A Computational Simulation of Children's Language Acquisition

Ben Ambridge
ESRC International Centre for Language and Communicative Development (LuCiD), University of Liverpool, UK

Abstract

Many modern NLP models are already close to simulating children's language acquisition; the main thing they currently lack is a "real world" representation of semantics that allows them to map from form to meaning and vice-versa. The aim of this "Crazy Idea" is to spark a discussion about how we might get there.

2012 ACM Subject Classification Theory of computation → Grammars and context-free languages

Keywords and phrases Child language acquisition, language development, deep learning, BERT, ELMo, GPT-3

Digital Object Identifier 10.4230/OASIcs.LDK.2021.4

Category Crazy New Idea

Funding This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 681296: CLASS). Ben Ambridge is a Professor in the International Centre for Language and Communicative Development (LuCiD) at The University of Liverpool. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

1 Crazy Idea

Modern NLP systems such as BERT, ELMo and GPT-3 have many potential applications in both industry and academia; but one that has barely been considered is simulating how children learn their native language. This question lies at the very heart of cognitive science – with at least five journals devoted solely to it – but has yet to be tackled with modern NLP approaches. Although modelling work is conducted in this domain, it typically uses small and simple models (e.g., three-layer connectionist networks) to tackle narrowly circumscribed problems (for example, children's acquisition of the English past-tense system; [10]).

But here's the thing: Models like BERT, ELMo and GPT-3 are, in many respects, nearly there. What the past 50 years of child language research have taught us is that learners store representations at every level from the concrete (e.g., the lexical string *cup+of+tea*) to the abstract (e.g., the *SUBJECT VERB OBJECT* transitive construction), and everything in between (see [1, 2] for reviews). That is, exemplars – utterances that children hear and store – are never discarded in favour of context-free symbolic rules. Rather these exemplars are re-represented at increasingly abstract levels, just as in BERT, ELMo, GPT-3 (and other deep-learning models in domains such as image-classification models; e.g., [14]).

Crucially, as I argued in [2], "this type of model is not just a metaphor ([5, 6, 9]). The brain really does contain multiple layers of units (i.e., neurons), each of which aggregates input signals using a nonlinear function and outputs signals to other units. While any particular artificial neural network model of language is only the clumsiest metaphor, the claim that language is represented as patterns of activation across 'dumb' neurons, each of which 'knows' nothing about nouns, verbs and all the rest of it is literally true, and quite beyond dispute".

3rd Conference on Language, Data and Knowledge (LDK 2021).
Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch; Article No. 4; pp. 4:1–4:3
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

So what's missing? Why aren't BERT and the like already viable candidates for models of children's language acquisition? The answer, of course, is that BERT lacks not only any kind of communicative goals, but any links to real-world meanings at all (e.g., [3, 11, 12]), with "meanings" represented solely as contextualized word embeddings. What we need, then, is a deep-learning model that learns like children; a model that – when "listening" – maps strings onto meanings and – when speaking – maps "meanings" onto strings.

Of course, this type of approach was tried in the earliest days of NLP, and swiftly abandoned as unworkable. And, indeed, if our goal is to translate from one natural language to another, to develop a predictive-text application, or to generate passages of text given a prompt (e.g., GPT-3), contextualized word embeddings will probably do a better job. But if our goal is to simulate children's language acquisition, we have to bite the bullet and develop "real-world" semantic representations (which, as Gary Marcus has often argued, are important for many practical applications of NLP too).

Indeed, simulating the first few years of language acquisition may be a useful way to take the first steps towards this much bigger problem. A typical two-year-old has a vocabulary of only a couple of hundred words; a typical three-year-old, a couple of thousand. The amount of input that children receive – around 10,000–20,000 words of speech per day – is also small by BERT standards, and most of it is relatively simple, concrete and highly repetitive ([4]). Thus, simulating the first few years of child language acquisition in its entirety is a realistic goal for an ambitious and well-funded research team, even if some hand coding of semantics is required (though a wrinkle here is the extent to which children's semantic representations are adultlike).

How should we go about this problem? This is where I hand over to you (and why I'm submitting this paper as a "Crazy Idea" for discussion). I'm a child language experimentalist, with only very limited experience of basic computational modelling. Could knowledge graphs represent the necessary semantic information? Would we need some additional hand-coding of at least the basic objects and actions in the child's world? And, if so, can we adopt a "view from nowhere", or do we need to take account of the fact that human cognition is embodied in our sensorimotor experience ([8]), perhaps by including something like sensorimotor norms (e.g., [7])?

Could neural-symbolic approaches (e.g., [13]) connect knowledge graphs with neural networks? And what other semantic representations are used in modern NLP? Or can we use some kind of vector representation after all, perhaps using principal component analysis to distil them into elements of meaning that can be used to encode semantic "messages". Or can we somehow represent meaning by leveraging techniques used in machine translation and using crosslinguistic vectors (e.g., the "meaning" of *cat* is the entity that stands in the same relationship to *dog* as does French *chat* to *chien*)? You tell me!

References

1 Ben Ambridge. Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, 40(5-6):509–559, 2020a.

2 Ben Ambridge. Abstractions made of exemplars or "you're all right, and i've changed my mind": Response to commentators. *First Language*, 40(5-6):640–659, 2020b.

3 Emily M Bender and Alexander Koller. Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, 2020.

4 Thea Cameron-Faulkner, Elena Lieven, and Michael Tomasello. A construction based analysis of child directed speech. *Cognitive Science*, 27(6):843–873, 2003.

5 Daniel C Dennett. *From bacteria to Bach and back: The evolution of minds*. WW Norton & Company, 2017.

6 Uri Hasson, Samuel A Nastase, and Ariel Goldstein. Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron*, 105(3):416–434, 2020.

7 Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, pages 1–21, 2019.

8 Jean M Mandler. How to build a baby: Ii. conceptual primitives. *Psychological Review*, 99(4):587, 1992.

9 Andrea E Martin. A compositional neural architecture for language. *Journal of Cognitive Neuroscience*, 32(8):1407–1427, 2020.

10 James L McClelland, David E Rumelhart, PDP Research Group, et al. *Parallel distributed processing*, volume 2. MIT press Cambridge, MA, 1986.

11 William Merrill, Yoav Goldberg, Roy Schwartz, and Noah A. Smith. Provable limitations of acquiring meaning from ungrounded form: What will future language models understand? *CoRR*, abs/2104.10809, 2021. `arXiv:2104.10809`.

12 Mitja Nikolaus and Abdellah Fourtassi. Evaluating the acquisition of semantic knowledge from cross-situational learning in artificial neural networks, May 2021. `doi:10.31234/osf.io/mbesf`.

13 Kexin Yi, Jiajun Wu, Chuang Gan, Antonio Torralba, Pushmeet Kohli, and Joshua B Tenenbaum. Neural-symbolic vqa: Disentangling reasoning from vision and language understanding. *arXiv preprint*, 2018. `arXiv:1810.02338`.

14 Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization, 2017. `arXiv:1611.03530`.