

- Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In *The role of speech perception in phonology*, ed. by Elizabeth V. Hume and Keith Johnson, 219–250. San Diego, CA: Academic Press.
- Steriade, Donca. 2001/2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In *The nature of the word: Studies in honor of Paul Kiparsky*, ed. by Kristin Hanson and Sharon Inkelas, 151–180. Cambridge, MA: MIT Press.
- Torres-Tamarit, Francesc, and Peter Jurgec. 2015. Lapsed derivations: Ternary stress in Harmonic Serialism. *Linguistic Inquiry* 46: 376–387.
- Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
- Wilson, Colin. 2000. Targeted constraints: An approach to contextual neutralization in Optimality Theory. Doctoral dissertation, Johns Hopkins University, Baltimore, MD.
- Wilson, Colin. 2001. Consonant cluster neutralization and targeted constraints. *Phonology* 18:147–197.
- Wolf, Matthew Adam. 2008. Optimal Interleaving: Serial phonology-morphology interaction in a constraint-based model. Doctoral dissertation, University of Massachusetts, Amherst. Rutgers Optimality Archive, ROA 996. <http://roa.rutgers.edu>.
- Yip, Moira. 2004. Phonological markedness and allomorph selection in Zahao. *Language and Linguistics* 5:969–1001.

ACCIDENTAL GAPS AND SURFACE-BASED PHONOTACTIC LEARNING: A CASE STUDY OF SOUTH BOLIVIAN QUECHUA
 Colin Wilson
 Johns Hopkins University
 Gillian Gallagher
 New York University

The lexicon of a natural language does not contain all of the phonological structures that are grammatical. This presents a fundamental challenge to the learner, who must distinguish linguistically significant restrictions from *accidental gaps* (Fischer-Jørgensen 1952, Halle 1962, Chomsky and Halle 1965, Pierrehumbert 1994, Frisch and Zawaydeh 2001, Iverson and Salmons 2005, Gorman 2013, Hayes and White 2013). The severity of the challenge depends on the size of the lexicon (Pierrehumbert 2001), the number of sounds and their frequency distribution (Sigurd 1968, Tambovtsev and Martindale 2007), and the complexity of the generalizations that learners must entertain (Pierrehumbert 1994, Hayes and Wilson 2008, Kager and Pater 2012, Jardine and Heinz 2016).

In this squib, we consider the problem that accidental gaps pose for learning phonotactic grammars stated on a single, surface level of representation. While the monostratal approach to phonology has considerable theoretical and computational appeal (Ellison 1993, Bird

We're grateful to Maria Gouskova and Paul Smolensky for comments on this work, as well as to audiences at the Annual Meeting on Phonology 2016 and SigMorPhon 2016.

and Ellison 1994, Scobbie, Coleman, and Bird 1996, Burzio 2002), little previous research has investigated how purely surface-based phonotactic grammars can be learned from natural lexicons (but cf. Hayes and Wilson 2008, Hayes and White 2013). The empirical basis of our study is the sound pattern of South Bolivian Quechua, with particular focus on the allophonic distribution of high and mid vowels. We show that, in characterizing the vowel distribution, a surface-based analysis must resort to generalizations of greater complexity than are needed in traditional accounts that derive outputs from underlying forms. This exacerbates the learning problem, because complex constraints are more likely to be surface-true by chance (i.e., the structures they prohibit are more likely to be accidentally absent from the lexicon). A comprehensive quantitative analysis of the Quechua lexicon and phonotactic system establishes that many accidental gaps of the relevant complexity level do indeed exist.

We propose that, to overcome this problem, surface-based phonotactic models should have two related properties: they should use distinctive features to state constraints at multiple levels of granularity, and they should select constraints of appropriate granularity by statistical comparison of observed and expected frequency distributions. The central idea is that actual gaps typically belong to statistically robust feature-based classes, whereas accidental gaps are more likely to be featurally isolated and to contain independently rare sounds. A maximum-entropy learning model that incorporates these two properties is shown to be effective at distinguishing systematic and accidental gaps in a whole-language phonotactic analysis of Quechua, outperforming minimally different models that lack features or perform nonstatistical induction.

