Language-specific factors influence learnability: Case study from contour tone licensing*

Charlie O'Hara

University of Southern California

1. Learnability shapes typology

In constraint-based phonology, typology is typically considered in terms of grammatical mappings. Languages are categorized in terms of the patterns of phonotactically licit forms, and what phonological processes occur in the input and output mapping. However, languages differ on more factors than just their grammatical patterns, and languages with particular traditionally non-grammatical factors may be more likely to show certain grammatical patterns than others. For example, Zhang (2002, 2004) argues that the phonetic duration of certain phonologically identical syllable types vary across languages, and syllables with longer phonetic duration are more likely to bear more tonal contrasts. Zhang argues for a "direct approach" to capturing these associations, by allowing the phonological grammar to access information about the phonetic realization of sounds, making phonetic duration effectively a "grammatical factor".

Here I argue that phonetic duration is not the only language-specific non-grammatical factor that can affect grammatical patterning, but other language-specific factors cannot be implemented as part of the grammar as easily as phonetic factors. Particularly, I focus on lexical frequency as a language-specific factor that may associate with certain grammatical patterns over others. Where the phonetic realization of a word is directly available in the speech stream, determining the frequency of a structure in a language requires computation across the whole language. As an alternative to the direct approach, I argue for a learnability approach that can capture the associations between grammatical patterns and types of lexical frequency distributions.

Recent work has highlighted the importance of learnability to shape typology. In some approaches, learnability acts as a categorical filter on typology, ruling out grammatically possible patterns that are completely unlearnable from realistic learning data, given a par-

*I would like to thank Karen Jesney, Rachel Walker, Caitlin Smith, Hayeun Jang, Betsy Pillion, Larry Hyman, and audiences at USC, UC Berkeley, Miami University, NELS 2019, and LSA 2020 for feedback on this project.

ticular learning algorithm (Boersma 2003, Heinz 2009). Even among learnable patterns, though, not all patterns are equally learnable, allowing for gradient typological predictions (Wilson 2006, Pater and Moreton 2012, Staubs 2014, Stanton 2016, Hughto 2019, O'Hara In prep, 2019).

Harder-to-learn patterns require more data to be learned to any chosen degree of accuracy. As a result, harder-to-learn patterns will be likely to be learned slightly less accurately than easier-to-learn patterns even when both patterns are learned extremely accurately. This difference in accuracy compounds across generations making harder-to-learn patterns more likely to change over time, and therefore be less frequent in typology.

This approach allows for non-grammatical factors that can influence learnability to affect typology. As a result, we can predict associations between grammatical patterns and non-grammatical language specific factors, like lexical frequency. Typology can be defined not simply in terms of the grammatical patterns generated by the factorial typology, but by grammar-external factors as well. In this paper, I show that lexical frequency can impact the learnability of tone licensing patterns. This difference in learnability leads to a statistical association between certain skewed frequency distributions and grammatical patterns. I show that Thai and Navajo both exhibit skewed frequency distributions that would foster the learning of the tone licensing pattern in each language, while depressing the learning of the inverse tone licensing pattern.

2. Tone licensing

It is possible for language specific factors such as lexical frequency to interact with a wide variety of phonotactic phenomena, here, I'll look particularly at tone licensing patterns following Zhang (2004, 2002). Many languages allow a restricted set of tones in some syllable types, while allowing a larger set of contrastive tones in other syllables. This paper focuses on two aspects of syllable type that can affect tone licensing: vowel length, and coda sonority. Zhang (2002) details a typological survey showing that many languages, including Navajo, restrict tone contrasts on syllables with short vowels, regardless of coda consonant; and that many languages, including Standard Thai, restrict tone contrasts on syllables with obstruent-coda syllables, regardless of vowel length.[1]

First, consider Navajo, which restricts contour tones only in short syllables. Navajo has a distinction between short and long vowels. Only a subset of Navajo consonants can appear in codas: [d s ʃ z ʒ tʃ ɬ l n h ʔ] appear in the corpus explored below. Of these, all but [l] and [n] are obstruents. Navajo has four contrastive tones: low, high, falling, and rising. The level tones (low and high) are available on any syllable of a word, but the contour tones (rising and falling) are only available on long vowels and diphthongs. Contour tones are available on long vowels regardless of the sonority of the syllable coda, whether open

