Cognition 151 (2016) 18–27

Contents lists available at ScienceDirect

Cognition

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

Word learning under infinite uncertainty

Richard A. Blythe [a,*], Andrew D.M. Smith [b], Kenny Smith [c]

[a] SUPA, School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3JZ, UK
[b] Literature and Languages, School of Arts and Humanities, University of Stirling, Stirling FK9 4LA, UK
[c] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9AD, UK

ARTICLE INFO

Article history:
Received 5 January 2015
Revised 16 February 2016
Accepted 21 February 2016
Available online 27 February 2016

Keywords:
Word learning
Cross-situational learning
Quine's Problem

ABSTRACT

Language learners must learn the meanings of many thousands of words, despite those words occurring in complex environments in which infinitely many meanings might be inferred by the learner as a word's true meaning. This problem of *infinite referential uncertainty* is often attributed to Willard Van Orman Quine. We provide a mathematical formalisation of an ideal cross-situational learner attempting to learn under infinite referential uncertainty, and identify conditions under which word learning is possible. As Quine's intuitions suggest, learning under infinite uncertainty is in fact possible, provided that learners have some means of ranking candidate word meanings in terms of their plausibility; furthermore, our analysis shows that this ranking could in fact be exceedingly weak, implying that constraints which allow learners to infer the plausibility of candidate word meanings could themselves be weak. This approach lifts the burden of explanation from 'smart' word learning constraints in learners, and suggests a programme of research into weak, unreliable, probabilistic constraints on the inference of word meaning in real word learners.

1. Word learning and indeterminacy of meaning

Children are prolific word learners, learning around 60,000 words by age 18 (Bloom, 2000). Their prodigious word-learning abilities are even more remarkable when we consider some of the challenges facing the word learner, including the need to segment words from connected speech (Saffran, Aslin, & Newport, 1996), to generalise word forms across speakers (Henderson & Graham, 2005), and to identify the syntactic properties of those words (Mintz, 2002). In this paper we focus on another aspect of the word learning problem: inferring word meaning. Children will typically encounter words in a complex environment. How do they know what these words mean? Every time a word is used, there may be many meanings which a learner could infer as the word's true meaning: the learner will face *referential uncertainty*. As discussed below, a widespread observation in the literature is that there are potentially *infinitely* many candidate meanings which would be consistent with any given situation of usage. This idea, which we will refer to as *infinite referential uncertainty*, is commonly attributed to Quine (1960), although on our reading (discussed below) we think Quine's central point was rather different.

Regardless of its provenance, the infinite referential uncertainty hypothesis has been crucial in the development of two approaches to word learning, which differ in emphasis but are entirely compatible in content. One position emphasises the importance of heuristics which guide word learning, serving to reduce referential uncertainty and allow the learner to make accurate inferences about word meaning. These heuristics might include: exploiting the attentional focus of a speaker (Tomasello & Farrar, 1986); the assumption that words refer to whole objects (Macnamara, 1972); using knowledge of the meaning of other words to constrain hypotheses about the meaning of a new word, for example by assuming that words have mutually exclusive meanings (Markman & Wachtel, 1988); using argument structure and syntactic context to constrain the meaning of new words (Gillette, Gleitman, Gleitman, & Lederer, 1999). Heuristic-driven accounts emphasise how such constraints enable learners to eliminate uncertainty about word meaning and form good hypotheses about word meaning on even a single exposure to a word.

In the strongest accounts, these heuristics are hypothesised to eliminate all uncertainty. However, the possibility that word learners may be confronted with *some* residual referential uncertainty even after these heuristics have done their work has driven a recent burst of interest in a second approach to word learning, emphasising integration of information across multiple exposures

* Corresponding author.
E-mail address: r.a.blythe@ed.ac.uk (R.A. Blythe).

http://dx.doi.org/10.1016/j.cognition.2016.02.017
0010-0277/

as a means for learning in the face of referential uncertainty. *Cross-situational learning* comes in various flavours, from the classic formulation provided by e.g. Siskind (1996) to associationist treatments (Yu & Smith, 2007) to more minimal accounts (Medina, Snedeker, Trueswell, & Gleitman, 2011; Smith, Smith, & Blythe, 2011). For instance, in its most powerful instantiation (e.g. Siskind, 1996), cross-situational learning involves tracking the set of meanings which has been consistently inferred on every exposure to some target word: the word's true meaning should be a member of this set, which can be winnowed down across a series of exposures until it includes only the true meaning. Cross-situational learning accounts typically assume the presence of heuristics which serve to reduce referential uncertainty to manageable levels: rather than replacing heuristics, the contribution of this research is to explore the extent to which word learning is possible even given some residual (i.e. non-zero, but typically small) referential uncertainty.