1 Vowel Height Allophony in Quechua

1.1 *The Pattern and Traditional Analysis*

Descriptively, South Bolivian Quechua (henceforth Quechua; Bills, Troike, and Vallejo 1971, Laime Ajacopa 2007, Gallagher 2016) has three phonemic vowels /i u a/ with allophonic lowering of /i u/ to [e o] in the vicinity of uvulars /q q^h q'/. Mid vowels occur immediately following or preceding a uvular (1a–b), or preceding a uvular across an intervening coda (1c). High vowels occur in all other consonantal environments (2).

- (1) *Uvular contexts:* [e o] *[i u]
- | | | | | |
|----|---------------------------|------------|-----------------------|--------------|
| a. | q'epij (*q'ipij) | 'to carry' | q'ɔɲi (*q'ɔɲi) | 'hot' |
| b. | peqaj (*piqaj) | 'to grind' | noqa (*nuqa) | 'I' |
| c. | wesq'aj (*wisq'aj) | 'to close' | toɬqa (*tuɬqa) | 'son-in-law' |
- (2) *Elsewhere:* [i u] *[e o]
- | | | | | |
|--|---------------------|-------|-----------------------|----------------|
| | misi (*mese) | 'cat' | kuɬku (*koɬko) | 'type of bird' |
|--|---------------------|-------|-----------------------|----------------|

A traditional analysis of this pattern would assign the high vowels *elsewhere* status and specify the contexts in which mid vowels occur.

In an *SPE*-style account (Chomsky and Halle 1968), only high vowels would appear in underlying representations and mid vowels would be derived by rule. In Optimality Theory/Harmonic Grammar (OT/HG), the complementary distribution would follow from subordinating a general constraint against mid vowels (*E) to specific constraints against high vowels before or after a uvular (3). (Natural classes are abbreviated throughout as follows: K = velar consonant, Q = uvular consonant, I = high vowel, E = mid vowel, V = any vowel. The constraints in (3) apply on the dorsal tier, as specified in table 1 below.)

| (3) | /piqaj/ | *QI, *IQ | *E | Id[high] | /mesi/ | *QI, *IQ | *E | Id[high] |
|-----|---------|----------|----|----------|--------|----------|----|----------|
| | piqaj | *! | | | → misi | | | * |
| | → peqaj | | * | * | mesi | | *! | |

1.2 Analysis with Surface-Based Constraints

The Quechua distribution of high and mid vowels, despite being a typologically unremarkable instance of allophonic variation, is more difficult to analyze in monostratal phonology. The restrictions on high vowels can remain as above (*QI, *IQ). But an inviolable surface constraint *E would be overly restrictive, as mid vowels do occur in grammatical forms. Even in a surface-based model with violable constraints, *E would still be inadequate: it would penalize all surface mid vowels equally and thus miss the crucial difference between (for example) grammatical [qena] and ungrammatical *[pena].

In place of *E, there must be several surface constraints against mid vowels in the exhaustive set of “nonuvular” environments. We provide an example analysis in (4), with restrictions again stated on the dorsal tier (which contains velar and uvular consonants along with all vowels). This tier accounts for the segmentally nonlocal interaction between a uvular and a nonhigh vowel across an intervening coda (e.g., [orqo] ‘mountain’) and also has the welcome side effect of reducing the number of “nonuvular” contexts that must be enumerated.

- (4) a. *Surface-true constraints on high vowels in uvular contexts (as in (3))*

*Q I *I Q

- b. *Surface-true constraints on mid vowels in “nonuvular” contexts*

*# E K *K E K *V E K

*# E # *K E # *V E #

*# E V *K E V *V E V

Trigram constraints like those of (4b) are necessary because a mid vowel can be licensed by a uvular on either side. To exclude mid vowels in unlicensed contexts only, the constraints must ensure that something other than a uvular both precedes and follows.

1.3 Accidental Gaps in Quechua

While the preceding analysis may be judged inelegant, our main concern is with the consequences it has for learning. Previous research on surface-based phonotactic induction has primarily limited constraints to a maximum length of two (i.e., unigrams or bigrams; Ellison 1993, Adriaans and Kager 2010, Jardine and Heinz 2016, Jardine and McMullin 2017; cf. Hayes and Wilson 2008).¹ However, trigram constraints are required to analyze patterns in which a sound can be conditioned by context on either side, as we have just shown for Quechua, and in the analysis of both-side conditioning such as intervocalic voicing and lenition. The general expectation is that permitting constraints of greater complexity will make the problem of distinguishing systematic and accidental gaps more challenging, as complex structures are more likely to be unattested by chance.

