# Arriving at Hidden Structure

Mutually Reinforcing Syllabification and Stress Learning

Kaustubh Ghoshal, New York University

Annual Meeting on Phonology, October 9-11, 2026

## 1 The problem

**Syllable weight affects stress position, and what counts as heavy varies across languages.**

(1) CVX heavy: Kabardian (Gordon, 2017)
Final stress if CVV or CVC, else penult

| | | |
|---|---|---|
| a. | tɐpʃə́g | 'plate' |
| b. | saːbíː | 'baby' |
| c. | máʃɐ | 'bear' |
| d. | χɐrzənɐ | 'good' |

(2) CVV heavy: Huallaga Quechua (Weber, 1983)
Final stress only if CVV

| | | |
|---|---|---|
| a. | ajwaʃáː | 'bear' |
| b. | ʧajʌaráː | 'just a moment ago' |
| c. | ajwájkaŋ | 'they are going' |
| d. | manáraχ | 'not yet' |

**Ambiguity in surface forms: Hidden structure.** The same surface form [pámpa] is consistent with three different structures, leading to different grammars:

| | | | | |
|---|---|---|---|---|
| (3) | a. | [pám$_{\mu}$.pa] | WSP ≫ AlignL/R[1] | (CVC is heavy) |
| | b. | [pám.pa] | AlignL ≫ AlignR | (left-aligned, weight unknown) |
| | c. | [pá.mpa] | NoCoda, AlignL ≫ AlignR | (left aligned, no coda) |

- Learner only sees the surface form [pámpa], does not know which of these grammars is correct.
- The learner needs both the *syllable parse* (where the consonant goes) and the *weight type* (whether a coda counts for syllable weight).

**Circularity problem**: Learning stress requires syllabification and weight type; but these lack acoustic cues and in turn require stress.

- **Syllabification**: general sonority effects (Clements, 1990) but no reliable acoustic correlates (Bosch, 2011; Fokes and Bond, 1993); speaker judgments vary (Derwing, 1992; Macaulay, 2024; Steriade, 1999; Treiman and Danis, 1988).
- **Weight type**: unclear acoustic correlates; claims rest on one to three speakers (Broselow et al., 1997; Gordon, 2002) with findings contested (Kelly, 2021). Few acoustic studies on weight-sensitive languages (cf. Gordon and Roettger, 2017).

**The Credit Problem** (Dresher, 1999, p. 28): "When there is a mismatch between a target form and a learner's grammar, there is no way of reliably knowing which parameters/constraints must be reset to yield a correct output."
**Hidden structure** (Tesar and Smolensky, 1998, 2000) is an instantiation of the Credit Problem (Jarosz, 2019; Nazarov and Jarosz, 2021).

[1] Weight-to-Stress (WSP): Every heavy syllable must be stressed. AlignL/R: Align the stressed syllable to the left/right edge of the prosodic word.

**The gap in previous work.**

- Stress and weight learners take syllabification as given (Boersma and Pater, 2016; Dresher and Kaye, 1990; Hayes and Wilson, 2008; Heinz, 2009; Nazarov, 2016; Nazarov and Jarosz, 2017, 2021; Pater and Prickett, 2022; Staubs, 2014; Tesar and Smolensky, 2000).
- Syllabification learners model segmental phonotactics without stress (Goldwater and Johnson, 2005; Jarosz, 2006; Tesar and Smolensky, 2000).
- The gap is stated explicitly in the literature: "the problem of hidden structure in other domains, such as syllabification, remains to be dealt with" (Jarosz, 2013, 38, fn. 3); see also Apoussidou (2007), Daland et al. (2011), Hayes and Wilson (2008), Jarosz (2019), and Lee et al. (2024).
- Hayes and Wilson (2008) predict that Maxent can be used for hidden structure learning: "the correct hidden structure will yield a tighter phonotactic characterization, which increases the probability of the learning data, a measurable quantity under the maxent approach."

