USING HMMS TO ATTRIBUTE STRUCTURE TO ARTIFICIAL LANGUAGES

KEREM ERYILMAZ, HANNAH LITTLE, BART DE BOER
Artificial Intelligence Lab, Vrije Universiteit Brussel
Brussels, Belgium
kerem@ai.vub.ac.be

Artificial languages have gained popularity as a way of empirically testing hypotheses on language evolution and language change (Scott-Phillips & Kirby, 2010), distancing the participants from the biases of their native languages as much as possible. Literature on artificial languages already contains studies on identifying the building blocks in repertoires of continuous signals. However, these rely on manual methods for identifying subparts of a signal that either use quantitative judgement or enforced boundaries between segments (Verhoef, Kirby, & Boer, 2013; Galantucci, 2005). Since these techniques build on the qualitative judgement of the researcher, they are time-consuming and difficult to replicate. We propose using computational models for this task.

The speech recognition literature uses positional HMMs' states as building blocks that stand in for different positions in a syllable (Rabiner & Juang, 1993), whereas the artificial languages literature tends to use segments of signals as building blocks (see Verhoef et al., 2013 for an example). As an alternative, we propose using computational models trained on an entire repertoire of signals in order to infer building blocks, and not imposing a pre-determined positional structure on the model so that the model structure is learned from the data. The formal process used guarantees reproducibility of the results, and creates an objective basis of comparison for signal repertoires created during artificial language experiments, both of which are improvements to the state of the art in artificial languages research to our knowledge.

We use HMM states as analogues for *areas* in the signal space. This approach is similar to the vowel quality literature which makes heavy use of the vowel triangle on the formant space, on which different areas correspond to different vowels. The extent of each area is defined by the covariance of the area's emission distributions. We use the number of states inferred as a measure of structure in the repertoire being modeled, analogous to phonetic inventory size.

The data is taken from an artificial language experiment, where participants create continuous acoustic signals using hand sensors to label a series of meanings (Little, Eryılmaz, & Boer, 2015). The participants are tested on their recall of the repertoire of signals they created using a forced-choice task at the end of each phase. The signals can be either one- or two-dimensional, with dimensions corresponding to pitch and amplitude of the signal derived from the coordinates of the participant's hand. The meanings are also "dimensional" in that they differ along one or more feature dimensions such as size or colour. The features can be discrete (e.g. simple polygons of different number of vertices) or continuous (e.g. real numbers). The combination of these two spaces uniquely defines a "phase", and the collection of signals that belong to that combination form a "repertoire". The training data comprises a series of real numbers or tuples of real numbers, corresponding to the series of amplitude and frequency values of the tones played to the participants.

The emission distributions of the Gaussian HMMs used are either univariate or bivariate continuous distributions, depending on the dimensionality of the repertoire. Models are trained using the Baum-Welch algorithm, and the number of states are selected by training HMMs of various sizes and picking the one with the lowest Bayesian Information Criterion. The validation of the models is done using linear mixed-effects regression which uses the number of states in the HMM i.e. the "inventory size" of the repertoire, and its interaction with the properties of each phase to predict the recall of the participant at each phase.

The regression model outperformed the baseline, and explained more than half the variation in participant scores ($R^2 = 0.616$), demonstrating the validity of our model. It also matched our predictions that:

1. Continuous meaning spaces are more suitable for iconicity.
2. Discrete meaning spaces are more robust against variation.
3. Uninterrupted strategies outperform interrupted strategies.

While this demonstrates our model captures some aspects of the phenomenon, a third subset of the data encouraging participants to use "duration" as a signal dimension could not be validly modelled. This is most likely due to HMMs' inability to explicitly model time spent in a certain state. To remedy this, we plan to use more advanced latent models such as Explicit Duration HMMs that eliminate self-transitions, or Hidden Semi-Markov Models, which relax the Markovian assumption, in order to estimate duration distributions for the states.

Acknowledgements

Financial support from the ERC starting grant, ABACUS, project number 283435.

References

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Little, H., Eryılmaz, K., & Boer, B. de. (2015). Linguistic modality affects the creation of structure and iconicity in signals. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, & P. Maglio (Eds.), *The 37th annual meeting of the cognitive science society (cogsci 2015)* (p. 1392-1398). Austin, TX: Cognitive Science Society.

Rabiner, L., & Juang, B.-H. (1993). Fundamentals of speech recognition.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, *14*(9), 411–417.

Verhoef, T., Kirby, S., & Boer, B. de. (2013). Combinatorial structure and iconicity in artificial whistled languages. In *Proc. cogsci* (Vol. 35).