To quantify the problem for Quechua, we constructed an exhaustive list of hypothetical CV(C)CV(C) forms (>560,000 items) for which the position-specific segments and medial consonant clusters are attested in roots. We then divided this list of hypothetical roots into those that satisfy the known phonotactic generalizations given in table 1 and those that violate one or more generalizations. (For further discussion of the phonotactic restrictions in Quechua, and of the laryngeal cooccurrence restrictions in particular, see MacEachern 1999 and Gallagher 2011, 2016.) We compared the tier-based trigrams that occur in the list of legal hypothetical roots with those in a lexicon of 1,104 actual roots compiled from the Laime Ajacopa (2007) dictionary and

Table 1
Quechua tiers and phonotactic generalizations

| Tier | Projected segments | Phonotactic generalizations |
|---------------------------------|---|---|
| Dorsal C-dorsal ^a | Dorsal consonants, vowels Dorsal consonants, + | High-mid vowel allophony *K . . . Q, *Q . . . K (within morphemes) |
| Laryngeal Segmental | Stops, affricates, h, ? All, + | Laryngeal cooccurrence restrictions *VV, *CCC, *wu, *wo |

^a The C-dorsal tier contains a morpheme boundary symbol (+), allowing the model to represent the fact that the restriction on uvular and velar consonant cooccurrence holds within morphemes but not across them.

¹ Many additional computational approaches to phonotactics, speech segmentation, and phoneme learning have also in practice limited generalizations to unigrams and bigrams (Vitevitch and Luce 2004, Peperkamp et al. 2006, Kirby and Yu 2007, Albright 2009, Adriaans and Kager 2010, Heinz 2010, Daland and Pierrehumbert 2011, Kempton and Moore 2014, Calamaro and Jarosz 2015; cf. Martin, Peperkamp, and Dupoux 2013) and would have to be extended to account for Quechua. See Pierrehumbert 1994 and Kager and Pater 2012 for evidence of speaker knowledge of complex phonotactic restrictions.

verified with a native speaker.² On the segmental tier, there are 2,966 unique trigram sequences in the hypothetical roots, but only 1,472 (49%) of them are attested in the root lexicon. That is, on the segmental tier there are about as many trigram accidental gaps as there are attested trigrams. The ratio of attested to legal trigrams is higher on the other tiers, but several accidental gaps nevertheless exist for each one (dorsal tier: 193 attested / 204 legal; C-dorsal tier: 17 / 19; laryngeal tier: 167 / 176).

The unattested legal trigrams typically contain rare segments. For example, the sequence [eq^ho] is legal but unattested, reflecting the fact that aspirated dorsals are infrequent in medial position generally and that [e] is the least common surface vowel in the language. However, not all unattested sequences containing rare parts are accidental gaps. For example, [k^h] is among the rarest segments in Quechua, and the sequence [k^hek] has zero frequency like [eq^ho], but in this case the gap is principled (see section 2.1). Considerations of this sort indicate that a surface-based phonotactic learner should induce constraints on a statistical basis, so that it can avoid penalizing sequences like [eq^ho] that are unlikely to occur simply by chance, and that the learned constraints should be stated with features, so that sequences like [k^hek] can be brought under more general constraints (e.g., *KEK) rather than being ruled out individually.

2 Surface-Based Phonotactic Learning Models

2.1 Feature-Based Statistical Model

The maximum-entropy phonotactic model developed by Hayes and Wilson (2008) has both of the properties that we have identified as important for overcoming the problem of accidental gaps. Constraints in the model are stated with features and can range from segment-specific (e.g., “No [ʔ] in noninitial position”) to very broad (e.g., “No initial vowels”). Among other criteria, the model prefers to induce new constraints that have low ratios of observed to expected violation frequencies (where expected violations are statistical quantities determined by the current grammar of weighted constraints).