**Hidden structure as parameters, stress grammar by induction.**

- Much work on hidden structure learning in phonology uses fixed parameters or constraints to model the hypothesis space (Dresher, 1999; Dresher and Kaye, 1990; Jarosz, 2006, 2015; Lee et al., 2024; Nazarov and Jarosz, 2021; Pater and Prickett, 2022; Pater et al., 2012; Tesar and Smolensky, 2000).
- Here, syllabification and weight are modeled using parameters, and stress position is left to an $n$-gram constraint grammar built and fit on the data (Hayes and Wilson, 2008; Staubs, 2014), following the hybrid UG+statistics blueprint in Pearl, 2021.

▷ **Current proposal:** Syllabification and weight are selected by the fit of the stress grammar.

*Mutually reinforcing*: The stress grammar itself is induced from the data based on syllabification and weight settings.

## 2 How the learner works: step by step

**Input:** unsyllabified segment strings, one stress mark per word, and a sonority value for each segment. The words below are from Petrarki (Result 1), an artificial language.

(4) **Generating rule (RDTS):** stress the rightmost heavy (CVX) syllable, else the final syllable.
**Generating syllabification (or):** obstruent+$r$ is an onset (V.TRV); $r$+obstruent is split (VR.TV).

The learner tries every combination $\Theta$ of syllabification and weight settings (27 in all; see Appendix A).

(5) Consonant parameters (three settings: (Max Onset, Onset Rise) = (0,1), (1,1), (1,0))
   a. Max Onset: prefer a complex onset over a coda. (Itô, 1988)
   Max Onset=1: [a.tra] Max Onset=0: [at.ra]
   b. Onset Rise: complex onsets must rise in sonority. (Clements, 1990)
   Onset Rise=1: [ar.ta] Onset Rise=0: [a.rta]

(6) Weight parameters (three settings: none, CVV, CVX) (Dresher and Kaye, 1990; McCarthy, 1979)
   a. CVV heavy: syllables with a long vowel are heavy.
   b. CVC heavy: closed syllables are also heavy (only together with CVV heavy).

**Step 1: each setting parses the same strings differently.** Taking all stress candidates of training items (one stress per word), different settings feed different parses of the stress template. The heavy tier filters to stress values of heavy syllables only (here = CVX).

| | | **Correct: Θ = (1,1), CVX** | | | | **Wrong: Θ = (0,1), CVX** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Parse | Weight | Stress tier0 | Heavy tier | Parse | Weight | Stress tier0 | Heavy tier |
| ✓ | kakatrí | ka.ka.trí | L L L | #0 0 1# | ## | ka.kat.rí | L H L | #0 0 1# | #0# |
| | kakátri | ka.ká.tri | L L L | #0 1 0# | ## | ka.kát.ri | L H L | #0 1 0# | #1# |
| ✓ | kapárto | ka.pár.to | L H L | #0 1 0# | #1# | ka.pár.to | L H L | #0 1 0# | #1# |
| | kapartó | ka.par.tó | L H L | #0 0 1# | #0# | ka.par.tó | L H L | #0 0 1# | #0# |
| ✓ | kakató | ka.ka.tó | L L L | #0 0 1# | ## | ka.ka.tó | L L L | #0 0 1# | ## |

Table 1: Syllabification changes the stress template (`0` unstressed, `1` stressed, # word boundary)

**Correct parse:** in attested candidates, the rightmost heavy syllable is always stressed; if no heavy syllables (# #), final stress.
**Wrong parse:** splitting TR clusters creates additional closed syllables (*kat* in ka.kat.rí), so rightmost heavy may be unstressed (# 0 #).

