Symposium: Learning Lexical Specificity in Phonology

Organizers

Stephanie S Shih
University of California, Merced
shih@ucmerced.edu

Claire Moore-Cantwell
University of Connecticut
clairemoorecantwell@gmail.com

The interaction of the phonological grammar with the lexicon is a necessary component in the phonological acquisition process and its end state, since the lexicon shapes and is shaped by phonology at potentially every stage of learning. Newly available empirical tools such as corpus methods, machine learning, and experimental techniques have recently accelerated investigations in many of the long-standing issues of lexicon-phonology interaction. The aims of this symposium are to bring together work that leverages these modern empirical approaches and to situate this work within the broader landscape of phonological theory. The symposium will foster deeper understanding of the long-standing issues in learning lexical specificity in phonology.

The phonological grammar and lexicon share a complex relationship, as illustrated by the numerous phenomena in which phonological behavior exhibits lexical specificity: morphologically-conditioned phonology (Anttila 2002; Inkelas 2014), lexical class-sensitive phonology (Smith 2002, Ito & Mester 1999, Sonderegger & Niyogi 2013), lexical exceptions to phonological patterns (Inkelas et al. 1997; Zuraw 2000), and phonological variation in the lexicon (Becker & Gouskova, to appear). Lexically-specific phonology has over the years significantly influenced the development of morpho-phonological theory: e.g., Lexical Morphology & Phonology and Stratal OT (Kiparsky 1982, 2000), cophonologies (Inkelas & Zoll 2005), indexed constraints (Itô & Mester 1994; Becker 2009; Pater & Coetzee 2005; Pater 2009). The current state of the field presents new challenges in the consideration of lexicon-grammar architecture. Access to natural language quantitative data now allows us to observe not only the empirical extent of lexical specificity across a phonological system but also the push-pull between massive variation and systematicity that exists in natural languages. Developments in experimental and computational methodologies and theories have advanced the study of learning and acquisition (Becker & Tessier 2011; Storkel & Lee 2011), as have developments in understanding psycholinguistic influences on phonology (Cohen Priva 2012; Coetzee & Kawahara 2013).

Now is the time to bring together researchers who have been instrumental in building theories of the lexicon-grammar interface and those taking new approaches to investigating learning of natural language lexical specificity in phonology. The symposium will address the following issues of the interaction between the developing lexicon and the developing grammar, with particular focus on the role of variation in learning lexical specificity:

- When and how does a learner learn lexical specificity?
- How does the learner manage lexical specificity and natural language variation?
- How does lexical sensitivity differ or remain the same for learning alternations and allomorphy versus static lexical phonotactics?
- What are the relevant lexical items and categories for phonology? How specific does lexical specificity have to be--e.g., per lexical item? per lexical category?
- What is the optimal balance in grammatical design between representational efficiency and predictive accuracy and robustness? How is the trade-off between complexity and adequacy managed in grammar and learning of lexically-sensitive phonological patterns?
- How do the developing grammar and developing lexicon interact in learning? How does the lexicon shape the phonological grammar, and how does the grammar shape the lexicon?
- How do features of the lexicon influence the grammar (e.g., lexical frequency)?

1 SYMPOSIUM STRUCTURE

We propose a 3-hour session, with 1 introduction, 4 talks, and 2 discussion sections (2 discussants per section). A 10-minute introduction will overview the current state of the field of phonological learning and lexically-conditioned phonology, to frame the prevailing research questions for the session. Each talk will be 20 minutes + 5 minute open Q&A. While the talks will all address questions that cross-cut issues in learning and lexical-specific phonology, we plan for the session to be split into two halves based on empirical and theoretical focuses: one half on lexical specificity in morphophonological phenomena (e.g., alternations, allomorphy), and the other half on the learning of lexical items and lexical classes over phonotactic phenomena. Discussants will prepare 8-10 minute responses to the two papers in their respective session halves, with the aim of highlighting overarching threads, providing theoretical or empirical context, bringing additional relevant evidence to bear, and seeding discussion towards understanding the learning of lexical specificity in phonology. Each discussion section will be 35 minutes total (two 10-minute responses, 15 minute open discussion).