This model has supported attempts to learn the entire phonotactic pattern of a natural language (i.e., Hayes and Wilson 2008 on Wargamay and Hayes and White 2013 on English). However, the attempts

² While this is a relatively small number of forms, there is reason to believe that it provides a fair representation of the range of phonotactic possibilities that Quechua learners would encounter. Because the root list is derived from a dictionary, it includes many rare words that may provide the single instance of a particular sequence. A corpus of ~10,000 forms, recently compiled by one of the authors from the newspaper *Conosur Ñawpaqman* (<http://www.cenda.org/periodico-conosur>), had substantially less phonotactic diversity than the small root list. For example, the cluster [ʃtʃ^h] is attested in a single root in the dictionary list, but this root is absent from the newspaper corpus.

have been only partly successful: constraints corresponding to previous phonotactic descriptions were learned, but the induced grammars also included many constraints that penalize accidental gaps. In pursuit of the present goal of learning all and only the phonotactic restrictions of Quechua, we modified the original model in two ways.

(i) *Initialization.* Grammars were initialized with a separate violable constraint for each segment in the Quechua inventory. When appropriately weighted, these constraints are equivalent to a unigram stochastic model of the learning data. The general purpose of this kind of initialization is to prevent the model from learning complex constraints against sequences that contain rare segments (e.g., as in the [eq^ho] example above).

(ii) *Gain-based constraint selection.* A grammar was induced one constraint at a time by calculating the *gain* of each surface-true constraint. The gain of constraint C is proportional to the highest log probability (of the learning data) that could be obtained by adding C to the grammar while holding all other constraints and their weights fixed (see Della Pietra, Della Pietra, and Lafferty 1997). This criterion favors constraints that are violated substantially less often in the data than their expected values would predict: the current grammar assigns too much probability to structures that violate such constraints, probability that could be profitably reallocated to attested structures. On each round of constraint selection, the constraint on each tier with highest gain above a fixed threshold γ was added to the grammar (we evaluated various thresholds and report results for $\gamma = 100.0$). Learning halted when no constraint had sufficient gain.³

For the purpose of comparison to alternatives below, we refer to this model as *Maxent-Ftr*. In principle, the model could induce gradient phonotactic grammars containing violable constraints. However, the Quechua phonotactics described earlier are categorical, and one of the alternative models discussed below is inherently nongradient; therefore, we required all of the constraints induced by *Maxent-Ftr* to be surface-true.

2.2 Alternative Models

The first alternative we considered, called *Maxent-Seg*, is identical to *Maxent-Ftr* except that constraints are stated over segments. While features are traditionally used in phonotactic descriptions, some recent models eschew them (e.g., Heinz 2010, Heinz and Rogers 2010; cf.

³ Application of the gain threshold is related to L1 regularization (e.g., Perkins, Lacker, and Theiler 2003) and to the cost of constraints in MDL models (e.g., Rasin and Katzir 2016), because it penalizes the addition of a new constraint regardless of its weight. We also included a term that penalizes large weights, $\lambda \sum_i w_i^2$, but set λ equal to a small constant ($1.0e-5$) that made this penalty negligible.

Heinz and Koirala 2010).⁴ The comparison of Maxent-Ftr and Maxent-Seg provides a close examination of how learning is facilitated by allowing constraints to refer to segment classes. Maxent-Seg was initialized in the same way as Maxent-Ftr and used the same statistical criterion for inducing surface-true constraints.

The second alternative is a nonstatistical version of Maxent-Seg, referred to as *Memory-Seg*, that was inspired by recent formal language research (Heinz, Rawal, and Tanner 2011, Jardine and Heinz 2016, McMullin 2016, Jardine and McMullin 2017; see also de la Higuera 2010). While our presentation of the model draws upon that work, the research question addressed here is quite different. We are interested in the grammars that models learn from natural “gappy” data—not in the important but distinct question of what is provably learnable from hypothetical data in which all legal structures are exemplified.

