Cognition 148 (2016) 117–135

Contents lists available at ScienceDirect

Cognition

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

Perceiving referential intent: Dynamics of reference in natural parent–child interactions

John C. Trueswell [a,*], Yi Lin [a], Benjamin Armstrong III [a], Erica A. Cartmill [b], Susan Goldin-Meadow [c], Lila R. Gleitman [a,*]

[a] University of Pennsylvania, United States
[b] University of California, Los Angeles, United States
[c] University of Chicago, United States

ARTICLE INFO

Article history:
Received 3 August 2014
Revised 5 August 2015
Accepted 6 November 2015
Available online 8 January 2016

Keywords:
Psycholinguistics
Language development
Word learning
Reference

ABSTRACT

Two studies are presented which examined the temporal dynamics of the social-attentive behaviors that co-occur with referent identification during natural parent–child interactions in the home. Study 1 focused on 6.2 h of videos of 56 parents interacting during everyday activities with their 14–18 month-olds, during which parents uttered common nouns as parts of spontaneously occurring utterances. Trained coders recorded, on a second-by-second basis, parent and child attentional behaviors relevant to reference in the period (40 s) immediately surrounding parental naming. The referential transparency of each interaction was independently assessed by having naïve adult participants guess what word the parent had uttered in these video segments, but with the audio turned off, forcing them to use only non-linguistic evidence available in the ongoing stream of events. We found a great deal of ambiguity in the input along with a few potent moments of word-referent transparency; these transparent moments have a particular temporal signature with respect to parent and child attentive behavior: it was the object's appearance and/or the fact that it captured parent/child attention at the moment the word was uttered, not the presence of the object throughout the video, that predicted observers' accuracy. Study 2 experimentally investigated the precision of the timing relation, and whether it has an effect on observer accuracy, by disrupting the timing between when the word was uttered and the behaviors present in the videos as they were originally recorded. Disrupting timing by only ±1 to 2 s reduced participant confidence and significantly decreased their accuracy in word identification. The results enhance an expanding literature on how dyadic attentional factors can influence early vocabulary growth. By hypothesis, this kind of time-sensitive data-selection process operates as a filter on input, removing many extraneous and ill-supported word-meaning hypotheses from consideration during children's early vocabulary learning.

1. Introduction

Our intuitions tell us that infants likely learn the meanings of their very first words during moments when word and object happen to co-occur, e.g., when they hear the word "doggie" in the presence of a dog. And indeed, ample observational and experimental evidence supports this idea (e.g., Baldwin, 1991, 1993; Baldwin & Tomasello, 1998; Bloom, 2002; Brown, 1973; Hollich et al., 2000; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Smith, Colunga, & Yoshida, 2010). Yet this very same evidence tells us that mutual co-presence of word and thing is probabilistic and conditional, rather than necessary and sufficient, for an infant to identify a referent and learn a word's meaning. The referential context depicted in Fig. 1 is an example of one glaring problem that must be solved to make good on any word-to-referent scheme for lexical learning: there seem to be far too many hypotheses made available by the observed scene, and probably too many for a realistic full cross-situational comparison process to parse out across multiple observations (e.g., Medina, Snedeker,

* Corresponding authors at: Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Solomon Lab Bldg., Philadelphia, PA 19104-6241, United States.
E-mail addresses: trueswel@psych.upenn.edu (J.C. Trueswell), gleitman@psych.upenn.edu (L.R. Gleitman).

http://dx.doi.org/10.1016/j.cognition.2015.11.002
0010-0277/

Fig. 1. Example of a referential context. Photograph courtesy of Tamara Nicol Medina (Medina et al., 2011).