Introduction	Joe Pater (University of Massachusetts, Amherst)	10 minutes
Part 1. Allomorphy & Alternations		
Talks	Michael Becker (Stony Brook University) *Affix-specificity makes stress learnable*	25 minutes
	Brian Smith (University of California, Santa Cruz) *Using phonotactics to learn affix-specific phonology*	25 minutes
Discussion	Sharon Inkelas (University of California, Berkeley) Kie Zuraw (University of California, Los Angeles)	35 minutes
Part 2. Items & Classes		
Talks	Claire Moore-Cantwell (University of Connecticut) *Concurrent learning of the lexicon and phonology*	25 minutes
	Stephanie S Shih (University of California, Merced) *Learning lexical classes for class-sensitive phonology*	25 minutes
Discussion	Andries Coetzee (University of Michigan) Jen Smith (University of North Carolina at Chapel Hill)	35 minutes

2 TOPICS

The four talks all draw heavily from empirical data (either experimental, computational, or corpus-based) and address facets of lexical specificity in phonology. The first two talks (Becker, Smith) deal with how lexical specificity in phonology interacts with morphological items or processes, and what phonological evidence leads to the learning of lexical specificity in morphophonology. Smith asks how allomorphs are learned (i.e., chosen): given that the behavior of certain allomorphs is unlearnable from exposure to natural language data, Smith argues that phonological conditioning in allomorphy is bootstrapped on information from the phonotactic grammar and from learning idiosyncratic lexeme-specific 'baselines.' Becker asks how alternations are learned and proposes a model of sublexical phonology that relies on paradigmatic phonological comparisons in learning morphological alternations.

The second two talks (Moore-Cantwell, Shih) examine phonological patterns across lexicon, investigating how gradual probabilistic learning balances the developing lexicon alongside the developing grammar. Moore-Cantwell argues that exceptional phonological behavior for lexical items can be learned and then stored in the lexical representation, with dynamic, psycholinguistically-informed effects in the

phonological grammar. Shih addresses how groups of exceptional lexical items--i.e., lexical classes--can be learned efficiently from noisy, highly-variable surface input. While Moore-Cantwell and Shih both utilize versions of lexically-specific Maximum Entropy models, they place the locus of lexical specificity in different parts of the lexicon-grammar interface: the juxtaposition of these two proposals in this symposium will allow for comparison between the similar approaches.
The four talks all include data that cross-cuts well-known lexical-specific phenomena and novel observations (in English, Korean, Turkish, Egyptian, Polish, French, Mende). Using a mix of computational, experimental, and quantitative approaches, the talks together present a fresh, multifaceted view of the acquisition of lexical specificity in phonology.

The discussants for the symposium were selected given their long-standing contributions to understanding lexical specificity in phonology and their expertise in relevant aspects of learning, data, and/or methodology. Inkelas developed Cophonology Theory and also works on pathways of child phonology acquisition. Zuraw's research has addressed numerous phenomena of lexical exceptions in phonology, via computational modeling of corpus and experimental data. Coetzee's work deals with the quantitative modeling of lexical patterns, integrating psycholinguistic principles (e.g., frequency) in the phonological grammar. Smith has extensively studied lexical class-sensitive phonology using typological and experimental data. The symposium chair Pater, who will present the introduction to the state of the field, has significant work in acquisition, computational learning models, and lexical specificity in phonology.

In addition to research relevance, speakers and discussants were also chosen with an eye towards diversity. Two of four speakers and three of five discussants/introducers are women, to encourage equal participation in phonological discussion (see Pater 2016 for recent discussion: http://blogs.umass.edu/phonolist/2016/01/05/the-representation-of-women-in-phonological-discussion/). Speakers are all junior faculty, presenting state-of-the-art computational and experimental work. Discussants are senior faculty, who are optimally situated to contextualize and broaden discussion on lexically-conditioned phonology and learning.

Affix-specificity makes stress learnable

Generative analyses of stress are top-down, posting a uniform grammar and a set of URs for lexical items. Yet learning of these analyses has never scaled to realistic data (Tesar 2006). I propose a demonstrably learnable bottom-up approach, which additionally turn out to be simpler and more local.