[1]For simplicity, following Zhang (2002), I assume that these two aspects of syllable type affect tone licensing in similar ways, and target the same sets of tones, "contour tones". Morén and Zsiga (2006) argue that obstruent-coda syllables are associated with low tones, predicting a different set of tones to be restricted in obstruent-coda syllables than short syllables. Morén and Zsiga (2006) indeed capture the asymmetry between which tones are restricted in Thai and Navajo. I abstract away from these differences in tone restrictions in this paper, but the lexical frequency approach would perform similarly without the abstraction.

(CV(V)), obstruent-coda (CV(V)O) or sonorant-coda (CV(V)N). Examples of the attested contour tones in Navajo are presented in (1).

(1) *Navajo Contour Tones (Data from Zhang 2002)*

	CV(V)	CV(V)O	CV(V)N
Short	*[sá**nǐ**]	*[pì**tǐɬ**]	*[pì**kʰǐn**]
Long	[há**kǒː**nèːʔ] ‘let’s go’	[**tĕɪʒ**níːɬton] ‘they shot at him’	[**tĕɪl**ʔá] ‘they extend’

On the other hand, in Thai, tones are more restricted in obstruent-coda syllables rather than short syllables. Thai also has a contrast between long and short vowels, except in open syllables, where all vowels are long (Zhang 2004). Thai only allows [p t k m n ŋ w j] in coda position, with [m n ŋ w j] being sonorants, and [p t k] being obstruents. Thai has five contrastive tones: high, low, mid, falling, and rising. Of these, rising tones (and mid tones) are most restricted.[2] Rising tones are banned in obstruent-coda syllables (CV(V)O) regardless of vowel length, but are available in short sonorant-final syllables as seen in (2).

(2) *Thai Rising Tones (Data from Morén and Zsiga 2006)*

	CV(V)	CV(V)O	CV(V)N
Short	N/A	*[**lǎk**]	[**lǎŋ**] ‘back’
Long	[**nǎː**] ‘thick’	*[**lǎːk**]	[**lǎːŋ**] ‘grandchild’

Vowel length and coda sonority are orthogonal factors that can impact the licensing of tonal contrasts. In Navajo, vowel length affects tone licensing more than coda sonority, and in Thai coda sonority affects tone licensing more than vowel length.

3. Corpus study

If lexical frequency affects the tonal licensing patterns within a language, we would expect that in languages with more short syllables than obstruent-coda syllables, short syllables should license more tones than obstruent-coda syllables, and vice versa. This section reports on corpus analyses of Thai and Navajo, which confirm that the relative frequencies of short and obstruent-coda syllables pattern with tonal licensing in those syllable types.

In order to get the frequency of syllable types in Navajo, I used a wordlist extracted from Wiktionary as part of the SIGMORPHON shared task 2017 (Cotterell et al. 2017).[3]

[2]Falling and high tones are rare in native Thai words in CVO and CVVO syllables respectively, though Zhang (2004) notes that this is likely a result of a diachronic process, and loanwords have filled these gaps, suggesting that this is not a synchronic restriction.

[3]This wordlist is `navajo-train-high` from task1 available at `https://github.com/sigmorphon/conll2017/tree/master/all/task1` as of June 11, 2020.

This list is made up of 10,000 pairs of Navajo lemmas and inflected forms, including nouns and verbs. The 10,000 pairs of forms were sampled by the creators of the wordlist from 12,354 forms available in the February 2017 dump of Wiktionary. This list includes 20,000 words, but includes duplicates. Using a python script, I collected all unique words (9,255) and classified them according to the syllable type of their final syllable. I used only final syllables for two reasons: Thai words are largely monosyllabic, so this avoids confounds that could be created by comparing non-final syllables with final syllables, and the coda consonants of a final syllable are less ambiguous than those in non-final syllables. Syllables were classified as having obstruent codas if they ended with [d s z ʃ ʒ ɬ ʔ], and sonorant-final if the ended in a vowel, [l] or [n].