**Step 2: each setting induces its own Maxent stress grammar** (Goldwater and Johnson, 2003; Hayes and Wilson, 2008). All bigrams and trigrams on either tier become constraints (cf. Gouskova and Gallagher, 2020; Wilson and Gallagher, 2018 on trigrams; Pothos, 2010; Staubs, 2014 on $n \leq 3$).

| **Correct Θ** | | | **Wrong Θ** | | |
|---|---|---|---|---|---|
| Constraint | Penalizes | $w$ | Constraint | Penalizes | $w$ |
| heavy `*0#` | rightmost heavy unstressed | 12.19 | tier0 `*0#` | final unstressed (default) | 7.88 |
| tier0 `*0#` | final unstressed (default) | 7.17 | heavy `*0#` | rightmost heavy unstressed | 4.85 |
| heavy `*#0` | leftmost heavy unstressed | 3.17 | heavy `*#0` | leftmost heavy unstressed | 3.79 |
| heavy `*00` | two unstressed heavies | 2.98 | heavy `*00` | two unstressed heavies | 3.78 |
| | *6 more* | $\leq 0.39$ | heavy `*00#` | last two heavies unstressed | 1.23 |
| | | | | *5 more* | $\leq 0.61$ |

Table 2: Constraint grammars depend on the syllabification and weight settings (Θ)

- For both grammars, "stress the rightmost heavy syllable" (heavy `*0#`) and " stress the final syllable" (tier0 `*0#`) come out on top. But 'heavy' means different things for the two.

**Step 3: compare probability of the attested stress under each grammar.**

- Different representations for the same strings, so different violation profiles of each constraint.

**Evaluation measures** (held-out words, 20%):

(7) a. **Accuracy**: % words whose attested stress is the unique most probable candidate.
b. **Prediction**: mean probability of the attested stress.
c. **Perplexity**: $\exp(-LL/N)$, $N$ = number of words.

| | **Correct Θ** | **Wrong Θ** |
|---|---|---|
| ✓ kakatrí | .999 | .331 |
| kakátri | .0005 | **.669** |
| ✓ kapárto | .999 | .669 |
| kapartó | .0003 | .331 |
| ✓ kakató | .999 | .999 |
| Mean held-out accuracy / prediction | 100% / .9996 | 81.9% / .7411 |
| Held-out Perplexity | 1.0004 | 1.5319 |

Table 3: Probability of attested stress form varies by Θ

- The wrong Θ prefers **kakátri* (based on its parse [ka.kát.ri]) over the attested *kakatrí*.
- For the wrong Θ, because the rightmost heavy is not always stressed, the learner hedges and also has lower prediction even when it does pick the attested stress form *kapárto* (.669 vs. .9997 for right Θ).

**Two ways to choose among the parameter settings.**

- Settings that transform the training data identically are grouped, so the learner's hypothesis space is restricted to what the data has discriminating evidence for.

(8) a. **Argmax:** Fit a grammar for every hypothesis setting and keep the one that fits held-out words best. This is a post-hoc global comparison.
b. **Expectation Maximization (EM)** (Dempster et al., 1977): Keep a probability and an induced grammar for each hypothesis, and refit all the grammars together, with words weighted by how well each grammar explains them. The final probabilities show how certain the learner is.

- EM has been used before for hidden structure learning: syllable structure (Jarosz, 2006), underlying representations (Pater et al., 2012; Wang and Hayes, 2025), and feet (Pater and Prickett, 2022).
- Fitting uses L-BFGS-B (Byrd et al., 1995; Zhu et al., 1997), standard for phonology (Lee et al., 2024; Pater and Prickett, 2022; Pater et al., 2012). Each Θ's grammar fit is convex (Della Pietra et al., 1997; Wang and Hayes, 2025).

## 3 Result 1: stress grammar fit selects the syllabification and weight settings

Artificial language Petrarki (43,716 words), same stress rule (RDTS) with different syllabifications.

(9) a. **nocc**: all CC clusters split across syllables (VT.RV, VR.TV)
Θ: (Max Onset, Onset Rise) = (0,1), weight = CVX
b. **or**: obstruent+*r* forms complex onset (V.TRV), *r*+obstruent splits (VR.TV)
Θ: (Max Onset, Onset Rise) = (1,1), weight = CVX

▷ **The learner picks the syllabification and weight which makes stress predictable** (Table 4).