In the Memory-Seg model, a grammar is defined as a set G_t of legal substrings for each tier t . At the onset of learning, each G_t is empty. The sets are then updated with the substrings of forms that are encountered during learning. For example, the form [toʎqa] is represented as [oqa] on the dorsal tier and the substrings of length 1 to 3 that it contributes to G_{dorsal} are [o], [q], [a] (omitting word boundaries); [#o], [oq], [qa], [a#]; and [#oq], [oqa], [qa#]. In essence, learning involves memorizing the segment sequences that are observed on each tier. This requires significantly less computation than in Maxent-Ftr or Maxent-Seg, because expected frequencies need not be calculated. The one free parameter of the model is the maximum subsequence length (n), directly parallel to the maximum constraint length in the two Maxent models.

A form is ungrammatical with respect to a Memory-Seg grammar iff it contains at least one substring of length $1 \leq m \leq n$, on at least one tier t , that is not in G_t . For example, a grammar learned from the Quechua lexicon with $n = 3$ would not contain the substring [#ei] on the dorsal tier and would therefore correctly identify *[mesi] as illegal.

What about a third logically possible alternative, a nonstatistical model like Memory-Seg that learns by memorizing feature sequences (i.e., Memory-Ftr)? The immediate problem confronting such a model is that any given segment sequence has multiple different featural representations. For example, the attested dorsal-tier trigram [oqa] could be represented with very general classes (e.g., [+syll][−syll] [+syll] = VCV), with maximally specific classes (i.e., [+syll, −high, −low, +back][−cont, −son, +dorsal, −high, −cg][+syll, −high, +low] = [oqa]), or at intermediate levels of granularity (e.g., [+syll, −high, −low][−cont, −son, +dorsal, −high][+syll, −high, +low] = EQA).

⁴ Segmental n -gram models are also commonplace in natural language processing (e.g., Jurafsky and Martin 2000), but are typically limited to one segmental tier and contiguous sequences (but cf. Ron, Singer, and Tishby 1996).

If hypothetical Memory-Ftr judged a substring to be legal as long as it satisfied *any* attested featural description, it would tolerate (among other structures) every VCV trigram and thus massively overgeneralize. If the model instead required *all* feature representations of a substring to be attested, it would be equivalent to Memory-Seg (as segments correspond to singleton classes). Lacking a method for deciding *which* representations are relevant for assessing well-formedness—precisely the role played by statistics in Maxent-Ftr—learning in Memory-Ftr is doomed.

3 Results

We evaluated Maxent-Ftr, Maxent-Seg, and Memory-Seg with five-fold cross-validation (e.g., Hastie, Tibshirani, and Friedman 2001, Mohri, Rostamizadeh, and Talwalkar 2012). The complete learning data consisted of the Quechua root lexicon (section 1.3), as well as forms derived from the roots by adding phonologically representative suffixes (*-nku* ‘3pl present’, *-spa* ‘gerund’, *-rqa* ‘3sg past’) and applying vowel lowering when appropriate. The lexicon was divided into five parts, or folds, of roughly equal size (approx. 870 forms). Each fold served as a set of legal held-out test forms for models trained on the combination of the other four folds. Testing also included the exhaustive set of CV(C)CV(C) nonce roots discussed earlier (section 1.3), each categorized as legal or illegal according to table 1.

The models were provided with the same tiers and allowed to learn generalizations up to length 3.⁵ For Maxent-Ftr/-Seg, a test form was grammatical iff it satisfied all of the learned constraints. For Memory-Seg, grammaticality was determined as described above. Table 2 shows the proportion of test forms judged grammatical by each model. The models performed comparably on attested but held-out forms. Maxent-Ftr generalized far more successfully to nonce roots, ruling out essentially all of the illegal forms and accepting the great majority

Table 2

Proportion of test forms judged grammatical by each model

| | Held-out forms | Legal nonce roots | Illegal nonce roots |
|------------|----------------|-------------------|---------------------|
| Maxent-Ftr | 99.8% | 82.2% | 1.9% |
| Maxent-Seg | 99.7% | 71.5% | 45.4% |
| Memory-Seg | 96.7% | 18.8% | 0.1% |

⁵ In principle, the learner could discover both the need for tiers and their contents (Goldsmith and Riggle 2012, Jardine and Heinz 2016, Jardine and McMullin 2017), though we anticipate that tier induction mechanisms will also be highly sensitive to accidental gaps in the data. We are unaware of attested phonotactic generalizations in Quechua or other languages that exceed the trigram limit adopted here.

