Homework assignments (20%)
- Lead discussion of one paper in the required reading list (10%)
- A final term paper and its presentation (50%)
- Readings and class participation (20%)

Final term paper

- Select topic and meet with me by Oct. 28.
- Submit a one-page outline of the paper by Nov. 18.
- Paper Presentation: Last week (Dec. 11)
- Paper Due: Dec. 14

If you have any questions on the required reading papers and the problem, you can talk to me. For short questions, you can call me; for longer discussions, you can either take my office hours or make an appointment with me. E-mail may be useful for both purposes. Surely, you can come by my office with questions.

Tentative Class Schedule (★student presentation)

Week	Topics	Readings
1. 9/2	Introduction: basics of statistics	Johnson (2008, chs 1-3)
2. 9/9	Regression	Johnson (2008, chs 4-5)
3. 9/16	Maxent (1)	Goldwater & Johnson (2003), Hayes & Wilson (2008)
4. 9/23	Maxent (2)	
5. 9/30	Overview of variation	Coetzee & Pater 2011
6. 10/7	Partially-ordered constraint theory	Anttila (1997b)
7. 10/14	Stochastic OT Noisy HG	Boersma & Hayes (2001), Coetzee (2009a)
8. 10/21	Mid-term (no class)	
9. 10/28	Productivity and frequency matching (1)	Pater et al. (2012), ★Ernestus & Baayen. (2003), ★Becker et al. (2011)
10. 11/4	Productivity and frequency matching (2)	★Hayes & Londe (2006), ★Hayes et al. (2009), Zuraw (2010)
11. 11/11	Variation in Korean (1)	Zuraw (2011), ★Ito (2014)
12. 11/18	Variation in Korean (2)	Jun (to appear), Park (2015)
13. 11/25	Phonotactics and phonotactic learning (1)	★Hayes & Wilson (2008), ★Hayes & White (2013)
14. 12/2	Phonotactics and phonotactic learning (2)	Wilson (2006) ★Gallagher (2013)

15. 12/9	Paper presentation	
	Paper Due (12/14)	

Bibliography

Anttila, Arto (1997a) *Variation in Finnish Phonology and Morphology*. Doctoral dissertation. Stanford, CA: Stanford University.

Anttila, Arto (1997b) Deriving variation from grammar. *CILT* 146.

Anttila, Arto (2002) Morphologically Conditioned Phonological alternations. *Natural Language and Linguitic Theory* 20.1: 1-42. [available at http://www.stanford.edu/~anttila/research/papers.html]

Boersma, Paul & Bruce Hayes (2001) Empirical tests of the gradual learning algorithm. *LI* 32.1.

Becker, Michael; Nihan Ketrez; and Andrew Nevins. (2011) The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87.1: 84-125.

Coetzee, Andries W. (2009a) An integrated grammatical/non-grammatical model of phonological variation. In Young-Se Kang, Jong-Yurl Yoon, Hyunkyung Yoo, Sze-Wing Tang, Yong-Soon Kang, Youngjun Jang, Chul Kim, Kyoung-Ae Kim, & Hye-Kyung Kang, eds. *Current Issues in Linguistic Interfaces. Volume 2.* Seoul: Hankookmunhwasa. p. 267-294.

Coetzee, Andries W. (2009b) Learning lexical indexation. *Phonology* 26.1, 109-145.

Coetzee, Andries & Joe Pater (2011) The place of variation in phonological theory. In J. Goldsmith, J. Riggle & A. Yu (eds.) *The Handbook of Phonological Theory*, 2nd edition. Malden, MA and Oxford, UK: Blackwell, 401-434. [ROA-946]

Hayes, Bruce and Zsuzsa Londe (2006) Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23.1.

Hayes, Bruce & James White (2013) Phonological naturalness and phonotactic learning. *LI* 44: 45-75.

Hayes, Bruce & Colin Wilson (2008) A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379-440.

Hayes, Bruce; Kie Zuraw; Péter Siptár, and Zsuzsa Londe (2009) Natural and unnatural constraints in Hungarian Vowel Harmony. *Language* 85: 822-863.

Hayes, Bruce & Kie Zuraw (2013) Ling 251A/B: Topics in Phonetics and Phonology—Variation in phonology. Class handouts, UCLA. (http://www.linguistics.ucla.edu/people/zuraw/251_2013/)

Ernestus, Mirjam & R. Harald Baayen. (2003) Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.

Gallagher, Gillian (2013) Learning the identity effect as an artificial language: bias and generalization. *Phonology* 30.2: 253-295.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Andres Eriksson & Ö sten Dahl (eds.) *Proceedings of the Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111-120.

Inkelas, Sharon (2000) Phonotactic blocking through structural immunity. In Siebels and Wunderlich (eds.), *Lexicon in focus. Studia Grammatica* 45. Berlin: Akademie-Verlarg, 7-40.

Ito, Chiyuki (2014) Compound tensification and laryngeal co-occurrence restrictions in Yanbian Korean. *Phonology* 31.3: 349-398.

Jun, Jongho (forthcoming) Korean n-insertion: a mismatch between data and learning. *Phonology*.

Kiparsky, Paul (1993) An OT perspective on Phonological variation. Handout. [available at http://www.stanford.edu/~kiparsky/]

Park, Seon (2015) Phonological variation of Korean /h/-deletion. [written in Korean] MA thesis. Seoul National University.

Pater, Joe (2000) Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* 17.2, 237-274.

Pater, Joe (2008) Gradual learning and convergence. *LI* 39.2, 334-345.

Pater, Joe (2009) Morpheme-specific phonology: constraint indexation and inconsistency resolution. In Steve Parker, (ed.) *Phonological Argumentation: Essays on Evidence and Motivation*. London: Equinox. 123-154. [downloadable from http://people.umass.edu/pater/index-5.html]

Pater, Joe, Robert Staubs, Karen Jesney & Brian Smith (2012). Learning probabilities over underlying representations. In *Proceedings of the Twelfth Meeting of ACL-SIGMORPHON: Computational Research in Phonetics, Phonology, and Morphology*. Learning probabilities over underlying

representations.

Wilson, Colin (2006) Learning phonology with substanctive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30: 945-982.

Zuraw, Kie. (2000) *Patterned exceptions in phonology*. PhD dissertation, UCLA.

Zuraw, Kie. (2002) Aggressive reduplication. *Phonology* 19.3: 395-539.

Zuraw, Kie (2010) A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28.2: 417-472.

Zuraw, Kie. (2011) Predicting sai-siot in Korean compound nouns: Phonological and non-phonological factors. Handout presented at the 21st Japanese/Korean Linguistic Conference, Seoul National University.

Zuraw, Kie (2012) Studies in Phonology 2: Quantitative models of phonological variation. Seoul National University block seminar (fall 2012) class handouts.
(http://www.linguistics.ucla.edu/people/zuraw/courses/SNU.html)