Computational Modeling of Human Language Acquisition

Afra Alishahi
Department of Computational Linguistics and Phonetics
Saarland University, Germany
`afra@coli.uni-saarland.de`

1   Introduction

The nature and amount of information needed for learning a natural language, and the underlying mechanisms involved in this process, are the subject of much debate: is it possible to learn a language from usage data only, or some sort of innate knowledge and/or bias is needed to boost the process? This is a topic of interest to (psycho)linguists who study human language acquisition, as well as computational linguists who develop the knowledge sources necessary for largescale natural language processing systems. Children are a source of inspiration for any such study of language learnability. They learn language with ease, and their acquired knowledge of language is flexible and robust.

Human language acquisition has been studied for centuries, but using computational modeling for such studies is a relatively recent trend. However, computational approaches to language learning have become increasing popular, mainly due to the advances in developing machine learning techniques, and the availability of vast collections of experimental data on child language learning and child-adult interaction. Many of the existing computational models attempt to study the complex task of learning a language under the cognitive plausibility criteria (such as memory and processing limitations that humans face), as well as to explain the developmental patterns observed in children. Such computational studies can provide insight into the plausible mechanisms involved in human language acquisition, and be a source of inspiration for developing better language models and techniques.

2   Content Overview

This tutorial will discuss the main research questions that the researchers in the field of computational language acquisition are concerned with, and will review common approaches and techniques used in developing such models. Computational modeling has been vastly applied to different domains of language acquisition, including word segmentation and phonology, morphology, syntax, semantics and discourse. However, due to time restrictions, the focus of the tutorial will be on the acquisition of word meaning, syntax, and the relationship between syntax and semantics.

The first part of the tutorial focuses on some of the fundamental issues in the study of human language acquisition, and the role of computational modeling in addressing these issues. Specifically, we discuss language modularity, i.e. the representation and acquisition of various aspects of language, and the interaction between these aspects. We also review the major arguments on language learnability and innateness. We then give a general overview of how computational modeling is used for investigating different views on each of these topics, how the theoretical assumptions are integrated into computational models, and how such models are evaluated based on the experimental observations.

In the second part of the tutorial, we will take a closer look at some of the existing models of language learning. We discuss general trends in computational modeling over the past decades, including symbolic, connectionist, and probabilistic modeling. We review a number of more influential models of the acquisition of syntax and semantics, and the link between the two. Finally, we explore some of the available tools and resources for implementing and evaluating computational models of language acquisition.