The frequencies of the different syllable types are Navajo is presented in (3). It is clear that obstruent-coda syllables are more frequent in Navajo final syllables than short vowels. This is because there are nearly three times as many long obstruent-coda syllables (CVVO) than there are short sonorant-final syllables (CV(N)). Recall that Navajo bans contour tones on short syllables, but allows them on long syllables, even if they have obstruent codas. The more frequent syllable type (CVVO) licenses more contexts than the less frequent syllable type (CV(N)), consistent with the hypothesis that lexical frequency is associated with grammatical patterns.

(3) *Frequency of Syllable Types in final syllables of Navajo*

	Obstruent-Final	Sonorant-Final	Total
Short	CVO 32.9%	CV(N) 15.3%	48.2%
Long	CVVO 45.4%	CVV(N) 6.4%	51.8%
Total	78.3%	21.7%	100%

For Thai, I extracted 2,961 unique words[4] of child-directed speech from the CRSLP-MARCS corpus (Luksaneeyanawin 2000) on Childes (MacWhinney 2000), a phonologically transcribed corpus of Thai. This was done by collecting all speech by participants other than the infant in the corpus, including mothers, fathers, grandparents, researchers, and siblings. This type of corpus is a more realistic type of corpus to use for this type of study than the wordlist used for Navajo, because it generates a corpus of words that we know the children observe in the process of acquisition.[5] I classified words according to the length of the vowel and the sonority of coda consonants. Because Zhang (2004) shows that open syllables transcribed with short vowels in Thai are phonetically longer than long

[4]Over 99% of the words in the Thai corpus were monosyllabic, so I removed the nine hapax bisyllabic words found in the corpus.

[5]This corpus also allows for comparison between type and token frequencies of syllable types. I use type frequencies, to compare with the Navajo lexical corpus. Further, a number of studies have suggested that type frequencies have greater impact on grammatical generalization that token frequencies (Bailey and Hahn 2001, Archer and Curtin 2011). The token frequencies found in this corpus show a stronger skew towards short sonorant-final syllables than observed in the type frequencies, enhancing any of the effects under investigation in this paper.

vowels in closed syllables, I classified orthographically short open syllables as long.[6] Syllables were considered sonorant-final if they ended in a vowel or [m n ŋ j w], they were considered to have obstruent codas if they ended in [p t k].

The frequencies of the different syllable types in Thai are presented in (4). The opposite skew is observed relative to what was seen in Navajo. Short vowels are more frequent in Thai than obstruent-coda syllables. This is because there are nearly twice as many short sonorant-final syllables (CVN) than there are long obstruent-coda syllables (CVVO). Recall that Thai bans contour tones on obstruent-coda syllables, but allows them on sonorant-final syllables, even if they are short. The more frequent syllable type (CVN) licenses more contexts than the less frequent syllable type (CVVO), consistent with the hypothesis that lexical frequency is associated with grammatical patterns.

(4) *Frequency of Syllable Types in final syllables of Thai*

	Obstruent-Final	SonorantFinal	Total
Short	CVO 12%	CV(N) 25%	37%
Long	CVVO 13%	CVV(N) 50%	63%
Total	25%	75%	100%

4. Learning model

As an explicit model to demonstrate how lexical frequency can affect the learnability of different patterns, I present a set of generational learning simulations based on a MaxEnt grammar and the Perceptron learning algorithm, following Staubs (2014), Hughto (2019), O'Hara (In prep).

The grammatical model considered here is a MaxEnt Harmonic Grammar (Goldwater and Johnson 2003), with a set of four constraints: *CONTOUR, *CONTOUR/SHORT, *CONTOUR/OBS-CODA, and IDENT-TONE, defined in (5).