Trueswell, & Gleitman, 2011). If the learner's task in word learning actually required completely open-minded referent selection from this set of presented alternatives, surely language would be very difficult if not impossible to learn. However, paradoxically enough, Fig. 1 points to approaches for solving the very question it poses. After all, the infant in this picture is looking at the shoe beneath his walker. If parents tend to talk about what their children are attending to, the reference problem seems more tractable (Bruner, 1974/1975). Indeed, even outside observers of this snapshot of parent–child interaction guess quite often – and correctly – that the mother was uttering "shoe" at the moment the picture was taken. From this perspective, it seems hardly to matter how many objects, qualia, etc., are in reach of the visual scan – be it 10 or 1000 alternatives – what matters most for communication is the immediate "common ground", the focus of joint attention for the interlocutors (e.g., Grice, 1975, 1989; Lyons, 1999; see also Brown-Schmidt & Tanenhaus, 2008; Yoshida & Smith, 2008).

In the studies presented here, we aim to investigate the properties and behaviors present in parent-infant interactions that are informative for identifying the intended referent of child-directed speech. To do this, we examine parent-infant visual attention, gesture, and object manipulation as words are uttered under typical conversational circumstances in the home. Importantly, and as we describe further below (see Section 2.1), we take advantage of a particular property of our corpus: it includes an independent estimate of the referential transparency of each exchange. In particular, adult observers watched muted versions of these videos and guessed what words the parent was uttering, in a procedure known as the Human Simulation Paradigm (HSP, Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2003). This procedure provides us with an estimate of referential transparency as inferred from the extralinguistic cues present in each interaction – words that are easily guessed are assumed to have been uttered in more transparent circumstances than words that are more difficult to guess.

Our focus is on two interrelated questions. First, just how referentially ambiguous is the infant's (sampled) learning environment, operationalized as the HSP observers' ability to reconstruct the intended referent of words from whatever extralinguistic cues are present. Our second focus is on the role of the *temporal dynamics* of these interactions, i.e., how these extralinguistic cues intercalate in time with the word utterance itself. That is, following a venerable theme from David Hume (1748), we ask how precise temporally contiguous cues have to be for an observer to conclude that there is a cause-effect relation between input words and the nonlinguistic behavior of the speaker. Is the timing relation systematic and tight enough to support a learner's choice of referent among all those that are in principle available when scanning the passing scene?[1]

We are by no means the first to address these questions. The topic of joint attention and its explanatory role in language acquisition was introduced into the current experimental literature in a seminal paper by Bruner (1974/75) who suggested that joint attention and joint reference likely provided an important early mechanism for linguistic and social learning; parents might do much of the work of referent identification by talking about what children are attending to. These comments led to substantial observational research examining interactional cues to learning (Moore & Dunham, 1995, and papers therein), which revealed the social-attentive behaviors that arise in spontaneous parent–child interactions during object play, as recorded either in the lab or home (e.g., Harris, Jones, Brookes, & Grant, 1986; Tomasello & Farrar, 1986; Tomasello, Mannle, & Kruger, 1986; Tomasello & Todd, 1983). These now classic studies established that not all parental word utterances are created equal when it comes to their ability to predict child vocabulary growth and, by implication, to facilitate accurate referent identification. In particular, parents who engaged more in follow-in labeling – labeling what the child was currently attending to – had children whose vocabulary growth outpaced that of children who were exposed to proportionally more discrepant labeling situations, with the latter being negatively correlated with vocabulary growth (e.g., Tomasello & Farrar, 1986). This work suggests that, at least during controlled object play, referent identification is

[1] From the way we have just set our problem space, it should be clear that our primary interest in the present paper is the very beginnings of vocabulary learning, which relies much more on evidence from the co-present referent world. It is now well established that children make inferences about word meaning based not only on reference but on, e.g., collateral distributional and syntactic evidence (e.g., Chomsky, 1969; Landau & Gleitman, 1985; Lidz, Waxman, & Freedman, 2003; Naigles, 1990, inter alia). Yet, these linguistic resources cannot themselves be mobilized until a "seed" vocabulary, mainly of whole-object nominals are acquired by perceptual observation, and used to build distributional libraries and the syntactic structures of the exposure language (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005). Reference finding via extralinguistic cue structure is only one evidentiary source for lexical learning but it is necessarily the earliest step, on which later accomplishments hinge.