Subregular Languages – Pop Math Article  
Jacob Johnson

1. Hook: I will present the reader with an example from Lai (2015) and Avcu & Hestvik (2020) as if the reader were a participant in these studies. The rest of the article will build up to the point where the reader could understand the idea behind this study.

2. Bridge: I will ask them to consider the way they approached it and start to motivate this study with English language examples of “plausible nonwords” (i.e., words that have no meaning but sound like they could be an English word) and “implausible nonwords” (i.e., words that do not sound like English words; they violate some English phonotactics rule). I will hint at the connection between patterns, computation, and language and suggest that these studies illustrate something important, without directly explaining the significance of the studies. Rather, I will use the studies and the other examples I mention throughout the article to motivate the broader topics they will be learning.

3. Exploration: I will first provide an intro/refresher on Phonotactics (for those without a Linguistics background) and Computability (for those without a Computer Science background). Throughout both, I will likely not give academic peer-reviewed sources (because this is very introductory material) but if I can find something I will try to point them to some general sources, like a textbook or lecture slides if I can find it.

Intro to Phonotactics: I will revisit my own examples from the Introduction and point out the relevant English Language phonotactics rule that they violate. I will then go on to share some examples of other phonotactics rules from other languages, citing Santa Ynez & Applegate (1972) and Hyman (1995)/Walker (2000) for examples.

Intro to Computability: I will introduce the idea of Finite-State-Automata and the idea of “regular languages” (i.e. patterns that can be detected by Finite-State-Automata) and briefly introduce that set of languages as only one of a wide range of patterns in the “Chomsky Hierarchy” (I will find a textbook or at least Wikipedia article to cite). I will bring up Grammatical Inference (building a grammar based on samples), citing Gold (1967) and and De La Higuera (2010). Then I will enter the subregular end of the complexity hierarchy.

Subregular Languages: This is where my article will begin to most look like a pop version of a literature review. I will provide characterizations of the Phonotactic rules from above as Finite-State-Automata and describe how examples like these have motivated the subregular hypothesis. Throughout, I will cite Heinz, Rawal, & Tanner (2011) for the first paper in this field, as well as Jardine&Heinz (2016) (for the first learning algorithms); De Santo & Graf (2019) (for more advanced subsets of these languages); Aksënova (2020) (for the implementation and evaluation of these learning algorithms); De Santo & Aksënova (2021) (for another advanced subset of these languages, with a corresponding learning algorithm); Lambert, Rawski, & Heinz (2021) (for more formal characterizations and analyses of these languages and especially Lambert (2021) and McMullin, Aksënova, and De Santo (2019) (for Learning algorithms for these languages). All of these, together, give recent ideas from Grammatical Inference in Subregular Languages. From these last two, I will focus on their characterizations of “Multiple Tier-based Strictly Local” (MTSL) language grammatical inference (which the reader will understand by the time they have read the preceding material). I will not include the proofs from the papers, but the readers can read on in the original papers.

4. Point: I will return to the experiment from the hook of the article and explain what the researchers predicted, based on the subregular hypothesis, and invite the reader to compare their answers with the researchers’ predictions. I will point out to the reader that it is interesting what results you get when you try to mathematically characterize a system even as messy as language.

Sources:

Aksënova, Alëna. Tool-assisted induction of subregular languages and mappings. Diss. State University of New York at Stony Brook, 2020.

Avcu, Enes, and Arild Hestvik. "Unlearnable phonotactics." Glossa: a journal of general linguistics 5.1 (2020).

De la Higuera, Colin. Grammatical inference: learning automata and grammars. Cambridge University Press, 2010.

De Santo, Aniello, and Alëna Aksënova. "Learning Interactions of Local and Non-Local Phonotactic Constraints from Positive Input." Proceedings of the Society for Computation in Linguistics 2021. 2021.

De Santo, Aniello, and Thomas Graf. "Structure sensitive tier projection: Applications and formal properties." Formal Grammar: 24th International Conference, FG 2019, Riga, Latvia, August 11, 2019, Proceedings 24. Springer Berlin Heidelberg, 2019.

Gold, E. Mark. "Language identification in the limit." Information and control 10.5 (1967): 447-474.

Heinz, Jeffrey, Chetan Rawal, and Herbert G. Tanner. "Tier-based strictly local constraints for phonology." Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human language technologies. 2011.

Hyman, Larry M. "Nasal consonant harmony at a distance the case of Yaka." Studies in African Linguistics 24.1 (1995): 6-30.

Jardine, Adam, and Jeffrey Heinz. "Learning tier-based strictly 2-local languages." Transactions of the Association for Computational Linguistics 4 (2016): 87-98.

Lai, Regine. "Learnable vs. unlearnable harmony patterns." Linguistic Inquiry 46.3 (2015): 425-451.

Lambert, Dakotah. "Grammar interpretations and learning TSL online." International Conference on Grammatical Inference. PMLR, 2021.

Lambert, Dakotah, Jonathan Rawski, and Jeffrey Heinz. "Typology emerges from simplicity in representations and learning." Journal of Language Modelling 9 (2021).

McMullin, Kevin, Alëna Aksënova, and Aniello De Santo. "Learning phonotactic restrictions on multiple tiers." Proceedings of the Society for Computation in Linguistics 2.1 (2019): 377-378.

Santa Ynez Band of Chumash Mission Indians of the Santa Ynez Reservation, California, and Richard B. Applegate. Samala-English dictionary: a guide to the Samala language of the Ineseno Chumash People. Santa Ynez Band of Chumash Indians, 2007.

Walker, Rachel. "Yaka nasal harmony: Spreading or segmental correspondence?." Annual meeting of the Berkeley linguistics society. Vol. 26. No. 1. 2000.