- Both learners choose the correct syllabification and weight; EM puts posterior 1.0 (unanimous) on it.
- The same holds for leftmost heavy, else initial stress (LDTS); and for CVV-weight corpora where syllabification is correctly left undecided. See Appendix B for full results.

| Corpus | Analysis | | Perplexity | Accuracy | Prediction |
|---|---|---|---|---|---|
| nocc_rdts | (0,1) CVX | **correct** | **1.0004** | **100%** | **.9996** |
| | (1,1) CVX | wrong syllabification | 1.5861 | 79.7% | .7575 |
| | (0,1) CVV | wrong weight | 2.2083 | 65.4% | .5663 |
| | no weight | | 3.1019 | 46.1% | .3654 |
| or_rdts | (1,1) CVX | **correct** | **1.0004** | **100%** | **.9996** |
| | (0,1) CVX | wrong syllabification | 1.5320 | 81.9% | .7411 |
| | (1,1) CVV | wrong weight | 1.8160 | 80.1% | .6812 |
| | no weight | | 2.9873 | 56.3% | .4057 |

Table 4: Grammar perplexity is lowest, accuracy and prediction are highest, for the correct settings

- Where the data give no evidence, the learner reflects that: on purely positional corpora (e.g. final stress), analyses tie (same accuracy, perplexity within $10^{-7}$); EM spreads probability evenly.

## 4 Result 2: how much the parameter settings matter depends on input frequencies

Petrarki or_rdts words without long vowels (40,178 words after removing duplicates).

▷ **Without long vowels, the wrong parse gets even worse.** Every heavy syllable is now CVC, so weight depends *entirely* on syllabification.

| | | With long vowels | | Without long vowels | |
|---|---|---|---|---|---|
| **Analysis** | | **Accuracy** | **Prediction** | **Accuracy** | **Prediction** |
| (1,1) CVX | **correct** | **100%** | **.9996** | **100%** | **.9997** |
| (0,1) CVX | wrong syllabification | 81.9% | .7411 | 68.7% | .6415 |
| no weight | | 56.3% | .4057 | 67.7% | .5015 |
| % heavy syllables (correct parse) | | 37.4% | | 16.2% | |
| % default (final) stress | | 17.3% | | 49.6% | |

Table 5: How the wrong Θ performs depends on input frequencies

**With long vowels**: about two thirds of heavy syllables are CVV, which doesn't depend on the parse, so a wrong parse still gets 81.9% accuracy.

**Without long vowels**: wrong parse with weight similar to no weight (68.7% vs. 67.7%).
Fewer heavy syllables, more default final stress, so no-weight grammar does better (56.3% → 67.7%).

## 5 Relation to previous proposals

- **Dresher's objection to global fit** (Dresher, 1999): a wrong setting can mask a right one, so overall extensional fit may not be a reliable metric for navigating the hypothesis space, see also Yang, 2002. Here every configuration gets its own fitted grammar, and each fit is convex, so there are no local optima within a Θ. On every weight-sensitive corpus tested, the best global fit is the generating analysis or an exact tie with it.
- **Global vs. local:** both learners compare whole analyses, unlike cue-based learning (Dresher and Kaye, 1990) and EDL (Jarosz, 2015; Nazarov and Jarosz, 2017, 2021). On credit granularity: the

NPL gives credit to the whole grammar (Yang, 2002); EDL per parameter or per constraint pair; Argmax per $\Theta$ over the corpus; EM per $\Theta$ for each word.

- **Domain-general vs. specific** (Chomsky, 2005; Yang, 2010): the learning strategy is domain-general. The representational information is domain-specific: the syllabification and weight parameters, and the candidate space of one primary stress per word.
- **Hybrid model** Following the UG plus statistics blueprint in Pearl, 2021. Here: UG parameters for representations, statistical induction for stress learning; reduces UG and search space.
- The learner builds a stress grammar from the data, not by searching a theoretical stress parameter space. The EDPL (Nazarov and Jarosz, 2021) learns 93% of the 302 systems predicted by Dresher and Kaye, 1990, but only 48 are attested in StressTyp2 (Goedemans et al., 2015).

