

Models of Phonological Learning: Overview and Syllabus

Joe Pater

Ling 751, Tuesday, January 26, 2009

1 The theories

The last 10-15 years have seen a dramatic increase in the interaction between research in phonological theory and the study of phonological acquisition, thanks in large part to OT. However, there are some core outstanding issues in the use of OT for the study of phonological acquisition:

1. Where do the constraints come from? Innateness isn't a good enough answer, given that many processes in child phonology have no parallel in adult phonology (see esp. Buckley 2003). But we do also see parallels (see esp. Boersma and Levelt 2003). We might say that (some) constraints are constructed (see e.g. Pater 2002), but can we actually provide a more explicit account?
2. How do we handle the gradualness, variation, exceptionality (and regularization) seen in acquisition? None of these is captured by the original version of OT (Prince and Smolensky 1993/2004) or by its original learning algorithm (Tesar and Smolensky 2000).
3. The most successful OT account of gradualness and variation in acquisition (Boersma 1997 *et seq.*) adds numerical values and a stochastic component. Some critics have said this is no longer OT, but connectionism; more accurate would be to say that it is a hybrid. What happens when we go all the way, and drop ranking completely? We get a convergent gradual algorithm (Jäger 2006; Boersma and Pater 2008; cf. Pater 2008a on OT-GLA). But what else? (see Farris-Trimble 2008; Jesney and Tessier 2008).
4. As we'll see, exceptions in acquisition are harder to get a handle on. Some proposals exist for the learning of exceptions in OT (e.g. Pater 2005; Tesar 2006; Apoussidou 2007; Pater 2008b), but they don't (yet?) have anything to say about the gradual spread of a rule through the lexicon, in child phonology or in historical phonology. Potentially relevant here are stochastic approaches to phonotactics, such as Hayes and Wilson (2008), and to alternations, like those reviewed in Coetzee and Pater (2008).
5. There exists considerable support for the notion that "simple" patterns are learned more readily than "complex" ones, both in natural language acquisition (e.g. Payne 1980; Chambers 2000) and in experimental studies (e.g. Pycha et al. 2003; Moreton 2008). Simplification or reanalysis is also often invoked in discussions of historical change (see e.g. Kiparsky 2000; Wheeler 1979). OT has no simplicity metric, and as far as I know, the SPE one was used to choose between possible analyses, not to account for ease of acquisition. We've made some baby steps in working on this (Pater et al. 2008); much bigger ones seem possible.

If we have a theory of phonological grammar in which constraints are learned, their weights are adjusted in a gradual fashion, and variation and regularity/exceptionality are dealt in a stochastic fashion, we need to ask what the relationship of such a theory is to connectionist, analogical, and exemplarist models. (Given the popularity of these models in cognitive science in general, and in phonology in particular, it's worth learning about them anyway!)

2 The data

If we are going to advance our understanding of theories of phonological learning and of their relationship to theories of phonology, we also need to make progress on the data. There are in fact non-trivial questions about what data a theory of phonological learning should account for. Here are some answers.

1. *None.* A significant amount of research in this area does not concern itself with acquisition data at all, but instead tries to answer the question of how a learner could in principle acquire the types of grammars posited in phonological theory. This is research on *learnability*. This research is important, and we will be discussing it in most detail with respect to the interesting work that has been done on stress (including Apoussidou 2007). There are also published articles that draw a competence-performance distinction in such a way as to put seemingly all acquisition data outside of the domain of phonological theory (Hale and Reiss 1998). Insofar as this research offers no results in either learnability or in acquisition, I see this as outside of the domain of our course (unless you insist...)
2. *Naturalistic production data.* The bulk of OT research on phonological acquisition has used ranked constraints to account for the phonological shape of children's utterances. We'll spend some time looking directly at data of this type, including some from a database collected by A.J. Compton and colleagues in the 1970's, which I put into a spreadsheet when I wrote my dissertation. Like Hale and Reiss, you may well ask how we know that these data are the product of phonological grammars, rather than immature motor systems. I would ask you to consider the criteria that in general determine whether data belong to phonology rather than phonetics, and when looking at the data, see if it looks like phonology to you! My own take on this is that if our theories shed new light on these data, then that's justification enough. But there's likely a good research question or two lurking in there somewhere...
3. *Artificial language learning.* Many questions about language learning can only be answered under controlled laboratory conditions. There has recently emerged a substantial body of research in which subjects are taught miniature languages in the laboratory (see Pater and Tessier 2006; Moreton 2008 for overviews). We'll be looking at some of this work, and will have the good fortune to be able to talk about it with the experts - Angela Carpenter (Carpenter 2008) and Anne Pycha (Pycha et al. 2003) as well as Elliott Moreton.
4. *"Wug" testing.* A methodology invented by Jean Berko Gleason to determine whether knowledge of morphophonology is productive. Michael Becker, Bruce Hayes, Kie Zuraw and others have been putting this methodology to good use in arguing that the generalizations learners make are sometimes different from the data they are exposed to. This has also been used in testing the types of generalizations learners make in acquiring the past tense alternations in English - a battleground for the connectionism *vs.* rules debate.

