Cognition 164 (2017) 116–143

Contents lists available at ScienceDirect

Cognition

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

Bootstrapping language acquisition ☆

Omri Abend *,[1], Tom Kwiatkowski [2], Nathaniel J. Smith [3], Sharon Goldwater, Mark Steedman

Informatics, University of Edinburgh, United Kingdom

ARTICLE INFO

Article history:
Received 10 February 2016
Revised 1 November 2016
Accepted 22 February 2017
Available online 13 April 2017

Parts of this work were previously presented at EMNLP 2010, EACL 2012, and appeared in Kwiatkowski's doctoral dissertation

Keywords:
Language acquisition
Syntactic bootstrapping
Semantic bootstrapping
Computational modeling
Bayesian model
Cross-situational learning

ABSTRACT

The semantic bootstrapping hypothesis proposes that children acquire their native language through exposure to sentences of the language paired with structured representations of their meaning, whose component substructures can be associated with words and syntactic structures used to express these concepts. The child's task is then to learn a language-specific grammar and lexicon based on (probably contextually ambiguous, possibly somewhat noisy) pairs of sentences and their meaning representations (logical forms).

Starting from these assumptions, we develop a Bayesian probabilistic account of semantically bootstrapped first-language acquisition in the child, based on techniques from computational parsing and interpretation of unrestricted text. Our learner jointly models (a) word learning: the mapping between components of the given sentential meaning and lexical words (or phrases) of the language, and (b) syntax learning: the projection of lexical elements onto sentences by universal construction-free syntactic rules. Using an incremental learning algorithm, we apply the model to a dataset of real syntactically complex child-directed utterances and (pseudo) logical forms, the latter including contextually plausible but irrelevant distractors. Taking the Eve section of the CHILDES corpus as input, the model simulates several well-documented phenomena from the developmental literature. In particular, the model exhibits syntactic bootstrapping effects (in which previously learned constructions facilitate the learning of novel words), sudden jumps in learning without explicit parameter setting, acceleration of word-learning (the "vocabulary spurt"), an initial bias favoring the learning of nouns over verbs, and one-shot learning of words and their meanings. The learner thus demonstrates how statistical learning over structured representations can provide a unified account for these seemingly disparate phenomena.

1. Introduction

One of the fundamental challenges facing a child language learner is the problem of generalizing beyond the input. Using various social and other extralinguistic cues, a child may be able to work out the meaning of particular utterances they hear, like "you read the book" or "Eve will read *Lassie*", if these are encountered in the appropriate contexts. But merely memorizing and reproducing earlier utterances is not enough: children must also somehow use these experiences to learn to produce and interpret novel utterances, like "you read *Lassie*" and "show me the book". There are many proposals for how this might be achieved, but abstractly speaking it seems to require the ability to explicitly or implicitly (a) decompose the utterance's form into syntactic units, (b) decompose the utterance's meaning into semantic units, (c) learn lexical mappings between these syntactic and semantic units, and (d) learn the language-specific patterns that guide their recombination (so that e.g. "Eve will read *Lassie* to Fraser", "will Eve read Fraser *Lassie*?", and "will Fraser read Eve *Lassie*?" have different meanings, despite using the same or nearly the same words). A further challenge is that even in child-directed speech, many sentences are more complex than "you read *Lassie*"; the child's input consists of a mixture of high- and low-frequency words falling into a variety of syntactic categories and arranged into a variety of more or less complex syntactic constructions.

In this work, we present a Bayesian language-learning model focused on the acquisition of *compositional* syntax and semantics in an *incremental, naturalistic setting*. That is, our model receives training examples consisting of whole utterances paired with noisy

☆ The work was supported in part by ERC Advanced Fellowship 249520 GRAMPLUS, EU IST Cognitive Systems IP EC-FP7-270273 Xperience, ARC Discovery grants DP 110102506 and 160102156, and a James S McDonnell Foundation Scholar Award.

* Corresponding author at: School of Computer Science and Engineering, Rothberg Building A527, Edmond Safra Campus, Givat Ram, The Hebrew University, Jerusalem 91904, Israel.

E-mail address: oabend@cs.huji.ac.il (O. Abend).

[1] Now at the Departments of Computer Science & Cognitive Science, The Hebrew University of Jerusalem.

[2] Now at Google Research.

[3] Now at the Berkeley Institute of Data Science, University of California, Berkeley.

http://dx.doi.org/10.1016/j.cognition.2017.02.009
0010-0277/

representations of the whole utterance's meaning, and from these it learns probabilistic representations of the semantics and syntax of individual words, in such a way that it becomes able to recombine these words to understand novel utterances and express novel meanings. This requires that the model simultaneously learn how to parse syntactic constructions, assign meaning to specific words, and use syntactic regularities (for example, in verb argument structure) to guide interpretation of ambiguous input. Our training data consists of real, syntactically complex child-directed utterances drawn from a single child in the CHILDES corpus, and our training is incremental in the sense that the model is presented with each utterance exactly once, in the same order that the child actually encountered them.