of the legals. Nearly all (96%) of the false negatives made by this model involved forms with root-final consonants ([n s r w j x]), which occur rarely in the lexicon and may in fact be phonotactically marginal.⁶ Maxent-Seg both undergenerated and massively overgenerated (e.g., it accepted illegal forms such as *[p'ap'a] that violate laryngeal cooccurrence restrictions). This comparison indicates that statistical calculations alone, in the absence of features or the classes they define, do not suffice for phonotactic learning. Finally, the Memory-Seg grammars showed extreme undergeneralization: by judging forms as ungrammatical if they contained accidental gaps such as [eq^ho], these grammars accepted less than 20% of the novel legal roots.

The Maxent-Ftr model largely refrained from inducing constraints against [eq^ho] and other gaps because the relevant segments are rare (i.e., the expected violations of segment-specific constraints are too small) and because more general constraints (e.g., *EQE) are violated by many attested forms (i.e., their observed violations are too large). The inclusion of featural representations allows the model to generalize the patterning of more frequent [q q'] to less frequent [q^h] and the patterning of more frequent [o] to less frequent [e]. Stating generalizations over the classes of “uvular consonants” and “mid vowels” is necessary to avoid overgeneralization and accurately capture the phonotactics of the language beyond the strictly attested trigram sequences.

4 Conclusion

Most research on phonotactic learning has not thoroughly considered the problem of accidental gaps. Previous work that does address the problem has focused on specific constraints of English (e.g., Pierrehumbert 1994, Gorman 2013, Hayes and White 2013) or achieved limited results on other languages (Hayes and Wilson 2008). In our case study of Quechua, we have shown that a surface-based analysis of the phonotactic pattern must include several trigram constraints and that, when analyzed at the level of segmental trigrams, the lexicon is full of accidental gaps. Because the phonotactic distribution of Quechua is typologically unexceptional, we anticipate that quantitative comparison of attested and legal sound structures in other languages will yield similar results.⁷

We have further shown that learning the phonotactic pattern of Quechua is within reach of a model that employs statistical computations to induce constraints over featural representations. A minimally

⁶ Word-final consonants are frequent in the language as a whole, but they are rare in roots and in our learning set (which did not include any consonant-final suffixes).

⁷ Indeed, the “sparse data” problem addressed here becomes more severe, and may arise even for bigram constraints, as more detail is represented in surface forms. For Quechua, we transcribed allophonic vowel height, but not the tense/lax vowel distinction conditioned by syllable structure nor the variable

different segmental model overgeneralized, because some parochial constraints were too statistically weak to meet the induction criterion. A model that forgoes both statistics and features had the opposite problem: ignoring the possibility that some sequences will be absent by chance, it incorrectly penalized all unattested trigrams. The stark differences in model performance underscore the value of studying phonotactic learning from inevitably sparse natural language data.

While features are a traditional component of phonological analysis, and the granularity and statistical character of phonotactic restrictions have been discussed previously (Coleman and Pierrehumbert 1997, Pierrehumbert 2001, 2003, Moreton and Pater 2012), our results provide support for the claim that features and statistics are jointly required for phonotactic learning (Hayes and Wilson 2008, Albright 2009, Adriaans and Kager 2010). Together with other representational devices, such as identity relations (Berent et al. 2012), features allow multiple segment-level restrictions to be combined into single phonotactic generalizations. For example, the similar patterning of [q], [q'], and [q^h] in the sound pattern of Quechua can be captured with a class that includes these sounds while excluding other stops and other dorsals. This is not only a matter of economy or elegance: feature-based generalizations can have statistical properties that their individual segment-specific instantiations do not share. The segment [q^h] may be too rare to support strong generalizations on its own, but its phonotactic behavior can be analyzed and learned in terms of the active classes to which it belongs (i.e., uvular stops, aspirated obstruents, etc.).