(5)
a. *CONTOUR - Assign a violation for a syllable with a contour tone.
b. *CONTOUR/SHORT - Assign a violation for a syllable with a contour tone with a short vowel.
c. *CONTOUR/OBS-FINAL - Assign a violation for a syllable with a contour tone and an obstruent coda.
d. IDENT-TONE - Assign a violation for a syllable that has a different tone than its input correspondent has.

I make use of the truncated perceptron rule (Rosenblatt 1958, Boersma and Pater 2016, Magri 2015) (6). This is an error-driven learning algorithm. On each step of the learning procedure, an input form is randomly sampled from the teacher's grammar according to the

[6] This assumption actually reduces the number of possible short sonorant-final syllables, decreasing the frequency skew under investigation here.

lexical frequency condition. The learner and the teacher each produce an output form for that form according to the probability distribution generated by their grammar. If the forms differ, the learner raises the weight on all constraints that favor their own form, and lowers the weight of all constraints that favor the teacher's form (unless the constraint's weight would become less than zero).

(6) *Perceptron update rule*
$w_C(t+1) = w_C(t) + \eta * (C(i,y_L) - C(i,y_T))$
(η is a learning rate constant, here set to .05, $w_C(t)$ is the weight of a constraint C at learning step t. $C(i,y_L)$ is the number of violations assigned to the mapping between input i and the learner's output, y_L. y_T is the teacher's output.)

The model makes use of an iterated agent-based model (Kirby and Huford 2002, Hughto 2019), built up of a number of generations of learners. The first learner is exposed to a limited number of forms (here, 5000) sampled from the target pattern. Then, the second learner observes the same number of forms sampled from whatever grammar the first learner learned, and so on. After many generations (here, 40), small inaccuracies in learning expand. At the end of learning, I classify the grammar learned by the final learner according to the closest categorical grammar (defined as the categorical grammar denoted by these constraint weights in categorical harmonic grammar.)

The stability of a pattern given a lexical frequency distribution can be measured by performing multiple runs of the generational learning model starting with a particular categorical pattern, and counting the percentage of runs where the last learner's grammar is classified the same as the starting pattern. A harder-to-learn pattern will be less stable than an easier-to-learn pattern. The precise rates of stability are sensitive to parameters of the simulation such as the number of forms observed per generation, and the number of generations, but the relative stability rates remain consistent across a variety of parameter settings. The parameters used here were chosen so that there were enough forms per generation such that patterns were learned with over 99% accuracy in the first generation, and enough generations that there was a difference in stability between patterns.

5. Simulations

I performed learning simulations crossing three different lexical frequency conditions with two grammatical patterns. The two grammatical patterns I tested are the NoShortContours pattern (7a), where contour tones are banned on short syllables, but allowed on long syllables regardless of coda sonority; and the NoObsCodaContours pattern (7b), where contour tones are banned on obstruent-coda syllables regardless of vowel length.

(7)

a. NoShortContours

	Obs-Coda	Son-Coda
Short	*pǎt	*pǎn
Long	pǎːt	pǎːn

b. NoObsCodaContours

	Obs-Coda	Son-Coda
Short	*pǎt	pǎn
Long	*pǎːt	pǎːn

Four types of input forms were sampled at each iteration: /pǎt/, /pǎːt/, /pǎn/, or /pǎːn/. The grammar consisted of two candidates for each input, a faithful candidate, and one that replaced the contour tone with a low tone. In the first generation, if an input form was sampled with a contour tone that was not licensed in the target pattern (for ex: [pǎt] is illicit in both patterns), the teacher produced the unfaithful candidate ([pàt]).

I measured the stability and learnability of both of these patterns paired with each of three frequency conditions, by running 50 generational simulations where the first generation sampled forms from the target pattern. The three frequency conditions represented three types of frequency skews: in the Control condition, all four forms equally likely to be sampled on each iteration, in the Thai and Navajo conditions, I used the frequencies found in the corpus studies in section 3: the Thai frequency distribution has more short syllables than obstruent-coda syllables whereas the Navajo frequency distribution has more obstruent-coda syllables than short syllables.

