Quantifying the Tradeoff Between Two Types of Morphological Complexity

Ryan Cotterell[1] **Christo Kirov**[1] **Mans Hulden**[2] **Jason Eisner**[1]
[1]Johns Hopkins University [2]University of Colorado

Introduction. A number of metrics have been proposed for determining the complexity of a language's inflectional morphology. However, many of the metrics put forth are subject to sources of bias that can significantly skew the results. We argue that modern machine learning methods can greatly reduce at least one major source of bias: the analytical bias introduced by linguists when formally describing the transformations relating different inflected forms that comprise a paradigm. In short, we propose a black-box approach to morphological complexity: a language is *less complex* if it is *easier to learn* with a domain-general string-to-string transduction method.

Neural networks have shown state-of-the-art performance in learning and generalizing string-to-string morphological transformations (Cotterell et al., 2017). We argue that how well these networks predict held-out data is a good measure for the complexity of an inflection. We provide results on 31 typologically diverse languages as evidence for a more general version of the low-entropy complexity hypothesis first proposed in Ackerman and Malouf (2013). We show that morphological systems evince a type of Pareto complexity: a language's inflectional paradigms may be either large in size or highly irregular, but never both.

Defining Morphological Complexity. Ackerman and Malouf (2013) introduce a distinction between two types of morphological complexity.[1] The first type, enumerative complexity (e-complexity), is the number of morpho-syntactic distinctions a word overtly marks (the number of morpho-syntactic functions times the number of unique exponents for each function). E-complexity varies dramatically across languages. While the regular English verb paradigm contains four forms, the Archi verb will have thousands. Despite the plethora of forms, however, the Archi system is highly regular. Integrative complexity (i-complexity) is based on the idea of quantifying this predictability. Building on Ackerman and Malouf (2013), we repose our measure of i-complexity on information theory. We estimate the entropy of the distribution from which an entire paradigm is generated, by assuming it takes the form of a tree-structured graphical model with neural network factors (Fig. 2).

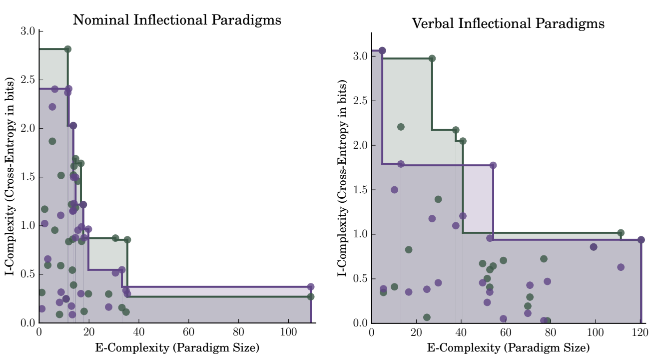

Figure 1: The y-axis is cross-entropy, an approximation to the paradigm entropy and i-complexity. The x-axis is the size of the paradigm, a measure of e-complexity. The purple points are languages with models trained with a fixed number of observed paradigms. The green points are languages with models trained with a fixed number of observed individual morphological transformations.

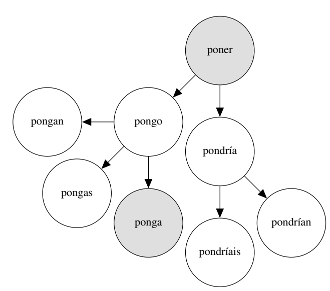

Figure 2: A partial directed graphical model for a paradigm completion task with Spanish verbs. The individual predictions are made by a neural network trained to predict inflected forms from other inflected forms.

[1]For a more general overview of morphological complexity, we refer the reader to Baerman et al. (2015).

Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 209-210.
Salt Lake City, Utah, January 4-7, 2018

Avoiding Bias in Complexity Calculations. Previous attempts to calculate morphological complexity have relied on linguists to analyze the morphological distinctions in a language, but this can lead to theory-dependent calculations. For example, in regular English plural formation, the speaker has three choices: [z], [s] and [ɨz]. One the one hand, we may treat this as a case of three unrelated suffixes. Under such an analysis, the entropy will reflect their empirical distribution: roughly, $1/4 \log 1/4 + 3/8 \log 3/8 + 3/8 \log 3/8 \approx 1.56127$. On the other hand, if we assume a unique underlying affix /z/, which is attached and then converted to either [z], [s] or [ɨz] by an application of perfectly regular phonology, this part of the morphological system of English has an entropy of 0. We attempt to mitigate this problem by leaving all analyses (beyond initial data selection) to be discovered by the neural model. The networks themselves operate on full unanalyzed strings. While it is true that neural nets, like all models that can generalize to new data, *must* have learning biases, we argue that by using the same model for all languages, we can at least ensure that any bias is cross-linguistically consistent.

Model and Results. Starting with raw paradigm tables, we apply the model of Kann and Schütze (2016), to learn morphological transformations in 31 languages. The model is a recurrent sequence-to-sequence neural architecture with attention (Bahdanau et al., 2015).[2] For each language, a model was trained on input-output pairings like the following:

```
IN_NUM=SG OUT_NUM=PL c a t → c a t s
```

The input features indicate the morpho-syntactic inflection of the input form, while the output features indicate the desired morpho-syntax of the output form to be produced by the network. In Fig. 1, we plot the paradigm size for each language (its e-complexity) against our entropy-based estimate of i-complexity. A clear trade-off is visible: most languages exhibit low entropy with moderate paradigm sizes. A few languages have small but more irregular paradigms, or large but highly regular paradigms. Languages with large *and* unpredictable paradigms do not seem to exist (a non-parametric test confirms that the upper-right quadrant is not likely to be empty by chance, with $p < 0.05$).

Conclusions. New methods that emerge from statistical pattern recognition and deep learning tend to find their way into natural language processing (NLP) relatively quickly (Goldberg, 2017). On the other hand, it takes longer for such methods to trickle their way into linguistics, the scientific study of language. In this work, we take a small step to encourage a tighter integration of NLP and linguistics: we apply modern machine learning to theoretical issues in morphology with a cross-lingual study using recurrent neural graphical modeling to empirically determine the complexity of inflectional systems. To the best of our knowledge, this is the largest analysis on morphological complexity in the literature.

References

Ackerman, F. and Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*.

Baerman, M., Brown, D., and Corbett, G. G. (2015). Understanding and measuring morphological complexity: An introduction.

Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations (ICLR)*.

Cotterell, R., Kirov, C., Sylak-Glassman, J., Walther, G., Vylomova, E., Xia, P., Faruqui, M., Kübler, S., Yarowsky, D., Eisner, J., and Hulden, M. (2017). The CoNLL-SIGMORPHON 2017 shared task. In *CoNLL-SIGMORPHON 2017 Shared Task*.

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*.

Kann, K. and Schütze, H. (2016). Single-model encoder-decoder with explicit morphological representation for reinflection. In *ACL*.

[2]All networks had 100 hidden units in both encoder and decoder, 300-unit embeddings for input symbols. They were trained using Adadelta (learning rate 1 and dropout 0.3). Decoding was performed using beam search with beam size 12.