ELSEVIER

The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition ☆

Padraic Monaghan [a,*], Morten H. Christiansen [b], Nick Chater [c]

[a] Department of Psychology, University of York, York, YO10 5DD, UK
[b] Department of Psychology, Cornell University, Ithaca, NY 14853, USA
[c] Department of Psychology, University College London, Gower Street, London, WC1E 6BT, UK

Abstract

Several phonological and prosodic properties of words have been shown to relate to differences between grammatical categories. Distributional information about grammatical categories is also a rich source in the child's language environment. In this paper we hypothesise that such cues operate in tandem for developing the child's knowledge about grammatical categories. We term this the Phonological-Distributional Coherence Hypothesis (PDCH). We tested the PDCH by analysing phonological and distributional information in distinguishing open from closed class words and nouns from verbs in four languages: English, Dutch, French, and Japanese. We found an interaction between phonological and distributional cues for all four languages indicating that when distributional cues were less reliable, phonological cues were stronger. This provides converging evidence that language is structured such that language learning benefits from the integration of information about category from contextual and sound-based sources, and that the child's language environment is less impoverished than we might suspect.
© 2007 Elsevier Inc. All rights reserved.

1. Introduction

Learning grammatical categories is essential in order for the child to develop an understanding of the relationships between sounds in a spoken sentence and objects and actions in the world around them (Gentner, 1982). Knowledge of the patterns determining which words can relate to objects, which to actions, and which modify the relationships between these objects and actions is an imperative in language development (Pinker, 1984).

One view of this acquisition process is that the child has innate constraints that facilitate this development. Some theorists argue that these constraints encode a complete grammar of human natural language, aside from a finite set of parametric variations that define the structural differences between languages (e.g., Baker, 2001; Chomsky, 1965, 1981; Crain & Lillo-Martin, 1999). From this perspective, the entire grammatical machinery of natural language is innate—and hence the set of possible syntactic categories, including nouns, verbs, adjectives, and so on, must similarly be innate. The child's task, under this view, is to learn which words belong to which syntactic categories.

Alternatively, Pinker's (1984) semantic bootstrapping hypothesis predicts rather that certain *semantic* referents are innately specified, and reflected in the surface properties of the language in terms of distributional co-occurrence information. Thus, for the noun/verb distinction, the child has innately specified information in terms of nouns referring to objects, and verbs referring to actions. These semantic referents then constrain the child's search for relevant correlations in the language to which she is exposed, and also, according to Pinker, provide an explanation for why such correlations between surface distributional properties and semantic features occur in natural languages (e.g., that nouns and verbs occur in different distributional contexts). Pinker (1984, p.43) states "it [semantic bootstrapping] claims that children always give priority to distributionally based analyses, and is intended to explain how the child knows which distributional contexts are the relevant ones to examine." Whether the innately specified language structure is syntactic or semantic, the child also faces a further task: learning which grammatical categories are realized in the language, given that not all possible categories occur in all languages (e.g., Croft, 2003; Dixon, 1977). According to some recent, and influential, linguistic analyses, the child's task, under the nativist position, may be more complex than previously assumed due to the extraordinary variety of fine-grained syntactic categories in natural language (e.g., Culicover, 1999).

The view that some knowledge about the language, or grammatical categories of the language, is innately specified is typically based, at least in part, on the assumption that there is insufficient evidence in the child's language environment to enable these properties to be learned from the language itself. That is, nativist viewpoints concerning the origins of syntactic categories typically rely, to some degree, on arguments from the "poverty of the stimulus" (e.g., Chomsky, 1980; though see Pullum & Scholz, 2002). Under the semantic bootstrapping account, for instance, it is claimed that learning the correlations between grammatical categories and distributional information of their usage ought to be impossible as the search for correlations is too unconstrained. Yet, a study by Gerken, Wilson, and Lewis (2005) demonstrated that such learning *is* possible in children younger than two