The Sweet Spot Effect: Rare Phonotactic Patterns Require Specific Lexical Frequencies

Charlie O'Hara - University of Southern California

Overview. Recent work has suggested that learnability can help shape phonological typology (Pater & Moreton, 2012; Staubs, 2014; Stanton, 2016). Rare patterns are less likely to be learned stably across generations, or to be innovated when another pattern is mislearned. A typological survey of 77 genetically-diverse languages that allow [k p] and [t] in word-initial position reveals that Finnish is alone in allowing only [t] in word-final position (Figure 1). If learning was to blame for the rarity of the [t]-final pattern, one might expect Finnish would be unstable, losing this pattern over a few real life generations. In spite of this prediction, Finnish has maintained this typologically rare pattern for at least 500 years. This paper shows that the lexical frequency of coda stops greatly influence the stability of different phonotactic patterns. It is shown that the Finnic languages fall neatly in the sweet-spot for the [t]-final pattern; there is a small and narrow range of lexical frequencies within which the [t]-final pattern is likely to be observed. The relative smallness of the [t]-final sweet spot is hypothesized to be responsible for the underattestation of this pattern crosslinguistically.

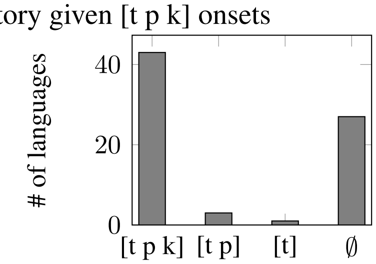
How Stability Relates to Typological Frequency. The [t]-final pattern serves as an intermediate pattern between two well-attested patterns, all-finals, and no-finals. When a language decays from the all-finals pattern to the no-finals pattern, theories of generational learning that incorporate the relative unmarkedness of coronals (Kean, 1975; de Lacy, 2006) predict that the language is likely to pass through [t]-final. At each generation, there is some probability that the language decays, conditioned by the current pattern and the lexical frequencies. The typological likelihood of [t]-final is conditional on two factors, having sufficiently high inflow to the pattern, and sufficiently low outflow. This paper shows that the lexical frequencies present in four different languages predict the inflow/outflows that lead to the patterns of change observed within each language's family.

Probability of language change were measured using a generational stabilty MaxEnt model based on Staubs (2014); Hughto (2018). Each run consists of 40 generations of learners who are trained on a limited number of forms (4600) from the previous generation's final grammar. Within each generation, the forms are randomly selected based on the lexical frequencies present in the tested language. 50 runs were performed for each initial target pattern. The specific results presented here (in Figure 2) are dependent on the number of generations and iterations per run, but the relative stability between patterns remain robust across different parameter settings.

Table 1: Normalized frequencies for tested languages

Language	$n =$	Initial			Final		
		p(tV)	p(pV)	p(kV)	p(Vt)	p(Vp)	p(Vk)
Gela	720	.19	.34	.22	.10	.03	.12
English	1321	.15	.16	.2	.30	.06	.14
Finnish	44040	.23	.21	.31	.25	.00	.01
Estonian	15472	.19	.23	.36	.17	.01	.04

Figure 1: Word-final stop inventory given [t p k] onsets

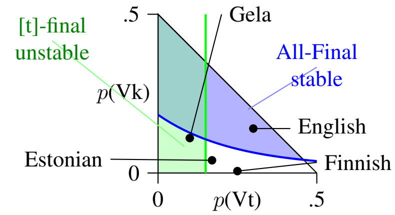


Gela. In the Oceanic language family, some languages like Mota (Vanuatu) have maintained the all-finals pattern that was present in Proto-Oceanic, but other languages like Gela (Solomon Islands) have lost all final stops. Crucially, no Oceanic language shows the [t]-final pattern. To model the lexical frequencies of stops before Gela lost them, proto-forms from the Comparative Austronesian Dictionary (Blust & Trussel, 2010 (2018)) were used. Simulations reveal that the inflow to the [t]-final pattern is large (.7), but the outflow is greater (1.0), predicting that Oceanic languages would rarely exhibit the [t]-final pattern. Thus this model matches the observed result that Oceanic languages show both the all-final and no-final patterns.

