



Morphosyntactic cues to noun categorization in English child-directed speech



Sara Feijoo^{a,*,1}, Carme Muñoz^{a,1}, Elisabet Serrat^{b,2}

^a English Department, Faculty of Philology, University of Barcelona, Barcelona, Spain

^b Department of Psychology, Faculty of Education and Psychology, University of Girona, Girona, Spain

ARTICLE INFO

Article history:

Available online 5 August 2015

Keywords:

Distributional cues
Child-directed speech
Word categorization
Statistical learning

ABSTRACT

Mainstream research in Linguistics claims that grammatical regularities are scarcely represented in the linguistic input to which children are exposed. However, recent empirical research shows that child-directed speech contains a series of reliable cues that might assist young language learners in language development. The present study aims at testing whether English child-directed speech contains morphosyntactic regularities which might be robust enough for infants to group nouns in their grammatical category. The results from the study show that, in fact, the kind of input available to English-learning infants contains reliable and consistent distributional cues to account for most of the nouns to which children are exposed.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Knowledge of the grammatical category membership of words is an important and essential part in language development, since it is a prerequisite to know how to use words in the language and produce grammatically correct sentences. Without an accurate categorization of words, language acquisition cannot proceed. But how do children learn the basic grammatical categories of their language?

Several studies have already highlighted the usefulness of distributional information for the accomplishment of several tasks during language development (cf. Elman et al., 1996; Plunkett, 1995). For example, distributional information has shown to be a very powerful cue that assists young language learners when they face the task of segmenting the continuous speech stream into word-like units (Brent and Cartwright, 1996; Johnson and Jusczyk, 2001; Saffran et al., 1996a, 1996b). In word segmentation tasks, distributional cues take the form of complex statistical information whereby sounds found within words will be strongly correlated (since they are part of the same word), while sounds found across word boundaries will be weakly correlated (since they occur one after the other only by chance). Saffran et al. (1996a) suggest that if children can learn words by recording frequent sound sequences, they might learn grammar in the same way.

Thus, when turning to the task of categorizing word-like units into their grammatical category, the same type of information might still be useful. In this case, the context of a word with respect to other words in the same sentence might provide indications about the category of that word in English. For example, English nouns are typically preceded by

* Corresponding author. English Department, Faculty of Philology, University of Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain.
E-mail addresses: sfeijoo@ub.edu (S. Feijoo), munoz@ub.edu (C. Muñoz), elisabet.serrat@udg.edu (E. Serrat).

¹ Facultat de Filologia, Dept. Filologia Anglesa i Alemanya, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain.

² Facultat d'Educació i Psicologia, Dept. Psicologia, Pl. Sant Domènec, 9, 17071 Girona, Spain.

determiners and followed by nominal morphology, while verbs are typically preceded by auxiliaries or strong subject pronouns and followed by verbal morphology. In this way, while in word segmentation the syllable *ba* is a high predictor of the syllable *by* (as in the word *baby*) and other syllables are not, in word categorization a determiner like *the* is equally a high predictor of a nominal element (as in the phrase *the baby*) and other words like *have* are not. In what follows, we shall examine the possibilities that distributional cues offer for word categorization.

2. Distributional cues and word categorization

As Mintz (2003) proposed, distributional information of the kind that can be found in the co-occurrence of patterns of words in sentences could provide a great deal of information relevant to the grammatical categories to which words belong. Studies on computer simulations have provided evidence for the usefulness of distributional and positional information for an initial categorization of words in the absence of semantic or referential information (Brent, 1996; Cartwright and Brent, 1997; Redington et al., 1998). Such distributional information appears to be available not only to adult speakers but also to young language learners (Gerken et al., 2005; Hsu et al., 2014; Mintz, 2002; Mintz et al. 2002; Mintz et al., 2014; Monaghan et al., 2005; Monaghan et al., 2007; Saffran et al., 1996b).

What type of distributional information is especially useful, and what kinds of distributional cues infants and young children are sensitive to and use in categorizing words? Authors distinguish between two different contexts: bigrams (Mintz et al., 2002; Redington et al., 1998) and frames (Mintz, 2003; Monaghan et al., 2007).

Bigrams are defined as pairs of elements where the categorizing element would either precede or follow the word to be categorized. In the case of English, for example, a word such as *come* could be successfully categorized as a verb on the basis of the bigram *they come* (where the strong subject pronoun would be the predictor of a coming verbal element) or on the basis of the bigram *coming* (where the verbal morpheme *-ing* that follows the target word would be the element that yields categorization).

Frames, on the other hand, are defined as “ordered pairs of words that frequently co-occur with exactly one word position intervening (occupied by any word).” (Mintz, 2003: 93). For example, the sequence *they come here* could be considered a frame, where both the strong subject pronoun (i.e. *they*) as well as the following adverb (i.e. *here*) would be strong indicators that the intervening word in the middle is a verb.

Empirical evidence shows that bigrams and frames behave differently in terms of word categorization, as each distributional context gives very different results. Thus, while frames appear to be particularly good categorization contexts in terms of accuracy, those higher accuracy scores using frames are obtained at the expense of losing completeness strength (i.e. fewer elements are captured by frames than by bigrams). Therefore, frames provide high accuracy but weak completeness scores, while the opposite pattern is true of bigrams: they give higher completeness scores but less overall accuracy. In fact, one of the problems of distributional contexts that has been pointed out is that, when using frames for word categorization tasks, several grammatical categories might emerge, instead of a single adult-like one (Mintz, 2003).

A further limitation that a learner who relies on distributional information is likely to encounter is that of non-immediate adjacency. As noted by some authors (e.g. Chomsky, 1975; Pinker, 1987), distributional regularities in English are not always local, but can occur over a variable distance, as shown in the examples below:

- (1) a. The house has two floors.
- b. The lovely big old white wooden house has two floors.

Patterns of lexical adjacency are variable in English. Thus, in the case of nouns and determiner-noun adjacencies, there can be a variable number of intervening modifying elements between the determiner and the noun. A learner who relies on strictly local distributional information and categorizes only from fixed positions could get to the wrong generalization that, for instance, *lovely* in (1b) above is a noun, and not an adjective.

How does the learner know which environments are important (as the one in (1a)), and which ones should be ignored? What kind of distributional cues are there in the linguistic input that might help children group individual lexical items into a larger and more general grammatical category? How reliable are such cues? To which extent could children work out the category for nouns in English accurately and on the basis of distributional information alone? The present paper aims at examining the kind and the amount of distributional cues which are present in the input addressed to English-learning children and which would allow the successful categorization of nouns in English. In what follows, the details of our analysis and the results obtained will be discussed.

3. Objectives

Several authors have already pointed out the usefulness of syntactic or distributional information for grammatical categorization (e.g. Cartwright and Brent, 1997; Mintz, 2003; Redington et al., 1998). Most of the studies that have been carried out to date are based on English data. Due to the limited number of English inflectional morphemes, distributional contexts or frames that have proved to be effective in the categorization of nouns in English are often made up of closed class items (e.g. pronouns, prepositions determiners) as well as open class words (e.g. lexical verbs or adjectives). As a consequence, a considerable number of possible categorizing frames emerge, and they are different across different corpora or different

Download English Version:

<https://daneshyari.com/en/article/934674>

Download Persian Version:

<https://daneshyari.com/article/934674>

[Daneshyari.com](https://daneshyari.com)