Proceedings of the Society for Computation in Linguistics

Volume 4 Article 52

2021

Deep neural networks easily learn unnatural infixation and reduplication patterns

Coleman Haley
coleman.c.haley@gmail.com

Colin Wilson
Johns Hopkins University, colin.wilson@jhu.edu

Deep neural networks easily learn unnatural infixation and reduplication patterns

Coleman Haley
Independent Researcher
coleman.c.haley@gmail.com

Colin Wilson
Johns Hopkins University
colin.wilson@jhu.edu

1 Introduction

Morphological patterns can involve simple concatenation of fixed strings (e.g., *unkind*, *kindness*) or 'nonconcatenative' processes such as infixation (e.g., Chamorro *lumiʔeʔ* 'saw (actor-focus)', Topping, 1973) and reduplication (e.g., Amele *babagawen* 'as he came out', Roberts, 1987), among many others (e.g., Anderson, 1992; Inkelas, 2014). Recent work has established that deep neural networks are capable of inducing both concatenative and nonconatenative patterns (e.g., Kann and Schütze, 2017; Nelson et al., 2020). In this paper, we verify that encoder-decoder networks can learn and generalize attested types of infixation and reduplication from modest training sets. We show further that the same networks readily learn many infixation and reduplication patterns that are unattested in natural languages, raising questions about their relationship to linguistic theory and viability as models of human learning.

2 Infixation

Broad cross-linguistic surveys have identified a small number of edge-anchored positions at which infixes can be located (e.g., *before the first vowel* of the word; Yu, 2007). No known language places an infix such as *-um-* consistently after the *second* or *third segment*, or after the *second* or *third vowel* — patterns that are trivial to describe formally and that involve limited counting or memory — and no language places a non-reduplicative infix such as *-mu-* after the first Onsetless syllable (McCarthy and Prince, 1993). We implemented a generic encoder-decoder network with OpenNMT-py (Klein et al., 2017) and tested its ability to learn and generalize two attested infixation patterns and eight unattested patterns (see Table 1).

Because data is scarce for many languages with infixation and most of the patterns are hypothetical, we implemented the patterns using the phonological forms of 7000+ Spanish lemmas. From a larger set of lemmas (Sagot, 2018), we eliminated those that contained triconsonantal clusters, non-initial onset clusters, word-final coda clusters, or glide-vowel sequences. This made the syllable structure of the remaining lemmas somewhat simpler than that of Spanish, streamlining the definition and implementation of infixation (and reduplication) patterns. The remaining lemmas were randomly partitioned into 1000 train and 6000+ test inputs. For each infixation pattern, a custom regular expression was used to create outputs from the inputs.

There is some ambiguity in how unattested patterns that reference the second or third vowel (or consonant) should apply to inputs that do not contain the designated pivot. On one interpretation, the infix should 'back off' to the immediately preceding unit of the same type (e.g., appearing after the first vowel in a monosyllabic form). This is analogous to a stress pattern that typically targets the penultimate syllable but assigns final stress in monosyllables. Alternatively, the infix could default to a suffix. Under this interpretation, the infix necessarily 'skips' all elements until reaching its pivot, landing at the rightmost position when the pivot is absent.[1] Table 1 represents the results for the back-off interpretation while results for the skipping interpretation are provided in the appendix.

The model had an embedding dimension of 50 (approximately twice the number of unique phonological segments that appeared in the lemmas), a single-layer bidirectional LSTM or GRU encoder with 100 units, and a single-layer attention-based decoder with 100 units and copy attention. Ten simulations were conducted for each of several resource conditions (1000, 500, 100, or 50 of the

[1]This ambiguity is negligible for patterns referencing the first vowel or consonant, or one of the first three segments, which are essentially always present in the input.

Proceedings of the Society for Computation in Linguistics (SCiL) 2021, pages 427-433.
Held on-line February 14-19, 2021

Pattern	Train size:	LSTM 1000	500	100	50	GRU 1000	500	100	50
Before first V	(e.g., Chamorro *lumiʔeʔ*)	1.0	1.0	.98	.95	1.0	1.0	.98	**.92**
After first C	(e.g., Tagalog *gumradwet*)	1.0	1.0	1.0	.99	1.0	1.0	1.0	1.0
Before second V	C*VC*umV...	**.93**	.96	**.87**	**.73**	.98	.97	**.90**	**.87**
Before third V	C*VC*VC*umV...	**.89**	**.89**	**.82**	**.79**	**.92**	**.91**	**.85**	**.75**
After second C	V*CV*Cum...	.99	.99	.98	**.94**	.99	.99	.98	.95
After third C	V*CV*CV*Cum...	.97	.97	**.93**	**.78**	.97	.98	**.93**	**.79**
After second segment	XXum...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
After third segment	XXXum...	1.0	1.0	1.0	.99	1.0	1.0	1.0	1.0
After fourth segment	XXXXum...	1.0	1.0	1.0	.98	1.0	1.0	1.0	.99
After initial Onsetless syll.	(VC*)muCV...	.96	.95	**.90**	**.90**	1.0	1.0	.95	**.88**