5. *Historical change.* With a good theory of learning, we should be able to explain some aspects of historical change. Unfortunately, I don't know the historical literature very well, so it will largely be up to you to find relevant cases to look at (I'll try too!) Try looking for cases in which a pattern is said to have been "simplified" or "regularized". A lot of our discussion of learning will be focused on stress, so stress cases would be particularly useful.
6. *Infant speech perception and word segmentation.* Some of the most important recent findings about phonological learning have been in this area. This research looks at knowledge of segmental categories and phonotactic regularities, and infants' abilities to segment speech. We'll look at a small subset of this work in the area of word segmentation.

3 What you'll be doing

I want everyone in this class to do at least some computational modeling and analysis of acquisition data, and we'll start right into both of these the second and third weeks. For the classes in that week, and for some of our later classes, you'll complete small assignments in these areas.

For those of you who are registered, you should be thinking right away about what you might like to do as a term project, and start reading and talking to me and other participants in the course. Some possibilities:

1. Phonological analysis of production data
2. A pilot experiment
3. Learnability research/computational modeling

I would also like all registered students to make **two reports** to the rest of the class over the course of the semester. These will be *brief* presentations (15-20 mins.) of a research result, or of an article that relates to the topic of that day's class, or to your research project. If you present an article, you *should not try to summarize the whole article* - just convey to the rest of the class what the research is about, why it is important and/or how it can be improved. I'd like unregistered participants to do at least one of these reports.

I will lead the discussion of most of the papers in the class (though I will defer to the experts where possible!), but I am going to assume you've read the paper, and will be able to contribute to discussion (though I recognize that prior background will affect your ability to do so). The reading load is not particularly heavy for a graduate seminar (except for the first couple classes) - I'm assuming that you will read these papers carefully.

4 Plans

Here is how I'm envisioning the semester proceeding; this is subject to discussion and change.

Date	Subject	Readings
1/27	Course overview	This document
1/29	The study of phonological acquisition in OT	Boersma and Levelt 2003; Pater 2002
2/3	Other data, other perspectives	Buckley 2003; Macken 1995 Menn 2004; Velleman and Vihman 2002
2/5	Variation in phonological theory	Coetzee and Pater 2008
2/10	Analysis of variable child data	Appendix, Pater 1996
2/12	Stochastic phonotactics and constraint induction	Hayes and Wilson 2008
2/17	Another approach to constraint induction	Boersma and Pater 2007; Pater et al. 2008
2/19	Moreton guest lecture	Moreton 2008
2/24	Restrictive learning and child phonology	Jesney and Tessier 2008
2/26	Phonotactic learning and language change	Martin 2007
3/3	Velleman guest lecture	
3/5	Probabilistic approaches to language change	Zuraw 2003
3/10	Frequency and language change	Bybee 1976; Phillips 1998
3/12	OT and language change	Kiparsky 2000
	<i>Spring Break</i>	
3/24	Learning stress - RIP	Tesar 1998
3/26	Learning stress - URs	Apoussidou 2007
3/31	Stress and exemplar models	Gillis et al. 2000
4/2	Stress and analogical modelling	Eddington 2000
4/7	Stress and probabilistic grammar	Narad 2009?
4/9	Stress and word segmentation: Data	Jusczyk et al. 1999 Thiessen and Saffran 2003
4/14	Statistical models of word segmentation	Frank et al. 2007
4/16	Stress and word segmentation: Models 2	Gambell and Yang 2005
4/21	<i>No Class</i>	
4/23	Word segmentation errors in child speech	Chevrot et al. 2008
4/28	Morpheme segmentation and phonological change	Hale 1973
4/30	Artificial language learning 1: Stress	Carpenter 2008
5/5	Artificial language learning 2: Segmental phonology	Pycha et al. 2003
5/7	Wug testing 1: Strong UG	Becker et al. 2008
5/12	Wug testing 2: Weak UG	Hayes et al. 2008