Turkish: I offer a sublexical analysis (Becker & Gouskova to appear; Allen & Becker 2015) of Turkish stress, using the generalizations in Inkelas & Orgun (2003). In this approach, an affix and its distribution are learned by fitting separate phonotactic grammars for each surface form of the affix. Some affixes, such as the interrogative [mɯ], have one surface form (ignoring vowel harmony), and attach to all words regardless of the base's stress, e.g. ['penaltɯ-mɯ] 'penalty-INT' (1a).

(1)

		operation	stems
a.	interrogative	suffix [mɯ]	no restriction
b.	plural	suffix [lar]	final stress banned
c.		suffix ['lar], destress stem	final stress required

In contrast, the plural is formed by applying one of two operations: "suffix unstressed [lar]" (1b), e.g. ['penaltɯ-lar] 'penalty-PL', or "suffix stressed ['lar] and destress the stem", e.g. [araba-'lar] 'car-PL' (1c).

The sublexical learner extracts these affixes and operations from unanalyzed word pairs, and then fits a phonotactic grammar to stems that take each surface from of an affix. For the plural suffix, one grammar bans final stress (1b), and the other bans non-final stress (1c). These two grammars compete for influence over novel stems based on their surface stress, extending the generalization productively. All Turkish suffixes fall into one of the two categories in (1), such as the progressive, whose surface forms [ɯjor] and ['ɯjor] follow the same distribution as in (1b-c). The analysis also captures the innermost/leftmost preference of Inkelas & Orgun.

In its current form, the sublexical analysis misses at least one important generalization: words that take a stressed variant of one suffix, also take the stressed variant of all other suffixes, e.g. [ara'ba ∼ araba'lar, araba'sɯ, araba'ja, araba'dan], etc. I discuss the mechanism that is needed to capture this cross-affixal generalization, implemented as a defeasible bias rather than an ironclad rule.

Egyptian Arabic: Hayes (1995) analyzes stress in Egyptian Arabic using inaudible left-to-right trochees, creating a computationally demanding unbounded dependency between the left edge of the word and the stressed syllable (Graf 2010). Adding the exceptional lexical items that weren't covered by Hayes (1995), I provide a more descriptively adequate sublexical analysis that replaces the unbounded dependency with a local three-syllable window. Egyptian stress is entirely predictable when there is a heavy syllable in the window, but when the last three syllables are light, stress is sensitive to morphological structure. The analysis captures this by allowing morpheme-specific stress to hold only when more heavily weighted constraints are inactive.

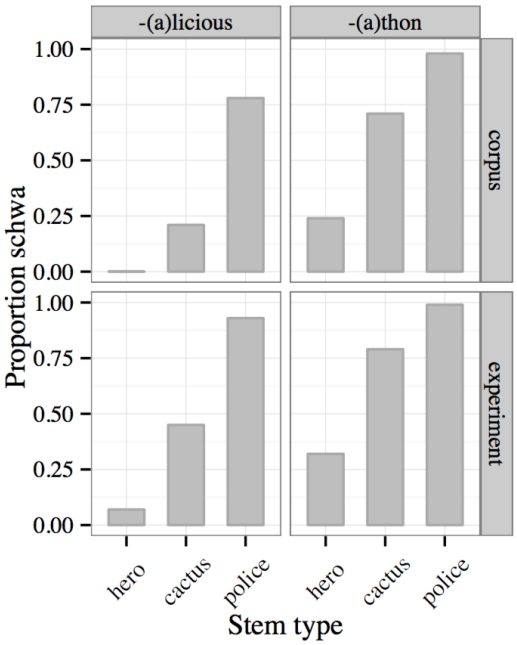
Conclusion: I offer an analysis of two intensely studied stress systems, in both cases focusing on the behavior of exceptional items from a sublexical analysis perspective. I show that these systems are demonstrably learnable from pairs of surface forms, reducing the complexity of the resulting grammars.