The work described here represents an advance over previous models that focused on learning *either* word meanings *or* syntax given the other (see below for a review). By developing a joint learning model we are able to explore how these phenomena interact during learning. A handful of other joint learning models have been presented in the literature, but these have either worked from synthetic input with varying degrees of realism (Beekhuizen, 2015; Maurits, Perfors, & Navarro, 2009) or have not yet been evaluated on specific phenomena known from child language acquisition, as we do here (Chrupała, Kádár, & Alishahi, 2015; Jones, 2015). In particular, we show in a series of simulations that our model exhibits syntactic bootstrapping effects (in which previously learned constructions facilitate the learning of novel words), sudden jumps in learning without explicit parameter setting, acceleration of word-learning (the "vocabulary spurt"), an initial bias favoring the learning of nouns over verbs, and one-shot learning of words and their meanings. These results suggest that there is no need to postulate distinct learning mechanisms to explain these various phenomena; rather they can all be explained through a single mechanism of statistical learning over structured representations.

1.1. Theoretical underpinnings

Our model falls under the general umbrella of "Semantic Bootstrapping" theory, which assumes that the child can access a structural representation of the intended semantics or conceptual content of the utterance, and that such representations are sufficiently homomorphic to the syntax of the adult language for a mapping from sentences to meanings to be determined (Bowerman, 1973; Brown, 1973; Clark, 1973; Grimshaw, 1981; Pinker, 1979; Schlesinger, 1971; cf. Wexler & Culicover, 1980:78–84; Berwick, 1985:22–24). By "homomorphic", we simply mean that meaning representation and syntax stand in a "type-to-type" relation, according to which every syntactic type (such as the English intransitive verb) corresponds to a semantic type (such as the predicate), and every rule (such as English $\mathsf{S} \rightarrow \mathsf{NP}\quad \mathsf{VP}$) corresponds to a semantic operation (such as function application of the predicate to the subject).

Early accounts of semantic bootstrapping (e.g. Berwick, 1985; Wexler & Culicover, 1980) assumed perfect access to a single meaning representation in the form of an *Aspects*-style Deep Structure already aligned to the words of the language. Yet, as we shall see, semantic bootstrapping is sufficiently powerful that such strong assumptions are unnecessary.

Since, on the surface, languages differ in many ways—for example with respect to the order of heads and complements, and in whether such aspects of meaning as tense, causality, evidentiality, and information structure are explicitly marked—the meaning representations must be expressed in a universal prelinguistic conceptual representation, in whose terms all such distinctions are expressable. The mapping must further be learned by general principles that apply to all languages. These general principles are often referred to as "universal grammar", although the term is somewhat misleading in the present context since the model we develop is agnostic as to whether these principles are unique to language or apply more generally in cognition.

A number of specific instantiations of the semantic bootstrapping theory have been proposed over the years. For example, "parameter setting" accounts of language acquisition assume, following Chomsky (1981), that grammars for each natural language can be described by a finite number of finitely-valued parameters, such as head-position, pro-drop, or polysynthesis (Hyams, 1986 and much subsequent work). Language acquisition then takes a form that has been likened to a game of Twenty-Questions (Yang, 2006 Ch:7), whereby parameters can be set when the child encounters "triggers", or sentences that can only be analyzed under one setting of a parameter. For example, for Hyams (1986), the fact that English has lexical expletive subjects (e.g., *it* in *it rained*) is unequivocal evidence that the pro-drop parameter is negative, while for others the position of the verb in simple intransitive sentences in Welsh is evidence for head-initiality. Such triggers are usually discussed in purely syntactic terms. However, in both examples, the child needs to know which of the words is the verb, which requires a prior stage of semantic bootstrapping at the level of the lexicon (Hyams, 1986:132–133).

Unfortunately, parameter setting seems to raise as many questions as it answers. First, there are a number of uncertainties concerning the way the learner initially identifies the syntactic categories of the words, the specific inventory of parameters that are needed, and the aspects of the data that "trigger" their setting (Fodor, 1998; Gibson & Wexler, 1994; Niyogi & Berwick, 1996). Second, several combinatoric problems arise from simplistic search strategies in this parameter space (Fodor & Sakas, 2005). Here, we will demonstrate that step-like learning curves used to argue for parameter-setting approaches (Thornton & Tesan, 2007) can be explained by a statistical model without explicit linguistic parameters.

A further variant of the semantic bootstrapping theory to be discussed below postulates a second, later, stage of "syntactic bootstrapping" (Braine, 1992; Gleitman, 1990; Landau & Gleitman, 1985; Trueswell & Gleitman, 2007), during which the existence of early semantically bootstrapped syntax allows rapid or even "one-shot" learning of lexical items, including ones for which the situation of utterance offers little or no direct evidence. Early discussions of syntactic bootstrapping implied that it is a learning mechanism in its own right, distinct from semantic bootstrapping. However, we will demonstrate that these effects attributed to syntactic bootstrapping emerge naturally under the theory presented here. That is, our learner exhibits syntactic bootstrapping *effects* (using syntax to accelerate word learning) without the need for a distinct *mechanism*: the mechanism of semantic bootstrapping is sufficient to engender the effects.

Although varieties of semantic bootstrapping carry considerable currency, some researchers have pursued an alternative *distributional* approach (Redington, Chater, & Finch, 1998), which assumes that grammatical structure can be inferred from statistical properties of strings alone. Many proponents of this approach invoke Artificial Neural Network (ANN) computational models as an explanation for how this could be done—see Elman et al. (1996) for examples—while others in both cognitive science and computer science have proposed methods using structured probabilistic models (Cohn, Blunsom, & Goldwater, 2010; Klein & Manning, 2004; Perfors, Tenenbaum, & Regier, 2011). The distributional approach is appealing to some because it avoids the assumption that the child can access meanings expressed in a language of mind that is homomorphic to spoken language in the sense defined above, but inaccessible to adult introspection and whose detailed character is otherwise unknown.