Word frequency as a cue for identifying function words in infancy

Jean-Rémy Hochmann [a],[*], Ansgar D. Endress [b], Jacques Mehler [a]

[a] SISSA, via Bonomea 265, 34136, Trieste TS, Italy
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar St, 46-4127, Cambridge, MA 02139, United States

ABSTRACT

While content words (e.g., 'dog') tend to carry meaning, function words (e.g., 'the') mainly serve syntactic purposes. Here, we ask whether 17-month old infants can use one language–universal cue to identify function word candidates: their high frequency of occurrence. In Experiment 1, infants listened to a series of short, naturally recorded sentences in a foreign language (i.e., in French). In these sentences, two determiners appeared much more frequently than any content word. Following this, infants were presented with a visual object, and simultaneously with a word pair composed of a determiner and a noun. Results showed that infants associated the object more strongly with the infrequent noun than with the frequent determiner. That is, when presented with both the old object and a novel object, infants were more likely to orient towards the old object when hearing a label with a new determiner and the old noun compared to a label with a new noun and the old determiner. In Experiment 2, infants were tested using the same procedure as in Experiment 1, but without the initial exposure to French sentences. Under these conditions, infants did not preferentially associate the object with nouns, suggesting that the preferential association between nouns and objects does not result from specific acoustic or phonological properties. In line with various biases and heuristics involved in acquiring content words, we provide the first direct evidence that infants can use distributional cues, especially the high frequency of occurrence, to identify potential function words.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Language acquisition involves learning both syntax and the lexicon. While these components are often studied separately, they might be intimately linked during language acquisition, and might cross-fertilize each other. Function words (such as determiners and prepositions) illustrate this point particularly well. They clearly are words that have to be acquired. However, in contrast to content words (which must be linked to some semantic referent), function words mainly serve syntactic rather than semantic purposes. On the one hand, function words might thus impair word learning – because they are words that children might try to learn, and yet they have no clear meaning they could be mapped onto. On the other hand, function words might facilitate word learning – by providing syntactic cues that might then be used for learning other (content) words. To use the syntactic cues associated with function words, however, infants need to *identify* them in the first place.

While different authors have uncovered different cues that tend to distinguish content and function words (Cutler, 1993; Shi, Morgan, & Allopenna, 1998), such proposals meet with two problems. First, to be useful for language acquisition, the cues must be available in any language a child might end up learning, and cannot be specific to a particular language (e.g., English). Second, early in life, infants need to be able to *use* such cues to identify function word candidates. Here, we start assessing these issues, asking whether infants can attribute different properties to

---

* Corresponding author.
  E-mail address: hochmann@sissa.it (J.-R. Hochmann).

potential content words and function words based on a language-independent distributional property of function words, namely their high frequency of occurrence.

2. Words, syntax, the chicken and the egg

Children acquire both the syntax and the lexicon of their native language. However, different theories disagree on the relation between the development of syntax and that of the lexicon. Specifically, proponents of semantic bootstrapping (Pinker, 1984) and usage-based theories of language acquisition (Dabrowska, 2001; Tomasello, 2003) hold that vocabulary acquisition facilitates syntax learning, while proponents of syntactic bootstrapping accounts (Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1994; Gleitman & Gleitman, 1992; Landau & Gleitman, 1985) propose that syntax boosts vocabulary acquisition. We will now briefly review both kinds of theories.

Semantic bootstrapping theories describe how infants can bootstrap the initial steps of syntax acquisition based on their knowledge of (a limited number of) words. For example, semantic categories such as objects and actions might initially be used to discover how syntactic categories such as nouns and verbs are implemented in the language. Specifically, infants might first acquire a few words related to the objects and actions they observe. Then, they might use these words to learn the corresponding syntactic categories. For example, object names might be mapped onto nouns, and words describing actions onto verbs. Based on such a mapping, infants might discover crucial aspects of the syntactic organization of their native language. For instance, knowing the verb 'eat' and the noun 'cookie' might be sufficient to decide whether the object comes after the verb (e.g., "eat cookies", corresponding to the canonical English word order), or whether the object precedes the verb (e.g., "cookies eat", corresponding to the canonical Japanese word order; Pinker, 1994, p. 112). On this view, infants can start acquiring syntax only after having learned a minimal set of words, because knowledge of these words is crucially required to bootstrap grammar acquisition.

