Cognition 173 (2018) 43–59

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner

Emmanuel Dupoux

EHESS, ENS, PSL Research University, CNRS, INRIA, France

ARTICLE INFO

Keywords:
Artificial intelligence
Speech
psycholinguistics
Computational modeling
Corpus analysis
Early language acquisition
Infant development
Language bootstrapping
Machine learning

ABSTRACT

Spectacular progress in the information processing sciences (machine learning, wearable sensors) promises to revolutionize the study of cognitive development. Here, we analyse the conditions under which 'reverse engineering' language development, i.e., building an effective system that mimics infant's achievements, can contribute to our scientific understanding of early language development. We argue that, on the computational side, it is important to move from toy problems to the full complexity of the learning situation, and take as input as faithful reconstructions of the sensory signals available to infants as possible. On the data side, accessible but privacy-preserving repositories of home data have to be setup. On the psycholinguistic side, specific tests have to be constructed to benchmark humans and machines at different linguistic levels. We discuss the feasibility of this approach and present an overview of current results.

1. Introduction

In recent years, artificial intelligence (AI) has been hitting the headlines with impressive achievements at matching or even beating humans in complex cognitive tasks (playing go or video games: Mnih et al., 2015; Silver et al., 2016; processing speech and natural language: Amodei et al., 2016; Ferrucci, 2012; recognizing objects and faces: He, Zhang, Ren, & Sun, 2015; Lu & Tang, 2014) and promising a revolution in manufacturing processes and human society at large. These successes show that with statistical learning techniques, powerful computers and large amounts of data, it is possible to mimic important components of human cognition. Shockingly, some of these achievements have been reached by throwing out some of the classical theories in linguistics and psychology, and by training relatively unstructured neural network systems on large amounts of data. What does it tell us about the underlying psychological and/or neural processes that are used by humans to solve these tasks? Can AI provide us with scientific insights about human learning and processing?

Here, we argue that developmental psychology and in particular, the study of language acquisition is one area where, indeed, AI and machine learning advances can be transformational, provided that the involved fields make significant adjustments in their practices in order to adopt what we call the *reverse engineering approach*. Specifically:

> The reverse engineering approach to the study of infant language acquisition consists in constructing *scalable* computational systems that can, when fed with *realistic* input data, *mimic* language acquisition as it is observed in infants.

The three italicised terms will be discussed at length in subsequent sections of the paper. For now, only an intuitive understanding of these terms will suffice. The idea of using machine learning or AI techniques as a means to study child's language learning is actually not new (to name a few: Anderson, 1975; Berwick, 1985; Kelley, 1967; Langley & Carbonell, 1987; Rumelhart & McClelland, 1987) although relatively few studies have concentrated on the early phases of language learning (see Brent, 1996b, for a pioneering collection of essays). What is new, however, is that whereas previous AI approaches were limited to proofs of principle on toy or miniature languages, modern AI techniques have scaled up so much that end-to-end language processing systems working with real inputs are now deployed commercially. This paper examines whether and how such unprecedented change in scale could be put to use to address lingering scientific questions in the field of language development.

The structure of the paper is as follows: In Section 2, we present two deep scientific puzzles that large scale modeling approaches could in principle address: solving the bootstrapping problem, accounting for developmental trajectories. In Section 3, we review past theoretical and modeling work, showing that these puzzles have not, so far, received an adequate answer. In Section 4, we argue that to answer them with reverse engineering, three requirements have to be addressed: (1) modeling should be computationally scalable, (2) it should be done on

E-mail address: emmanuel.dupoux@ens.fr.
URL: http://www.syntheticlearner.net.

https://doi.org/10.1016/j.cognition.2017.11.008
Received 28 October 2016; Received in revised form 10 November 2017; Accepted 22 November 2017

realistic data, (3) model performance should be compared with that of humans. In Section 5, recent progress in AI is reviewed in light of these three requirements. In Section 6, we assess the feasibility of the reverse engineering approach and lay out the road map that has to be followed to reach its objectives, and we conclude in Section 7.

2. Two deep puzzles of early language development

Language development is a theoretically important subfield within the study of human cognitive development for the following three reasons:

First, the linguistic system is uniquely *complex*: mastering a language implies mastering a combinatorial sound system (phonetics and phonology), an open ended morphologically structured lexicon, and a compositional syntax and semantics (e.g., Jackendoff, 1997). No other animal communication system uses such a complex multilayered organization (Hauser, Chomsky, & Fitch, 2002). On this basis, it has been claimed that humans have evolved (or acquired through a mutation) an innately specified computational architecture to process language (see Chomsky, 1965; Steedman, 2014).

Second, the overt manifestations of this system are extremely *variable* across languages and cultures. Language can be expressed through the oral or manual modality. In the oral modality, some languages use only 3 vowels, other more than 20. Consonants inventories vary from 6 to more than 100. Words can be mostly composed of a single syllable (as in Chinese) or long strings of stems and affixes (as in Turkish). Semantic roles can be identified through fixed positions within constituents, or be identified through functional morphemes, etc. (see Song, 2010, for a typology of language variation). Evidently, infants acquire the relevant variant through learning, not genetic transmission.