Figure 2: Transitional Probabilities that lead to underrepresentation of [t] Final

AllFinal	Gela: .78	[t]-Final	Gela: 1.0	NoFinal
tV pV kV	— English: .32 →	tV pV kV	— English: .02 →	tV pV kV
Vt Vp Vk	Finnish: 1.0	Vt ~~Vp~~ ~~Vk~~	Finnish: .06	~~Vt~~ ~~Vp~~ ~~Vk~~
	Estonian: 1.0		Estonian: .26	

Figure 3: Abstract representation of how lexical frequency affects stability



English. Compared to Gela, child directed speech in English (based on Bernstein-Ratner (1987); Brent & Cartwright (1996)) has a higher frequency of final [t], perhaps partially due to a large number of coronal final affixes. Unlike the Oceanic languages, it appears that all dialects of English and closely related languages have maintained the all-final pattern. Simulations again capture the trajectory of English. Even though the [t]-final pattern is more stable than the all-final pattern with English frequencies, English is more stable in the all-final pattern than any other language considered. As the amount of iterations increases, the inflow to the [t]-final pattern becomes effectively nonexistent, preventing the loss of final consonants in English and related languages. Thus, English thus falls beyond a sort of upper bound of the [t]-final sweet spot, having too many final stops prevents a language from ever decaying into the [t]-final pattern.

Finnish. Like English, Finnish has many coronal final affixes, leading to a high number of final [t]s. While Proto-Finnic had final [p] and [k], as observed today in Estonian, the varieties of Finnish have all lost non-coronal consonants word-finally (Itkonen, 1964). This [t]-final pattern appears to be stable since at least Old Written Finnish in the 1500s (Agricola, 1542 (2014)). The amount of final [t] present in Finnish results in stability of the [t]-final pattern in simulations, but the near absence of final [k] makes the All-final pattern quite unstable.

Estonian. To check that the frequencies of Proto-Finnic languages fall in the sweet-spot to produce the [t]-final pattern, type frequencies were extracted from several corpora of child-directed Estonian (Argus 1998, a.o). Simulations reveal that Estonian is stable in the [t]-final pattern, leading to little outflow, much like English and Finnish. However, the all-final pattern is much less stable in Estonian than English, resulting in inflow to the [t]-final pattern, that remains stable there. This is because Estonian has much less final [k] than English, leading to rarer data that helps learn the all-final pattern.

Conclusion. Figure 3 abstractly represents the space of lexical frequencies. Here, the influence of frequencies other than final [t] and [k] are factored out for ease of illustration. In order for a language to be likely to be (or become) the [t]-final pattern, it must be in the sweet-spot for the [t]-final pattern, represented by the white region. This region is bounded on the left by the green shaded region, languages that do not have enough final [t] to prevent outflow from the [t]-Final pattern. The sweet spot is bounded above by the blue region which represents those patterns that have enough final [k] to prevent inflow from the All-Final pattern. Because the sweet spot is small, narrow and far from the origin, it is unlikely that many languages will be in the sweet spot, explaining the underrepresentation of the [t]-final pattern.

References

AGRICOLA, MIKAEL. 1542 (2014). ABC-kirja ja Rukouskirjan alkuosa raamatullisten rukousten loppuun s. 1 - 344 riville 7. [tekstikorpus].

ARGUS, REILI. 1998. CHILDES'i eesti andmepank ja selle suhtluskeskne analüüs (Hendrik, 1.6-2.6). Tallinn: Tallinna Pedagoogikaülikool.

BERNSTEIN-RATNER, N. 1987. The phonology of parent-child speech. In: NELSON, K., & VAN KLEECK, A. (eds), Children's Language, vol. 6. Hillsdale, NJ: Erlbaum.

BLUST, ROBERT, & TRUSSEL, STEPHEN. 2010 (2018). Austronesian Comparative Dictionary, web edition.

BRENT, M.R., & CARTWRIGHT, T.A. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition, **61**, 93–125.

DE LACY, PAUL. 2006. Markedness: Reduction and Preservation in Phonology. Cambridge: Cambridge University Press.

HUGHTO, CORAL. 2018. Investigating the Consequences of Iterated Learning in Phonological Typology. In: Proceedings of the Society for Computation in Linguistics, vol. 1.

ITKONEN, TERHO. 1964. Proto-Finnic Final Consonants: Their History in the Finnic Languages with Particular Reference to the Finnish Dialects. Suomalais-ugrilaisen Seuran toimituksia, vol. 138, no. 1. Helsinki.

KEAN, MARY-LOUISE. 1975 (June). The Theory of Markedness in Generative Grammar. Ph.D. thesis, Massachusetts Institute of Technology.

PATER, JOE, & MORETON, ELLIOTT. 2012. Structurally biased phonology: complexity in language learning and typology. The EFL Journal, **3**(2), 1–44.

STANTON, JULIET. 2016. Learnability shapes typology: the case of the midpoint pathology. Language, **92**(4), 753–791.

STAUBS, ROBERT. 2014. Computational modeling of learning biases in stress typology. Ph.D. thesis, University of Massachusetts Amherst, Amherst.