Simulations were ran using a Python script available at the author's website (O'Hara 2020). Parameters were run according to the parameters in (8). Learners were initialized with faithfulness constraints weighted low and markedness constraints weighted high.

(8) *Parameters of Simulations*

Parameter	Value
Learning steps per generation	5000
Generations per run	40
Runs per condition	50
Learning Rate (η)	0.05
Initial Weight (Faith)	1
Initial Weight (Mark)	50

I report two measures of learnability: a metric of inter-generational stability, and a metric of learning speed in the first generation. The stability rates that were the outcome of these simulations are presented in (9). The learning speed metric is measured using the average number of learning steps it takes for the first generation's learner to match the teacher's grammar 99% of the time, presented in (10).

(9) *Stability after 40 generations across 50 runs*

	Pattern	
Frequency	NoShortContours	NoObsCodaContours
Control	98%	98%
Navajo	100%	**0%**
Thai	**30%**	98%

(10) *Average (and standard deviation) of learning steps for first generation to obtain 99% accuracy*[7]

	Pattern	
Frequency	NoShortContours	NoObsCodaContours
Control	2770±184	2743±197
Navajo	2145±196	**4539±203**
Thai	**3852±172**	2510±216

In the Control frequency condition, where all forms were observed equally frequently, there is no major distinction between the learning of the two patterns. Both patterns were stable in all but one run, and the average number of learning steps needed to learn the pattern are both in the mid 2700s. With the Navajo frequency distribution, which is skewed with more obstruent-coda syllables than short syllables, the NoShortContours pattern is learned easier on both metrics than the NoObsCodaContours pattern. The NoObsCodaContours pattern was not stable on any of the 50 runs, whereas the NoShortContours pattern was stable on all 50 runs. The NoShortContours pattern is learned in around 600 fewer learning steps with the Navajo frequency distribution than the Control distribution, but the NoObsCodaContours pattern requires around 1800 more learning steps. These results are consistent with the idea that the Navajo frequency distribution fosters the learning of the phonotactic pattern observed in Navajo, because Navajo exhibits the NoShortContours pattern.

With the Thai frequency distribution, which is skewed in the other direction, with more short syllables than syllables with obstruent codas, the NoObsCodaContours pattern is learned easier on both metrics. The NoShortContours pattern was stable on only 15 of the 50 runs, whereas the NoObsCodaContours pattern was stable on all but one of the 50 runs. The NoObsCodaContour pattern is learned in around 200 fewer learning steps with the Thai frequency distribution than the Control distribution, but the NoShortContours pattern requires around 1100 more learning steps. The Thai frequency distribution fosters the learning of the phonotactic pattern observed in Thai.

6. Discussion

The learnability differences between the grammatical patterns under skewed frequency distributions can lead to associations between grammatical patterns and frequency distributions over time. The grammatical framework alone allows for both grammatical patterns to be modeled under any of the lexical frequency distributions. However, language types with a frequency-pattern mismatch are more likely to be unstable than languages where frequent forms allow contour tones. As a result, even if mismatched and matched patterns were equally common at one point in time, after a number of generations, mismatched patterns would be underrepresented due to their instability.

A direct approach for capturing the associations between lexical frequency and grammatical patterns can be ruled out for two reasons. First, a direct approach would require

[7] All two way comparisons between conditions were found highly significant ($p < 0.0001$) by two-sided t-test, except the comparison between the two control conditions, which highly overlapped ($p = .47$).

including some constraint like *CONTOUR/INFREQUENT which is violated by contour tones on syllable types that are sufficiently infrequent. Computing the frequency of a syllable type would require a speaker to perform a corpus analysis on their entire lexicon in order to determine the phonotactics of a word, and the types of forms that violate this constraint would often change throughout the process of learning. Second, directly referencing frequency information introduces unnecessary complications to our model. It is uncontroversial that grammatical patterns must somehow be learned, and lexical frequency will indirectly affect the state of the grammar in this way under most error-driven learning algorithms.