All of the present results have been obtained within a monostratal theory of phonotactics. This approach has a parsimonious architecture and a straightforward computational implementation. In particular, a probability distribution over the set of grammatical forms can be represented by a weighted finite-state machine that provides for efficient computation of expected constraint violations (Eisner 2002). While related implementations are possible for rule-based systems (e.g., Bird and Ellison 1994, Kaplan and Kay 1994), there is no known method for converting an OT/HG grammar into a machine that explicitly represents the set of well-formed outputs or a distribution over them. This places a practical limit on the statistical computations that can be employed for learning in derivational constraint-based theories (but cf. Cotterell, Peng, and Eisner 2015). However, we have demonstrated that monostratal theories have the disadvantage that certain typologically common patterns, such as the allophonic variation of Quechua

production of uvular stops (i.e., /q/ may be [q], [ɬ], or [ɕ]; /q^h/ may be [q^h] or [χ]; /q'/ may be [q'] or [ç]). Transcribing all such distinctions would increase the number of possible surface sequences and compound the problem of distinguishing linguistically significant and accidental gaps (e.g., Martin, Pepercamp, and Dupoux 2013).

vowels, require analyses of greater complexity than in two-level theories. The development of phonotactic models should build on the computational strengths of single-level theories while striving for the analytic simplicity that is captured in models with input-output mappings.

References

- Adriaans, Frans, and René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62:311–331.
- Albright, Adam. 2009. Feature-based generalization as a source of gradient acceptability. *Phonology* 26:9–41.
- Berent, Iris, Colin Wilson, Gary F. Marcus, and Douglas K. Bemis. 2012. On the role of variables in phonology: Remarks on Hayes and Wilson 2008. *Linguistic Inquiry* 43:97–119.
- Bills, Garland, Rudolph C. Troike, and Bernardo Vallejo. 1971. *An introduction to spoken Bolivian Quechua*. Austin: University of Texas Press.
- Bird, Steven, and T. Mark Ellison. 1994. One-level phonology: Auto-segmental representations and rules as finite automata. *Computational Linguistics* 20:55–90.
- Burzio, Luigi. 2002. Surface-to-surface morphology: When your representations turn into constraints. In *Many morphologies*, ed. by Paul Boucher, 142–177. Somerville, MA: Cascadilla Press.
- Calamaro, Shira, and Gaja Jarosz. 2015. Learning general phonological rules from distributional information: A computational model. *Cognitive Science* 39:647–666.
- Chomsky, Noam, and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1:97–138.
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Coleman, John, and Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational phonology: The third meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics.
- Cotterell, Ryan, Nanyang Peng, and Jason Eisner. 2015. Modeling word forms using latent underlying morphs and phonology. *Transactions of the Association for Computational Linguistics* 3:433–447.
- Daland, Robert, and Janet Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 35:119–155.
- de la Higuera, Colin. 2010. *Grammatical inference: Learning automata and grammars*. Cambridge: Cambridge University Press.
- Della Pietra, Stephen, Vincent Della Pietra, and John Lafferty. 1997. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19:380–393.
- Eisner, Jason. 2002. Parameter estimation for probabilistic finite-state transducers. In *Proceedings of the 40th Annual Meeting of the*

- Association for Computational Linguistics*, 1–8. Association for Computational Linguistics.
- Ellison, T. Mark. 1993. The machine learning of phonological structure. Doctoral dissertation, University of Western Australia.
- Fischer-Jørgensen, Eli. 1952. On the definition of phoneme categories on a distributional basis. *Acta Linguistica* 7:8–39.
- Frisch, Stefan A., and Bushra A. Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77:91–106.
- Gallagher, Gillian. 2011. Acoustic and articulatory features in phonology—the case for [long VOT]. *The Linguistic Review* 28: 281–313.
- Gallagher, Gillian. 2016. Vowel height allophony and dorsal place contrasts in Cochabamba Quechua. *Phonetica* 73:101–119.
- Goldsmith, John, and Jason Riggle. 2012. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory* 30:859–896.
- Gorman, Kyle. 2013. Generative phonotactics. Doctoral dissertation, University of Pennsylvania, Philadelphia.
- Halle, Morris. 1962. Phonology in generative grammar. *Word* 18: 54–72.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. *The elements of statistical learning*. New York: Springer.
- Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44:45–75.
- Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.
- Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.
- Heinz, Jeffrey, and Cesar Koirala. 2010. Maximum likelihood estimation of feature based distributions. In *Proceedings of the 11th Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 28–37. Association for Computational Linguistics.
- Heinz, Jeffrey, Cheten Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers*, 2:58–64. Association for Computational Linguistics.
- Heinz, Jeffrey, and James Rogers. 2010. Estimating strictly piecewise distributions. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 886–896. Association for Computational Linguistics.
- Iverson, Gregory K., and Joseph C. Salmons. 2005. Filling the gap. *Journal of English Linguistics* 33:207–221.
- Jardine, Adam, and Jeffrey Heinz. 2016. Learning tier-based strictly 2-local languages. *Transactions of the Association for Computational Linguistics* 4:87–98.