Semantic information might help grammar acquisition in yet another way. According to usage-based theories of syntax acquisition (see e.g., Dabrowska, 2001; Tomasello, 2003), infants and children first learn specific word sequences, with very limited knowledge of their underlying structure. That is, they might remember words only in specific contexts, and assign meaning to words only within this context. Crucially, however, as they do not analyze sentences in terms of their underlying structure, they should be unable to use words in contexts that differ from those they have heard. For example, if they have heard the word "broke" only in the sentence "The window broke", they should be unable to use the word in new contexts such as "He broke it" or "The windows got broken" (e.g., Savage, Lieven, Theakston, & Tomasello, 2003; Tomasello, 2003; but see Lidz, Gleitman, & Gleitman, 2003; Thothathiri & Snedeker, 2008). As they get older, children should gradually discover that the sentences they have heard have in fact an underlying structure, eventually leading to the kind of abstract syntactic knowledge observed in mature, adult speakers. According to this theory, children thus need to acquire a substantial vocabulary before learning any syntactic regularity.

In contrast to such views, syntactic bootstrapping models hold that syntactic knowledge facilitates vocabulary acquisition (Gillette et al., 1999; Gleitman, 1994; Gleitman & Gleitman, 1992; Landau & Gleitman, 1985). For example, upon hearing a sentence like "the duck and the bunny are gorping", listeners as young as 2-year-olds are likely to conclude that "to gorp" must have an intransitive meaning, since it has no object. Upon hearing the sentence "the duck is gorping the bunny", in contrast, they tend to conclude that "to gorp" is transitive, since it now has a direct object (Naigles & Kako, 1993). Thus, a rather rudimentary syntactic analysis (such as counting the number of noun phrases and analyzing their positions) can constrain the interpretation of novel verbs.

Of course, semantic and syntactic bootstrapping accounts are not mutually exclusive, and infants might well use both routes in complementary ways. Both syntax and the lexicon might initially develop in parallel and cross-fertilize each other. This possibility is particularly important for the issue studied here, relating to how function words are acquired and used during language acquisition. From a syntactic bootstrapping perspective, the syntactic information carried by function words would be clearly helpful for learning new (content) words, as function words indicate syntactic roles and syntactic categories. For example, in a language like English, a word following a determiner is likely to be a noun, while a word following an auxiliary is likely to be a verb. Therefore, when hearing a novel word that is accompanied by a function word, infants might interpret it as referring to a novel object if the function word marks it as a noun (Brown, 1957), as referring to a novel action if the function word marks it as a verb (Bernal, Lidz, Millote, & Christophe, 2007; Brown, 1957), and as referring to a property when the function word marks it as an adjective (Waxman & Booth, 2001). This capacity seems to be present early in life, as infants as young as 14-month-old start using the syntactic information provided by function words to interpret new content words (Waxman & Booth, 2003).

While function words might facilitate the acquisition of content words by providing syntactic cues, they are more problematic from a vocabulary acquisition perspective, as they have no clear referents. As a result, unless infants can identify function words *as* function words, these words should impair vocabulary acquisition – because infants might consider them as meaningless "noise." In order to take advantage of the syntactic information provided by function words, infants thus need to identify them early on. In the next section, we will discuss a number of cues that might allow them to solve this problem.

3. Cues to identify function words

To identify function words, and to distinguish them from content words, infants might rely on two types of surface cues: phonological properties (Cutler, 1993; Shi, Werker, & Morgan, 1999; Shi et al., 1998), and distributional