## 6 Takeaways

- The fit of an induced stress grammar can select the syllabification and weight type that support it, learning from unsyllabified strings (Result 1).
- How impactful the selection is depends on the data: when weight comes only from codas, a wrong parse is more fatal (Result 2).
- Restricting UG to representations and inducing the stress grammar keeps the search space tractable.
- **Next:** real corpora with noisy data; secondary stress; lexical strata, which need split grammars; inducing syllabification constraints and sonority values from data; acquisition trajectory by tuning input size and frequency.

## A Appendix: Learner details

**Parameters.** 3 consonant × 3 nucleus × 3 weight settings = 27 combinations $\Theta$.

(10) a. Consonants (§2): MAX ONSET, ONSET RISE; settings (0,1), (1,1), (1,0).
b. Nuclei: settings (COMPLEX NUCLEI, HETERO NUCLEI) = (0,0), (1,0), (1,1).
(i) COMPLEX NUCLEI: two adjacent vowels can form one nucleus. [baa.rii] vs. [ba.a.ri.i]
(ii) HETERO NUCLEI: the two vowels can differ. [baa.rai] vs. [baa.ra.i]
c. Weight (§2): CVV HEAVY, CVC HEAVY; settings none, CVV, CVX.

**Grammar.** Constraints $C_i$ are all bigrams and trigrams over the base stress tier and the heavy tier (with edges #), with violations $v_i$. For a word $x$ with candidates $y$ (one stress per word):

$$H(y) = -\sum_i w_i\, v_i(y), \qquad P(y \mid x) = \frac{\exp H(y)}{\sum_{y' \in \mathrm{Cands}(x)} \exp H(y')}$$

Keeping weights $w_i \geq 0$, maximize the total log-likelihood of the training data minus an effect penalty.

*Effect penalty E*: $\lambda$ times the mean, over training words, of the variance of $H$ across the word's candidates.

*min_cost* weights: minimize $\sum_i \mathrm{cost}_i\, w_i$ (cost = n-gram length, +1 on the heavy tier) among them.

**EM.** Hypotheses $h$ are groups of $\Theta$ that transform the training data identically; prior $P(h)$ starts uniform as 1/number_of_groups. Probability of a word $x$ is:

$$P(x) = \sum_h P(h)\, P(x \mid \mathrm{grammar}_h)$$

E-step: calculate each word's posterior $P(h \mid x)$. M-step: update priors $P(h)$ = mean posterior of that $h$ over words; refit each grammar on the data weighted by posteriors.

## B Appendix: Petrarki consonant ambiguity full results

PetrarkiB (43,716 words); medial consonants: single (60%), $r$+obstruent (15%), obstruent+$r$ (25%); 25% of vowels are long; 15% of words end in a consonant.

| Corpus | Winner (both) | Best wrong syll. | Wrong weight | No weight |
|---|---|---|---|---|
| nocc_rdts_cvx | (0,1) CVX | (1,1): 1.586 / 79.7 / .7575 | 2.208 / 65.4 / .5663 | 3.102 / 46.1 / .3654 |
| or_rdts_cvx | (1,1) CVX | (1,0): 1.531 / 86.7 / .7800 | 1.816 / 80.1 / .6812 | 2.987 / 56.3 / .4057 |
| nocc_ldts_cvx | (0,1) CVX | (1,1): 1.737 / 75.5 / .6828 | 2.087 / 67.1 / .5836 | 2.562 / 63.3 / .4799 |
| or_ldts_cvx | (1,1) CVX | (1,0): 1.715 / 82.6 / .7120 | 1.851 / 79.2 / .6732 | 3.038 / 54.8 / .3952 |
| rdts_cvv | CVV, any syll. | — | 1.298 / 90.6 / .8527 | 2.923 / 59.3 / .4268 |
| ldts_cvv | CVV, any syll. | — | 1.121 / 93.9 / .9202 | 2.925 / 59.3 / .4266 |

Results format: Held-out perplexity / accuracy % / prediction. Winner scores 1.0004 / 100 / .9996 on every corpus.