References

- Apoussidou, Diana. 2007. The learnability of metrical phonology. Doctoral Dissertation, University of Amsterdam.
- Becker, Michael, Andrew Nevins, and Nihan Ketrez. 2008. The surfeit of the stimulus: Grammatical biases filter lexical statistics in turkish voicing deneutralization. URL <http://roa.rutgers.edu/view.php3?roa=1001>, ms, Reed College, Yale University and Harvard University.
- Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21:43–58. Available on Rutgers Optimality Archive, <http://ruccs.rutgers.edu/roa.html>.
- Boersma, Paul, and Clara Levelt. 2003. Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 1–50.
- Boersma, Paul, and Joe Pater. 2007. Constructing constraints from language data: The case of Canadian English diphthongs. Paper presented to the North Eastern Linguistic Society (handout available at <http://people.umass.edu/pater/boersma-pater-nels.pdf>).
- Boersma, Paul, and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Ms, University of Amsterdam and UMass Amherst. Available at <http://roa.rutgers.edu/>.
- Buckley, Eugene. 2003. Children’s unnatural phonology. In *Proceedings of the Berkeley Linguistics Society 29*, ed. Pawel Nowak and Corey Yoquelet, 523–534.
- Bybee, Joan. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In *Current progress in historical linguistics*, ed. William Christie, 95–105. Amsterdam: North Holland.
- Carpenter, Angela. 2008. *Learning artificial languages: The role of universal grammar*. VDM Verlag Dr. Mueller e.K. URL <http://people.umass.edu/pater/Carpenter.pdf>.
- Chambers, Jack. 2000. Dialect acquisition. *Language* 68:673–705.
- Chevrot, Jean-Pierre, Celine Dugua, and Michel Fayol. 2008. Liaison acquisition, word segmentation and construction in french: a usage-based account. *Journal of Child Language* URL <http://people.umass.edu/pater/chevrot2008.pdf>.
- Coetzee, Andries, and Joe Pater. 2008. The place of variation in phonological theory. In *The handbook of phonological theory, 2nd edition*, ed. John Goldsmith, Jason Riggle, and Alan Yu. Blackwell. To appear. Available at <http://roa.rutgers.edu/>.
- Eddington, David. 2000. Spanish stress assignment within the analogical modeling of language. *Language* 76:92–109. URL <http://linguistics.byu.edu/faculty/eddingtond/STRESS.pdf>.
- Farris-Trimble, Ashley. 2008. Cumulative faithfulness effects in phonology. Doctoral Dissertation, Indiana University. URL [ROA-991](http://roa.rutgers.edu/roa-991).

- Frank, Michael C., Sharon Goldwater, Vikash Mansinghka, Tom Griffiths, and Joshua Tenenbaum. 2007. Modeling human performance on statistical word segmentation tasks. . , 2007. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* .
- Gambell, Timothy, and Charles Yang. 2005. Word segmentation: Quick but not dirty. URL <http://www.ling.upenn.edu/~ycharles/papers/quick.pdf>, ms., Yale University.
- Gillis, Steven, Walter Daelemans, and Gert Durieux. 2000. Lazy learning: A comparison of natural and machine learning of stress. In *Models of language acquisition: Inductive and deductive approaches*, ed. P. Broeder and J. Murre, 76–99. Oxford University Press. URL <http://www.cnts.ua.ac.be/Publications/2000/GDD00>.
- Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. In *Current trends in linguistics*, ed. Thomas Sebeok, volume 9: Diachronic, Areal and Typological Linguistics, 401–458. The Hague: Mouton.
- Hale, Mark, and Charles Reiss. 1998. Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29:656–683. Available on Rutgers Optimality Archive, <http://ruccs.rutgers.edu/roa.html>.
- Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* To appear.
- Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Londe. 2008. Natural and unnatural constraints in hungarian vowel harmony. URL <http://www.linguistics.ucla.edu/people/hayes/papers/HayesZurawSiptarLondeNaturalUnnaturalConstraintsI> ms, UCLA.
- Jäger, Gerhard. 2006. Maximum entropy models and Stochastic Optimality Theory. In *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, ed. Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI. To appear.
- Jesney, Karen, and Anne-Michelle Tessier. 2008. Biases in harmonic grammar: On the road to restrictive learning. MS, University of Massachusetts, Amherst and University of Alberta.
- Jusczyk, Peter, Derek Houston, and M. Newsome. 1999. The beginnings of word segmentation in english-learning infants. *Cognitive Psychology* 39:159–207.
- Kiparsky, Paul. 2000. Analogy as optimization. In *Analogy, levelling, markedness: Principles of change in phonology and morphology*, ed. Aditi Lahiri, number 127 in Trends in Linguistics. Studies and Monographs. Mouton de Gruyter. URL <http://www.stanford.edu/~kiparsky/Papers/maurachpaper.pdf>.
- Macken, Marlys. 1995. Phonological acquisition. In *The handbook of phonological theory*, ed. John Goldsmith, 671–696. Oxford: Blackwell.
- Martin, Andy. 2007. The evolving lexicon. Doctoral Dissertation, UCLA.
- Menn, Lise. 2004. Saving the baby: Making sure that old data survive new theories. In *Constraints in phonological acquisition*, ed. Rene Kager, Joe Pater, and Wim Zonneveld. Cambridge: Cambridge University Press.

- Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25:83–127. URL <http://www.unc.edu/moreton/Papers/MoretonInPressPhonology.pdf>.
- Pater, Joe. 1996. Consequences of constraint ranking. Doctoral Dissertation, McGill University. URL <http://people.umass.edu/pater/dissertation.pdf>.
- Pater, Joe. 2002. Form and substance in phonological development. In *WCCFL 21 Proceedings*, ed. Line Mikkelsen and Christopher Potts, 348–372. Somerville, MA: Cascadilla Press.
- Pater, Joe. 2005. Learning a stratified grammar. In *Proceedings of the 29th boston university conference on language development*, ed. Alejna Brugos, Manuella R. Clark-Cotton, and Seungwan Ha, 482–492. Somerville, MA: Cascadilla.
- Pater, Joe. 2008a. Gradual learning and convergence. *Linguistic Inquiry* 39:334–345.
- Pater, Joe. 2008b. Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution. In *Phonological Argumentation: Essays on Evidence and Motivation*, ed. Steve Parker. London: Equinox. To appear.
- Pater, Joe, Elliott Moreton, and Michael Becker. 2008. Simplicity biases in structured statistical learning. URL <http://people.umass.edu/pater/pater-bucl1d.pdf>, poster presented at BUCLD.
- Pater, Joe, and Anne-Michelle Tessier. 2006. L1 phonotactic knowledge and the l2 acquisition of alternations. In *Inquiries in linguistic development: Studies in honor of lydia white.*, ed. Silvina Montrul Roumyana Slabakova and Philippe Prévost, 115–131. Amsterdam/Philadelphia: John Benjamins Publishing Company. URL R0A-752.
- Payne, Arvilla C. 1980. Factors controlling the acquisition of the philadelphia dialect by out-of-state children. In *Locating language in time and space*, ed. William Labov, 143–178. New York: Academic Press.
- Phillips, Betty S. 1998. Word frequency and lexical diffusion in english stress shifts. In *Historical linguistics 1995*, volume Volume 2: Germanic Linguistics. Amsterdam: Benjamins.
- Prince, Alan, and Paul Smolensky. 1993/2004. Optimality Theory: Constraint interaction in generative grammar. RuCCS Technical Report 2, Rutgers University, Piscataway, NJ: Rutgers University Center for Cognitive Science. Revised version published 2004 by Blackwell. Page references to the 2004 version.
- Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 22. (pp. 423-435). : Cascadilla Press.*, ed. G. Garding and M. Tsujimura, 423–435. Somerville, MA.
- Tesar, Bruce. 1998. An Iterative Strategy for Language Learning. *Lingua* 104:131–145.
- Tesar, Bruce. 2006. Faithful contrastive features in learning. *Cognitive Science* 30:5:863–903. URL <http://rucss.rutgers.edu/tesar/papers/cogsci2006.pdf>.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.
- Thiessen, Eric, and Jenny R. Saffran. 2003. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology* 39:706–716.

- Velleman, Shelley L., and Marilyn M. Vihman. 2002. Whole-word phonology and templates: Trap, bootstrap, or some of each? *Language, Speech and Hearing Services in Schools* 33:9–23.
- Wheeler, Deirdre. 1979. A historical explanation for final stress in tubatulabal. In *Papers presented to emmon bach by his students*, ed. Elisabet Engdahl and Mark Stein. Amherst, MA: GLSA.
- Zuraw, Kie. 2003. Probability in language change. In *Probabilistic linguistics*, ed. Rens Bod, Jennifer Hay, and Stefanie Jannedy, 139–176. Cambridge, MA: MIT Press. URL <http://www.linguistics.ucla.edu/people/zuraw/dnldpprs/ProbLangChange.pdf>.