

The typology of emergent languages

Edoardo M. Ponti, Dieuwke Hupkes, Elia Bruni

The emergence of communication in artificial agents is key to both achieving meaningful machine-machine and human-machine interactions and to shedding light on the evolutionary origins of the human faculty of language. This has led to the simulation of environments where artificial agents are encouraged to develop a communication protocol to perform some collaborative task (Kirby, 2002). Reinforcement learning and neural network models have been proven effective in inducing this sort of communication (Havrylov et al, 2017, *inter alia*).

An essential thread in this research consists in understanding how different task structures, number of agents, and levels of agent cooperation shape the semiotic properties of the emergent languages. Previous work has mostly focussed on compositionality, whereas we investigate the presence of other core properties of natural languages, such as lexical classes and dependency relations. These properties regulate the symbol ordering (syntax) and symbol reference (semantics). Words can be divided in lexical classes according to prototypical referents: nouns denote objects, verbs denote actions (Croft, 1991). Moreover, words can modify other words and predicate novel information (Tesnière, 1959).

Firstly, we inspect the emergent languages from a purely distributional perspective, using unsupervised NLP algorithms for part-of-speech and grammar induction that were originally developed for natural languages. To avoid making strong assumption about the presence and the number of lexical classes and dependency relations, we resort to non-parametric unsupervised algorithms. In particular, we use Hierarchical Dirichlet Processes to capture topical relatedness and infinite Hidden Markov Models (Beal et al, 2002) to capture the syntagmatic relatedness. Secondly, we inspect the referential properties of emergent languages, by seeking correlations between symbols and their referents (such as images) through multi-modal Convolutional Neural Networks and Deep Canonical Correlation Analysis (Yan et al, 2015). The results reveal the classes of symbols and symbol structures that lie behind the emergent languages.

Beal, M. J., Ghahramani, Z., & Rasmussen, C. E. 2002. The infinite hidden Markov model. In *Advances in neural information processing systems* (pp. 577-584).

Croft, W. (1991). *Syntactic Categories and Grammatical Relations*. University of Chicago Press.

Serhii Havrylov and Ivan Titov. 2017. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In *NIPS* (pp. 2149-2159).

Simon Kirby. 2002. Natural language from artificial life. *Artificial Life*, 8:185–215.

Lucien Tesnière. 1959. *Éléments de syntaxe structurale*, Klincksieck, Paris.

Yan, F., & Mikolajczyk, K. 2015. Deep correlation for matching images and text. *CVPR* (pp. 3441-3450).