Using phonotactics to learn affix-specific phonology

Affixes are often subject to selectional restrictions, selecting for stems of a specific shape. For example, English *–ize* requires stems with non-final stress (rándom-ize *corrúptize; Raffelsiefen 1996). One issue is how these requirements are learned. In this paper, I argue that phonological requirements sometimes don't need to be learned: they come from language-wide phonotactics. I present a case study of English suffixes *-alicious* and *-athon.* These can be learned in two steps: first, the phonotactic grammar is extended to the suffixes; then, affix-specific information, like a suffix's proclivity for schwa, is learned from observed data.

-alicious and -athon. Each of the suffixes has two forms, one with a schwa (-[əlɪʃəs]) and one without (-[lɪʃəs]). Both experimental and corpus data show that the form of the suffix tends to avoid stress clash and hiatus. The graph shows the results of a judgment experiment (2-alternative forced choice) and corpus search (corpus: GlOWbE, Davies 2013). In both, schwa is most likely after final stress C-final stems (*políce*), where it avoids a clash, and schwa is least likely with non-final stress V-final stems (*héro*), where it would create a hiatus. Stems like *cactus* (C-final, no final stress) are in between. In both the corpus and experiment, schwa is more likely in *-athon* than *-alicious.* This holds for 78/80 experimental items.

Lack of learning data. The fact that the experiment and corpus results are similar is surprising given the sparsity of learning data for speakers. In the corpus, many stem types are underrepresented: there are no stems longer than two syllables for *-alicious*, most C-final stems are monosyllabic (only 10% of stems are like *cactus*), and 58% of *-alicious* words only occur once in the 1.9 billion word corpus. Despite this, speakers agree about the use of *-alicious* and *-athon* with underrepresented stems, including trisyllabic stems and stems like *cactus.*



Phonotactics and baselines. The phonological conditioning of *-alicious* and *-athon* closely matches the phonotactics of English. Here I use the English phonotactic grammar of Hayes (2012). In both the phonotactic grammar and the suffixed words, *HIATUS and *CLASH play a role, and the weight of *HIATUS is greater than *CLASH. Using the constraint weights from the phonotactic grammar alone comes close to matching the experimental probabilities, but it has no way to account for the greater probability of schwa in *-athon*. This difference is represented in the grammar by suffix-specific constraints that encode baseline rates of schwa. Unlike the phonological conditioning, the suffixes' baseline likelihood of schwa is learnable from the corpus data: the most common *-alicious* words are schwaless, and the most common *-athon* words have schwa.

Conclusion. This paper argues that the distribution of *-alicious* and *-athon* follows from the combination of the phonotactic grammar and suffix-specific information. This provides a solution to the learning problem, and challenges analyses in which all selection requirements are learned on an affix-by-affix basis.

Word count: 499

Concurrent learning of the lexicon and phonology

During acquisition, the learner must memorize all the words of her language, as well as phonological generalizations over them. These tasks interact: the contents of the lexicon form the empirical basis for the acquisition of phonological grammar, and the contents of the phonological grammar can constrain how words are represented in the lexicon (e.g. Mehler, Dupoux, and Segui (1990)). This interaction becomes especially complex for patterns with exceptions. Fidelholz (1979) points out that patterns with just a few exceptions may be represented differently than patterns with many. I present a computational model in which features of a lexical item (like stress) can be represented with continuously-valued weights. These weights simulate the memory strength associated with each feature of a word. When the features of lexical items are learned concurrently with the phonological grammar, different degrees of exceptionality are represented differently, both in the grammar and in the lexical entries themselves.

Empirically, I focus on the interaction of lexical learning with the acquisition of stress. Peperkamp and Dupoux (2002) and Peperkamp et. al. (2010) argue based on stress 'deafness' facts that learners decide early in acquisition whether to represent stress on lexical items. Using a memory task, they find that speakers of languages with lexically specified stress, (Spanish and English), are much better able to recall stresses of novel words than speakers of languages with exceptionless stress patterns (French and Finnish). In Polish, where exceptions to the stress pattern exist, but are rare, speakers exhibit an intermediate degree of stress deafness. These results demonstrate (1) that the ability to represent stress is gradient, and (2) that it depends on the statistics of the lexicon.