Table 1: Average held-out test accuracy for attested (top) and unattested (bottom) infixation patterns

input/output training pairs sampled without replacement). In each simulation, the model was trained for 1000 epochs using Adagrad with an initial learning rate of 0.01. Most of the average test accuracies were above .95; bold cells identify the others.

With LSTM units, the model performed very highly on almost all of the infixation patterns when given at least 500 examples in training. The principal exception was infixation before the third vowel. For this pattern, the model nearly always produced an output that contained the infix (more than 99% of responses), but erred by placing the infix in the wrong position (either too early or to late in the output) or making spurious changes to the base of infixation (primarily deletion of one or two segments).

Performance degraded for many patterns in the very low resource conditions of 100 or 50 training examples, but not in a way that cleanly observed the attested vs. unattested divide. Counting patterns that depend on the distinction between consonants and vowels (e.g., after the third consonant) were difficult to learn from 50 examples, but unattested segment-counting infixation were learned at least as well as the Chamorro pattern. The results were overall similar with GRU units, notwithstanding their relatively limited counting abilities (Weiss et al., 2018).

3 Reduplication

We also compared learning performance for five attested reduplication patterns and five patterns that have been discussed in the theoretical literature as unattested and putatively impossible (see Table 2). The attested patterns included both light and heavy syllable reduplication, as in Amele and Ilokano respectively, and Foot reduplication as in Yidiɲ. We also considered patterns in which the reduplicant is reduced relative to the base, either by Onset simplification as in one pattern of light-syllable reduplication in Tagalog or by omission of any Foot-final Coda as in Dyirbal. The unattested patterns included copying of an initial string of segments, regardless of its prosodic composition (e.g., Marantz, 1982; McCarthy and Prince, 1993), and Foot reduplication patterns that differ minimally from Dyirbal in omitting internal or both internal and final Codas (e.g., McCarthy et al., 2012).

The same set of Spanish lemmas described above, split into 1000 train and 6000+ test inputs, was used to implement the reduplication patterns. As before, a custom regular expression was written for each pattern; when the base was shorter than the target reduplicant size, the default was always full copy (i.e., complete reduplication). Simulations were performed in the same way as for infixation.

The results indicate that unattested segment-counting patterns, such as reduplication of the first four segments, are at least as easy for the model to learn as attested syllable- and Foot- based patterns. Indeed, the average accuracy for segment counting was higher than for attested patterns in the low-resource conditions. While model performance on Foot reduplication with deletion of internal Codas was numerically lower than that of the two attested Foot-reduplication patterns, accuracy on Foot reduplication with omission of all Codas was higher than for intact Foot reduplication as in Yidiɲ.

4 Ultra-low resource experiments

To better understand which patterns are easiest for the model to learn, we ran additional simulations

Pattern	Train size:	LSTM 1000	LSTM 500	LSTM 100	LSTM 50	GRU 1000	GRU 500	GRU 100	GRU 50
Initial σ_μ	(e.g., Amele *babagawen*)	1.0	1.0	.96	.95	1.0	1.0	.97	**.92**
Initial σ_μ with Onset simplification	(e.g., Tagalog *tatrabaho*)	1.0	1.0	.96	**.92**	1.0	1.0	.97	**.92**
Initial $\sigma_{\mu\mu}$	(e.g., Ilokano *kalkaldiŋ*)	**.92**	**.94**	**.90**	**.84**	.95	.95	**.93**	**.81**
Initial Foot with deletion of final Coda	(e.g., Dyirbal *balgabalgan*)	.99	.99	.98	**.91**	1.0	.99	.99	**.93**
Initial Foot	(e.g., Yidiŋ *ŋalalŋalal*)	**.94**	**.94**	**.89**	**.75**	.99	.99	**.88**	**.71**
Initial 2 segments	$\underline{X_1X_2}X_1X_2\ldots$	1.0	1.0	1.0	.99	1.0	1.0	1.0	1.0
Initial 3 segments	$\underline{X_1X_2X_3}X_1X_2X_3\ldots$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Initial 4 segments	$\underline{X_1X_2X_3X_4}X_1X_2X_3X_4\ldots$	1.0	1.0	.98	.98	.98	.98	.98	.97
Initial Foot with deletion of internal Coda	(e.g., pseudo-Dyirbal *baganbalgan*)	**.90**	**.89**	**.85**	**.73**	.98	.99	**.90**	**.70**
Initial Foot with deletion of all Codas	(e.g., pseudo-Dyirbal *bagabalgan*)	.99	.99	.97	**.84**	.99	.99	.98	**.86**