Further, unlike the direct approach, the learnability approach may be able to capture both the effects of lexical frequency *and* phonetic duration. Zhang (2004) argues that contour tones are less accurately produced and perceived when they are on shorter duration syllables. An extension of this model of learning could include noisy transmission between teacher and learner, so that the learner did not always observe the contour tone when the teacher's grammar output one (see Boersma 2006 for a similar approach). If shorter duration contour tones are less likely to be observed accurately than longer duration ones, the learner will in effect observe fewer contour tones in shorter duration syllables than longer duration ones, skewing the frequencies that the learner observes in the same way lexical frequency skews do. Such simulations are outside of the scope of the current paper, but future work will investigate how learnability can allow a variety of non-grammatical language-specific factors to influence the grammatical patterns of languages.

References

Archer, Stephanie L., and Suzanne Curtin. 2011. Perceiving onset clusters in infancy. *Infant Behavior and Development* 34:534–540.

Bailey, Todd M., and Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44:568–591.

Boersma, Paul. 2003. Review of Bruce Tesar and Paul Smolensky 2000, Learnability in Optimality Theory. *Phonology* 20:436–446.

Boersma, Paul. 2006. The acquisition and evolution of faithfulness rankings. Presented at the Manchester Phonology Meeting, Available at `https://www.fon.hum.uva.nl/paul/presentations/BoersmaMFM14.pdf`.

Boersma, Paul, and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic Grammar and Harmonic Serialism*, ed. by John J. McCarthy and Joe Pater. Sheffield: Equinox.

Cotterell, Ryan, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kubler, David Yarowsky, Jason Eisner, and Mans Hulden. 2017. The CoNLL-SIGMORPHON 2017 shared task: Universal morphological reinflection in 52 languages. In *Proceedings of the CoNLL-SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, 1–30. Vancouver, Canada: Association for Computational Lingusitics.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University, Department of Linguistics.

Heinz, Jeffrey. 2009. On the role of locality in learning stress patterns. *Phonology* 26:303–351.

Hughto, Coral. 2019. Emergent typological effects of agent-based learning models in maximum entropy grammar. Doctoral dissertation, University of Massachusetts Amherst.

Kirby, Simon, and James Huford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language*, ed. by Angelo Cangelosi and Domenico Parisi, chapter 6, 121–148. London: Springer Verlag.

Luksaneeyanawin, S. 2000. Speech computing and speech technology in Thailand. *Interdisciplinary Approaches to Language Processing* 267–321.

MacWhinney, B. 2000. *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates, third edition.

Magri, Giorgio. 2015. How to keep the HG weights non-negative: the truncated perceptron reweighting rule. *Journal of Language Modeling* 3:345–375.

Morén, Bruce, and Elizabeth Zsiga. 2006. The lexical and post-lexical phonology of thai tones. *Natural Language & Linguistic Theory* 24:113–178.

O'Hara, Charlie. 2019. Learning prevents MaxEnt from giving probability to harmonically bounded candidates. Talk given at Annual Meeting on Phonology 2019. Slides available at `https://dornsife.usc.edu/assets/sites/837/docs/oharaamp2019.pdf`.

O'Hara, Charlie. 2020. Soft typology tool v. 0.3. Software package available at `http://dornsife.usc.edu/ohara/stt/`.

O'Hara, Charlie. In prep. Soft biases in phonology: Learnability meets grammar. Doctoral dissertation, University of Southern California.

Pater, Joe, and Elliott Moreton. 2012. Structurally biased phonology: complexity in language learning and typology. *The EFL Journal* 3:1–44.

Rosenblatt, Frank. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65:386–408.

Stanton, Juliet. 2016. Learnability shapes typology: the case of the midpoint pathology. *Language* 92:753–791.

Staubs, Robert. 2014. Computational modeling of learning biases in stress typology. Doctoral dissertation, University of Massachusetts Amherst.

Wilson, Colin. 2006. Learning phonology with a substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.

Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. New York: Routledge.

Zhang, Jie. 2004. The role of contrast-specific and language-specific phonetics in contour tone distribution. In *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 157–190. New York: Cambridge University Press.

Charlie O'Hara
charleso@usc.edu