Focussing on the interaction between heuristic and cross-situational approaches, in previous work (Blythe, Smith, & Smith, 2010) we applied mathematical techniques to quantify what residual level of referential uncertainty a cross-situational learner can tolerate and still learn a large lexicon in a reasonable timeframe. Our previous work focussed on calculating learning times for lexicons given finite meaning spaces and finite levels of referential uncertainty. In this paper we apply similar techniques to tackle the problem of cross-situational learning for infinite meaning spaces under infinite referential uncertainty. In doing so, we seek to address what is often (perhaps rather loosely) called "Quine's Problem" or "the gavagai problem", the notion that word learning under infinite referential uncertainty is impossible. We show that word learning under such conditions is in principle possible, provided that learners have heuristics which at least rank the plausibility of each candidate meaning at every exposure. Thus, as in fact envisaged by Quine in his exposition of the indeterminacy of translation, word learning is possible if learners know, of the infinitely many possible meanings a word could have on any given situation of usage, how plausible each of those meanings are, and that some are more plausible than others. Within this very general set of conditions, given enough time, cross-situational learning can be used to eliminate uncertainty. Furthermore, cross-situational learning will in principle be possible even if the learner's heuristics only impose very weak constraints on the ranking in terms of plausibility. This work therefore suggests similar conclusions to our previous work exploring finite referential uncertainty: word learning heuristics can in principle be far weaker than previously suggested and still allow word learning – in fact, those heuristics can be so weak as to admit *infinitely many* possible meanings on any given exposure to a word, which renders single-exposure word learning impossible. Importantly, we therefore directly overturn the commonly-held assumption that word learning is impossible in the face of infinite referential uncertainty. Furthermore, the fact that word-learning heuristics which provide only weak constraints on possible word meanings can nonetheless allow word learning has potential implications for our understanding of the heuristics and cognitive biases underpinning word learning, and therefore on the empirical research attempting to uncover those biases. Firstly, this moves the explanatory burden from 'smart' inference by learners to 'dumb' crunching of cross-situational statistics, therefore requiring us to assume less of word learners in terms of their ability to accurately infer word meaning. Second, word learning heuristics do not need to allow learners to make good guesses on a single exposure to a word, which is a standard diagnostic in experimental research: weaker, unreliable, probabilistic heuristics can also play a key role, and therefore merit investigation.

2. 'Quine's Problem': learning under infinite uncertainty

Words are used in complex environments, and each word could label any part of that complex environment. Worse, words can label objects and events which are not perceivable to speaker or hearer (e.g. events which are spatially or temporally distant from the time of speaking). And this is only considering the obvious possibilities – words might have 'strange' meanings (e.g. featuring disjunctions of the meanings of 'normal' words, meaning for instance "a spark plug *or* an elephant", "happiness *or* the number 17", etc.). This idea, commonly attributed to Quine's work on radical translation (of which more below), appeals to the notion that on any situation there will be *infinitely many* possible meanings that a novel word could have:

> "Even if we restrict ourselves to middle-sized objects . . .we are stuck with Quine's problem, which is that children who hear a word and know that it refers to a rabbit are still faced with an indefinite number of possible meanings for this word"
>
> (Bloom, 2000, p. 56)

> "Quine (1960) points out that there are an infinite number of true facts about the world that a learner might need to entertain as potential meanings of each utterance."
>
> (Siskind, 1996, p. 45)

> "Worse, or so philosophers tell us, learners might conjure up absurd and endlessly differing representations for those entities we adults call 'the cats.'"
>
> (Gillette et al., 1999, p. 136)

> "Famously articulated by Quine (1960), in any naming situation there are infinite interpretations for an unknown word. Thus, children face a daunting task of ambiguity resolution that they must solve thousands of times."
>
> (McMurray, Horst, & Samuelson, 2012, p. 831)

> "Word learning is often described as a difficult task because the world offers infants a seemingly infinite number of word-to-world mappings in just one moment in time (Quine, 1960)."
>
> (Vlach & Johnson, 2013, p. 375)

> "Determining the meaning of a newly encountered word should be extremely hard, due to the (in principle, unlimited) referential uncertainty inherent in the task (Quine, 1960)."
>
> (Smith et al., 2011, p. 480)

Such claims about infinite referential uncertainty are widespread in the literature, and have played an important role in the development of the theoretical motivation for research on the heuristics children use to eliminate uncertainty during word learning: seminal papers on the Mutual Exclusivity constraint (Markman & Wachtel, 1988), the shape bias (Landau, Smith, & Jones, 1988), joint attention (Baldwin, 1991), or lexical constraints in general (Golinkoff, Mervis, & Hirsh-Pasek, 1994) make explicit reference to the problem of there being "infinitely many" or "limitless" meanings a word could have. The consensus is that word learning is impossible given this infinite uncertainty, and that heuristics are required to eliminate some of these candidate meanings. In this paper we explore the validity of this widely-held and entirely reasonable intuition, and in particular show that it does not hold in a wide range of well-defined circumstances. However, before doing so it is worth briefly considering whether Quine's Problem was actually posed by Quine.

3. Quine on word learning

Quine (1960) introduces the problem not in terms of word learning, but in terms of "radical translation", his examination of