Table 2: Average held-out test accuracy for attested (top) and unattested (bottom) reduplication patterns

using LSTMs only on extremely low-resource settings of 5, 10, 25, 50, 75, or 100 examples, with 10 simulation runs each. Otherwise, identical settings to the previous experiments were used. Results on the held-out test set are summarized in Figure 1.

These results largely confirm the patterns observed in the previous experiments. While the spread of accuracies is quite wide for all patterns for 25 or fewer examples, the spread is typically small with higher amounts of training data, varying by about 10% or less. Notably, the unattested segment-counting patterns (e.g., infixation after the first two segments), are learned reliably with the least data for both infixation and reduplications, with average model accuracy above 0.8 with just 25 example. All other types of patterns (involving prosodic feet, syllables, consonants, or vowels), require at least 50 examples for this type of accuracy, suggesting that these patterns are in fact *easier* for the model to learn than those which are attested. No patterns could be learned reliably with fewer than 25 examples.

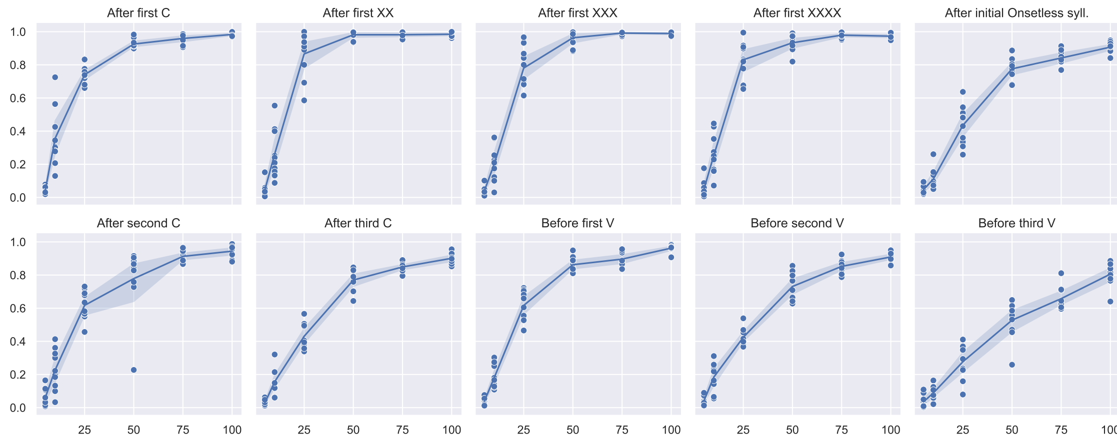
5 Discussion

A large body of research in theoretical linguistics has sought to develop restrictive approaches to infixation and reduplication (e.g., Moravcsik, 1978; Marantz, 1982; Steriade, 1988; McCarthy and Prince, 1986/1996, 1993; Raimy, 2000; Yu, 2007; McCarthy et al., 2012). For example, the nonexistence of reduplication patterns that consistently copy the first k segments provided motivation for the early C/V skeleton approach (Marantz, 1982) and for the even more restrictive framework of Prosodic Morphology (e.g., McCarthy and Prince, 1993). From the perspective of language acquisition, the principles of such theories can be considered as limits on the implicit hypothesis space considered by human learners.