- Both learners recover the generating syllabification and weight type, EM posterior 1.0 on the winner.
- For or, the third setting (1,0) (every cluster an onset) is the closest wrong syllabification; the main text uses (0,1) to keep the nocc/or contrast.

## References


Apoussidou, D. (2007). "The Learnability of Metrical Phonology". Universiteit van Amsterdam. 221 pp.

Boersma, P. and J. Pater (2016). "Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar". In: *Harmonic Grammar and Harmonic Serialism*. Ed. by J. J. McCarthy and J. Pater.

Bosch, A. R. K. (2011). "Syllable-Internal Structure". In: *The Blackwell Companion to Phonology*.

Broselow, E., S.-I. Chen, and M. Huffman (1997). "Syllable Weight: Convergence of Phonology and Phonetics". In: *Phonology* 14.1.

Byrd, R. H. et al. (1995). "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific Computing* 16.5.

Chomsky, N. (2005). "Three Factors in Language Design". In: *Linguistic Inquiry* 36.1.

Clements, G. N. (1990). "The Role of the Sonority Cycle in Core Syllabification". In: *Papers in Laboratory Phonology*. Ed. by J. Kingston and M. E. Beckman.

Daland, R. et al. (2011). "Explaining Sonority Projection Effects". In: *Phonology* 28.2.

Della Pietra, S., V. Della Pietra, and J. Lafferty (1997). "Inducing Features of Random Fields". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.4.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum Likelihood from Incomplete Data Via the EM Algorithm". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1.

Derwing, B. L. (1992). "A 'Pause-Break' Task For Eliciting Syllable Boundary Judgments from Literate and Illiterate Speakers: Preliminary Results for Five Diverse Languages". In: *Language and Speech* 35.1–2.

Dresher, B. E. (1999). "Charting the Learning Path: Cues to Parameter Setting". In: *Linguistic Inquiry* 30.1.

Dresher, B. E. and J. D. Kaye (1990). "A Computational Learning Model for Metrical Phonology". In: *Cognition* 34.2.

Fokes, J. and Z. Bond (1993). "The Elusive / Illusive Syllable". In: *Phonetica* 50.2.

Goedemans, R., J. Heinz, and H. Van Der Hulst (2015). *StressTyp2*. Version 1.

Goldwater, S. and M. Johnson (2003). "Learning OT Constraint Rankings Using a Maximum Entropy Model". In: *Proceedings of the Workshop on Variation within Optimality Theory*.

– (2005). "Representational Bias in Unsupervised Learning of Syllable Structure". In: *Proceedings of the Ninth Conference on Computational Natural Language Learning - CONLL*.

Gordon, M. (2002). "A Phonetically Driven Account of Syllable Weight". In: *Language* 78.1.

– (2017). "Syllable Weight: A Typological and Theoretical Overview". In: *Current Issues in Linguistic Theory*. Ed. by P. Newman. Vol. 338.

Gordon, M. and T. Roettger (2017). "Acoustic Correlates of Word Stress: A Cross-Linguistic Survey". In: *Linguistics Vanguard* 3.1.

Gouskova, M. and G. Gallagher (2020). "Inducing Nonlocal Constraints from Baseline Phonotactics". In: *Natural Language & Linguistic Theory* 38.1.

Hayes, B. and C. Wilson (2008). "A Maximum Entropy Model of Phonotactics and Phonotactic Learning". In: *Linguistic Inquiry* 39.3.

Heinz, J. (2009). "On the Role of Locality in Learning Stress Patterns". In: *Phonology* 26.2.