Following Moore-Cantwell and Pater (to appear), and similar to Shih and Inkelas (to appear) and Shih (this symposium), I use lexically specific constraints within a Maximum Entropy Grammar framework to model the concurrent development of the lexicon and the phonological grammar. In the present proposal, the lexically specific constraints are used to simulate the strength in the learner's memory of each lexical item's stress specification. These lexical specifications are induced when an error is made on a particular item, and their weights decay when the lexical item is not used. In a gradual error-driven learning simulation, the weights of stress specifications that are entirely predictable from the grammar decay away, while the stress specifications of exceptional words gradually acquire a high weight. Additionally, the learner's ability to induce new stress specifications is contingent on how many stress specifications it already uses. If no stress is specified (Finnish and French), the probability of specifying stress on a novel lexical item falls to close to 0. If stress is specified on many lexical items (Spanish), or just a few (Polish), the probability of specifying stress on a new lexical item is proportional. This model provides an explicit mechanism for the interaction of the developing lexicon and the developing grammar, and moreover predicts the gradient exceptionality actually observed in stress systems.

Learning lexical classes for class-sensitive phonology

One issue in lexically-sensitive phonology is how classes of lexical items that condition different phonological patterns are learned. In this paper, I present a computational approach based on unsupervised hierarchical clustering and information-theoretic model comparison that arrives at lexical classes from variable, noisy input. The lexical classes learned by the proposed approach are shown to be quantitatively better at capturing lexical class-conditioned surface patterns, as compared to classes assumed by long-standing top-down approaches. Given the results, I argue that utilizing hypothesis comparison and selection methods in conjunction with a probabilistic grammar allows learners to efficiently home in on the most robust generalizations of lexical classification over observed natural language evidence.

Lexical classes that can affect phonology range broadly from morphosyntactically-defined categories (parts-of-speech, lexical versus grammatical words) to etymologically-defined categories (strata; Itô & Mester 1999) to potentially arbitrary categories (Inkelas 2014:19ff). Given the large number of possible lexical classifications, constraining potentially unlimited class explosion is a problem, further amplified by the quantitative variability present within and between classes in natural language (Anttila 1997, 2002). From a gradual learning view, as speakers encounter lexical items, the number of possible classes relevant for phonology—as items, or as classes containing multiple items—increases at a greater-than-exponential rate (Bell 1934): the problem of how lexical classes are induced and learned from natural language surface input becomes computationally intensive.

This study revisits the supposed categorical distinction between English lexical and grammatical word, which demonstrate stress differences: i.e., grammatical words commonly reduce in context while lexical words rarely do (Kiparsky 1982; Selkirk 1984; Inkelas & Zec 1993). Data comes from the reduction of monosyllabic words in a corpus of conversational American English (n=206,858) (Pitt et al. 2007). Reduction was measured by segmental changes (*thing*s [θɪŋz]→[hŋs]; Johnson 2004) and word duration, as proxies for reduction. Without assuming *a priori* lexical classes, unsupervised agglomerative cluster modeling identifies probable hypotheses of lexical classes based on shared behavior in segmental and durational reduction. This computationally simple step allows for efficient dimensionality reduction based solely on surface information. The class hypotheses are tested in Maximum Entropy Harmonic Grammars with lexical class-specific constraints (Pater & Moore-Cantwell, to appear; Shih & Inkelas, to appear). Results show that an optimal lexical classification is obtained via information-theoretic multimodel comparison (Burnham & Anderson 2002) that significantly surpasses traditional English lexical-grammatical classifications in descriptive adequacy (ΔAIC_c=231.4). A gradual version of the learning model using stochastic data sampling also demonstrates a convergence to robust classes after an initial acquisition period of "exploring" potential lexical class hypotheses. Learning efficiency is adjusted by the weighting of lexical class "memory" from previous learning cycles (cf. memory decay in Moore-Cantwell; this symposium). Preliminary evidence from batch-learning for lexically-sensitive Mende tonotactics aligns with these results from English.

The approach here leverages the power of probabilistic grammars to test the efficacy of lexical classes in the learning process for lexically-sensitive phonology in natural language data, with potential ramifications for understanding the development of structure of the lexicon in the grammar acquisition process (Johnson 2016).

Word count: 500