Third, the human language capacity can be viewed as a finite computational system with the ability to generate a (virtual) infinity of utterances. This turns into a *learnability problem* for infants: on the basis of finite evidence, they have to induce the (virtual) infinity corresponding to their language. As has been discussed since Aristotle, such induction problems do not have a generally valid solution. Therefore, language is simultaneously a human-specific biological trait, a highly variable cultural production, and an apparently intractable learning problem.

Despite these complexities, most infants spontaneously learn their native(s) language(s) in a matter of a few years of immersion in a linguistic environment. The more we know about this simple fact, the more puzzling it appears. Specifically, we outline two deep scientific puzzles that a reverse engineering approach could, in principle help to solve: solving the bootstrapping problem and accounting for developmental trajectories. The first puzzle relates to the ultimate outcome of language learning: the so-called *stable state*, defined here as the stabilized language competence in the adult. The second puzzle relates to what we know of the intermediate steps in the acquisition process, and their variations as a function of language input.[1]

2.1. Solving the bootstrapping problem

The stable state is the operational knowledge which enables adults to process a virtual infinity of utterances in their native language. The most articulated description of this stable state has been offered by theoretical linguistics; it is viewed as a grammar comprising several components: phonetics, phonology, morphology, syntax, semantics, pragmatics.

The *bootstrapping problem* arises from the fact these different components appear *interdependent* from a learning point of view. For instance, the phoneme inventory of a language is defined through pairs of words that differ minimally in sounds (e.g., "light" vs "right"). This would suggest that to learn phonemes, infants need to first learn words. However, from a processing viewpoint, words are recognized through their phonological constituents (e.g., Cutler, 2012), suggesting that infants should learn phonemes before words. Similar paradoxical co-dependency issues have been noted between other linguistic levels (for instance, syntax and semantics: Pinker, 1987, prosody and syntax: Morgan & Demuth, 1996). In other words, in order to learn any one component of the language competence, many others need to be learned first, creating apparent circularities.

The bootstrapping problem is further compounded by the fact that infants do not have to be taught formal linguistics or language courses to learn their native language(s). As in other cases of animal communication, infants *spontaneously* acquire the language(s) of their community by merely being immersed in that community (Pinker, 1994). Experimental and observational studies have revealed that infants start acquiring elements of their language (phonetics, phonology, lexicon, syntax and semantics) even before they can talk (Hollich et al., 2000; Jusczyk, 1997; Werker & Curtin, 2005), and therefore before parents can give them much feedback about their progress into language learning. This suggests that language learning (at least the initial bootstrapping steps) occurs largely *without supervisory feedback*.[2]

The reverse engineering approach has the potential of solving this puzzle by providing a computational system that can demonstrably bootstrap into language when fed with similar, supervisory poor, inputs.[3]

2.2. Accounting for developmental trajectories

In the last forty years, a large body of empirical work has been collected regarding infant's language achievements during their first years of life. This work has only added more puzzlement.

First, given the multi-layered structure of language, one could expect a stage-like developmental tableau where acquisition would proceed as a discrete succession of learning phases organized logically or hierarchically (e.g., building linguistic structure from the low level to the high levels). This is not what is observed (see Fig. 1). For instance, infants start differentiating native from foreign consonants and vowels at 6 months, but continue to fine tune their phonetic categories well after the first year of life (e.g., Sundara, Polka, & Genesee, 2006). However, they start learning about the sequential structure of phonemes (phonotactics, see Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993) way *before* they are done acquiring the phoneme inventory (Werker & Tees, 1984). Even before that, they start acquiring the meaning of a small set of common words (e.g. Bergelson & Swingley, 2012). In other words, instead of a stage-like developmental tableau, the evidence shows that acquisition takes places at all levels more or less simultaneously, in a *gradual* and largely *overlapping* fashion.

Second, observational studies have revealed considerable *variations* in the *amount of language input* to infants across cultures (Shneidman & Goldin-Meadow, 2012) and across socio-economic strata (Hart & Risley, 1995), some of which can exceed an order of magnitude (Weisleder & Fernald, 2013, p. 2146; Cristia, Dupoux, Gurven, & Stieglitz, 2017; see also Supplementary Section S1). These variations do impact language achievement as measured by vocabulary size and syntactic complexity (Hoff, 2003; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges,

[1] The two puzzles are not independent as they are two facets of the same phenomenon. In practice, proposals for solving the bootstrapping problem may offer insights about the observed trajectories. Vice-versa, data on developmental trajectories may provide more manageable subgoals for the difficult task of solving the bootstrapping problem.

[2] Even in later acquisitions, the nature, universality and effectiveness of corrective feedback of children's outputs has been debated (see Brown, 1973; Chouinard & Clark, 2003; Clark & Lappin, 2011; Marcus, 1993; Pinker, 1989; Saxton, 1997).

[3] A sucessful system may not necessarily have the same architecture of components as described by theoretical linguists. It just needs to behave as humans do, i.e., pass the same behavioral tests. More on this in Section 4.3.