- Jardine, Adam, and Kevin McMullin. 2017. Efficient learning of tier-based strictly k-local languages. In *Language and Automata Theory and Applications: 11th International Conference*, ed. by Frank Drewes, Carlos Martín-Vide, and Bianca Truthe, 64–76. Cham: Springer.
- Jurafsky, Daniel, and James Martin. 2000. *Speech and language processing*. Upper Saddle River, NJ: Prentice Hall.
- Kager, René, and Joe Pater. 2012. Phonotactics as phonology: Knowledge of a complex restriction in Dutch. *Phonology* 29:81–111.
- Kaplan, Ronald M., and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20:331–378.
- Kempton, Timothy, and Roger Moore. 2014. Discovering the phoneme inventory of an unwritten language: A machine-assisted approach. *Speech Communication* 56:152–166.
- Kirby, James, and Alan Yu. 2007. Lexical and phonotactic effects on wordlikeness judgments in Cantonese. In *Proceedings of the Sixteenth International Congress of Phonetic Sciences*, 1389–1392. Saarbrücken: Universität des Saarlandes.
- Laime Ajacopa, Teofilo. 2007. *Diccionario Bilingüe, Iskay Simipi Yuyayk'anacha: Quechua–Castellano Castellano–Quechua*. La Paz, Bolivia.
- MacEachern, Margaret R. 1999. *Laryngeal cooccurrence restrictions*. New York: Garland.
- Martin, Andrew, Sharon Peperkamp, and Emmanuel Dupoux. 2013. Learning phonemes with a proto-lexicon. *Cognitive Science* 37: 103–124.
- McMullin, Kevin. 2016. Tier-based locality in long-distance phonotactics: Learnability and typology. Doctoral dissertation, University of British Columbia.
- Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. 2012. *Foundations of machine learning*. Cambridge, MA: MIT Press.
- Moreton, Elliott, and Joe Pater. 2012. Structure and substance in artificial-phonology learning. Part I: Structure. *Language and Linguistics Compass* 6:686–701.
- Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal, and Emmanuel Dupoux. 2006. The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition* 101: B31–B41.
- Perkins, Simon, Kevin Lacker, and James Theiler. 2003. Grafting: Fast, incremental feature selection by gradient descent in function space. *Journal of Machine Learning Research* 3: 1333–1356.
- Pierrehumbert, Janet. 1994. Syllable structure and word structure: A study of triconsonantal clusters in English. In *Phonological structure and phonetic form: Papers in laboratory phonology III*, ed. by Patricia A. Keating, 168–188. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet. 2001. Why phonological constraints are so coarse-grained. *Language and Cognitive Processes* 16:691–698.

- Pierrehumbert, Janet. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46:115–154.
- Rasin, Ezer, and Roni Katzir. 2016. On evaluation metrics in Optimality Theory. *Linguistic Inquiry* 47:235–282.
- Ron, Dana, Yoram Singer, and Naftali Tishby. 1996. The power of amnesia: Learning probabilistic automata with variable memory length. *Machine Learning* 25:117–149.
- Scobbie, James, John Coleman, and Steven Bird. 1996. Key aspects of declarative phonology. In *Current trends in phonology: Models and methods*, ed. by Jacques Durand and Bernard Laks, 2:685–709. Salford, Manchester: ESRI.
- Sigurd, Bengt. 1968. Rank order of consonants established by distributional criteria. *Studia Linguistica* 9:8–20.
- Tambovtsev, Yuri, and Colin Martindale. 2007. Phoneme frequencies follow a Yule distribution. *SKASE Journal of Theoretical Linguistics* 4:1–11.
- Vitevitch, Michael S., and Paul A. Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36:481–487.