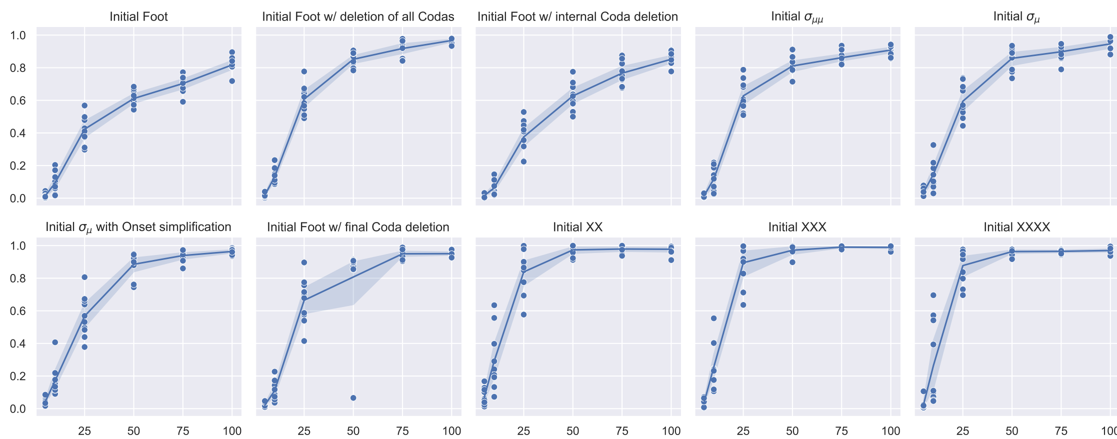
Our results suggest that such limits may be unnecessary for successful learning of infixation and reduplication, as long as a relatively modest amount of input data is available to the learner, and that networks of the kind considered here do not have soft biases that systematically favor natural over unnatural patterns. This could provide support for theories that eschew strong synchronic restrictions on morphophonological patterns, and which therefore presuppose robust learning mechanisms (e.g., ble). Alternatively, artificial-grammar or other experiments may reveal that human pattern learning is limited or biased in ways that generic deep neural networks cannot explain. For example, the networks show a preference for learning segment-counting patterns over patterns that take prosodic structure into account, despite these patterns being unattested. As such, these models do not provide an account of this apparent bias in human language.

Acknowledgments

We are grateful to the anonymous SCiL reviewers for useful feedback on this work.



(a) Infixation pattern accuracy



(b) Reduplication pattern accuracy

Figure 1: Held out test set accuracies in an ultra-low resource context (less than 100 examples). Each point represents a simulation run.

References

Evolutionary Phonology. New York.

Steven Anderson. 1992. *A-Morphous morphology*. Cambridge University Press, Cambridge.

Sharon Inkelas. 2014. *The interplay of morphology and phonology*. Oxford University Press, Oxford.

Katharina Kann and Hinrich Schütze. 2017. The LMU system for the CoNLL-SIGMORPHON 2017 shared task on universal morphological reinflection. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 40–48, Vancouver. Association for Computational Linguistics.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. OpenNMT: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada. Association for Computational Linguistics.

Alec Marantz. 1982. Re reduplication. *Linguistic Inquiry*, 13(3):435–482.

John J McCarthy, Wendell Kimper, and Kevin Mullin. 2012. Reduplication in harmonic serialism. *Morphology*, 22(2):173–232.

John J McCarthy and Alan Prince. 1986/1996. Prosodic Morphology 1986. *Linguistics Department Faculty Publication Series*, 14.

John J. McCarthy and Alan Prince. 1993. *Prosodic morphology I: Constraint interaction and satisfaction*. Rutgers University Center for Cognitive Science, New Brunswick.

Edith Moravcsik. 1978. Reduplicative constructions. In Joseph Greenberg, editor, *Universals of Human Language*, pages 297–334 (vol. 3). Stanford University Press, Stanford.

Max Nelson, Hossep Dolatian, Jonathan Rawski, and Brandon Prickett. 2020. Probing RNN encoder-decoder generalization of subregular functions using reduplication. In *Proceedings of the Society for Computation in Linguistics 2020*, pages 167–178, New York, New York. Association for Computational Linguistics.

Eric Raimy. 2000. *The phonology and morphology of reduplication*, volume 52. Walter de Gruyter, Berlin.

John R Roberts. 1987. *Amele*. Croom Helm, London.

Benoît Sagot. 2018. A multilingual collection of CoNLL-U-compatible morphological lexicons. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Donca Steriade. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology*, 5(1):73–155.

D. Topping. 1973. *Chamorro Reference Grammar*. University Press of Hawaii, Honolulu.

Gail Weiss, Yoav Goldberg, and Eran Yahav. 2018. On the practical computational power of finite precision RNNs for language recognition. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 740–745, Melbourne, Australia. Association for Computational Linguistics.

Alan C. L. Yu. 2007. *A natural history of infixation*. Oxford University Press, Oxford.

A Appendix A: Additional infixation results

		LSTM				GRU			
Pattern	Train size:	1000	500	100	50	1000	500	100	50
Before second V		**.90**	**.93**	**.89**	**.64**	.98	.97	**.80**	**.72**
Before third V		**.83**	**.88**	**.80**	**.62**	**.85**	**.85**	**.83**	**.68**
After second C		1.0	.99	.96	.95	1.0	1.0	.97	.97
After third C		.98	.98	**.92**	**.79**	.97	.98	**.94**	**.76**

Table 3: Average held-out test accuracy for unattested infixation patterns