Emergence of linguistic universals in neural agents via artificial language learning and communication

Lian, Y.

Summary

Human language is constantly evolving with its linguistic structure being shaped by language users at both individual and population levels. Recent developments of powerful neural learners in AI have drawn renewed interest in simulating language evolution with agent-based modeling, which is a powerful methodology for simulating the dynamic nature of human language.

A recent, but rich body of work known as 'emergent communication' has focused on letting neural network-based agents interact and develop novel communication protocols that allow them to successfully exchange information about simplified worlds. A key research focus in this area involves comparing the emergent language protocols with high-level properties of human languages to investigate whether communicative pressures can lead to human like linguistic patterns like compositional syntax. However, these protocols are initialized by sets of random symbols, and the agents develop their own 'closed codes', which makes it intrinsically difficult to analyze their languages and compare them against human productions at the level of specific language phenomena.

To address some of the challenges surrounding the standard emergent communication setup, this thesis proposed a Neural-agent Language Learning and Communication framework (NeLLCom). A crucial innovation includes training the agents first on a pre-defined artificial language, matching methods used in lab experiments with human participants. Specifically, NeLLCom agents first learn these artificial languages individually through supervised learning, and then interact in pairs or groups through a meaning reconstruction game, where they dynamically learn from these interactions through reinforcement learning.

Thus, NeLLCom provides a general language learning and communication procedure that can be used to study many language phenomena.

In this thesis, we focused on two widely attested language phenomena: (i) the trade-off between word order flexibility and case marking, and (ii) differential case marking. In both cases, we strictly followed the artificial language design as previously studied in human experiments. Our results show that neural agents can successfully replicate both phenomena in a human-like way through communication. Moreover, our new framework allows to extend prior results with humans to a larger scale, and we were able to show that a word-order/case-marking trade-off also emerges at the group level.

Consequently, focusing on the interplay between processes of language acquisition and communicative need in shaping human languages, this thesis provided a unified framework aligning with modern approaches in computational linguistics to simulate language learning and use. This framework can be used for conducting controlled experiments, complementing experimental research with humans on language evolution, to facilitate exploring the end goal of explaining why human languages look the way they do.