Itô, J. (1988). *Syllable Theory in Prosodic Phonology*. Routledge Library Editions: Phonetics and Phonology 10. 228 pp.

Jarosz, G. (2006). "Rich Lexicons and Restrictive Grammars: Maximum Likelihood Learning in Optimality Theory". PhD thesis. Johns Hopkins University.

– (2013). "Learning with Hidden Structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing". In: *Phonology* 30.1.

– (2015). "Expectation Driven Learning of Phonology". Unpublished ms.

– (2019). "Computational Modeling of Phonological Learning". In: *Annual Review of Linguistics* 5.

Kelly, N. (2021). "Syllable Weight, Vowel Length and Focus in Lebanese Arabic". In: *Glossa: a journal of general linguistics* 6.1.

Lee, S. S., J. Pater, and B. Prickett (2024). "Representing and Learning Stress: A MaxEnt Framework for Comparing Learning Across Grammatical Theories". Unpublished ms.

Macaulay, B. (2024). "Speaker Judgments Alone Cannot Diagnose Syllable Structure". In: *Theoretical Linguistics* 50.3–4.

McCarthy, J. J. (1979). "On Stress and Syllabification". In: *Linguistic Inquiry* 10.3.

Nazarov, A. (2016). "Extending Hidden Structure Learning: Features, Opacity, and Exceptions". PhD thesis. University of Massachusetts Amherst.

Nazarov, A. and G. Jarosz (2017). "Learning Parametric Stress without Domain-Specific Mechanisms". In: *Proceedings of the Annual Meetings on Phonology 2016*.

– (2021). "The Credit Problem in Parametric Stress: A Probabilistic Approach". In: *Glossa: a journal of general linguistics* 6.1.

Pater, J. and B. Prickett (2022). "Typological Gaps in Iambic Nonfinality Correlate with Learning Difficulty". In: *Proceedings of the Annual Meetings on Phonology 2021*.

Pater, J. et al. (2012). "Learning Probabilities over Underlying Representations". In: *Proceedings of the Twelfth Meeting of the ACL-SIGMORPHON*.

Pearl, L. S. (2021). "How Statistical Learning Can Play Well with Universal Grammar". In: *A Companion to Chomsky*. Ed. by N. Allott, T. Lohndal, and G. Rey.

Pothos, E. M. (2010). "An Entropy Model for Artificial Grammar Learning". In: *Frontiers in Psychology* 1.

Staubs, R. (2014). "Computational Modeling of Learning Biases in Stress Typology". PhD thesis. University of Massachusetts Amherst.

Steriade, D. (1999). "Alternatives to the Syllabic Interpretation of Consonantal Phonotactics". In: *Proceedings of the 1998 Linguistics and Phonetics Conference*. Ed. by O. Fujimura, B. Joseph, and B. Palek.

Tesar, B. and P. Smolensky (1998). "Learnability in Optimality Theory". In: *Linguistic Inquiry* 29.2.

– (2000). *Learnability in Optimality Theory*.

Treiman, R. and C. Danis (1988). "Syllabification of Intervocalic Consonants". In: *Journal of Memory and Language* 27.1.

Wang, Y. and B. Hayes (2025). "Learning Phonological Underlying Representations: The Role of Abstractness". In: *Linguistic Inquiry*.

Weber, D. J. (1983). "A Grammar of Huallaga (Huanuco) Quechua". PhD thesis. University of California, Los Angeles.

Wilson, C. and G. Gallagher (2018). "Accidental Gaps and Surface-Based Phonotactic Learning: A Case Study of South Bolivian Quechua". In: *Linguistic Inquiry* 49.3.

Yang, C. (2002). *Knowledge and Learning in Natural Language*.

– (2010). "Three Factors in Language Variation". In: *Lingua* 120.5.

Zhu, C. et al. (1997). "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization". In: *ACM Transactions on Mathematical Software (TOMS)* 23.4.