Journal of Memory and Language 63 (2010) 197–209

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Linking infants' distributional learning abilities to natural language acquisition

Marieke van Heugten, Elizabeth K. Johnson *

Department of Psychology, University of Toronto, 3359 Mississauga Road North, Mississauga, Ontario, Canada L5L 1C6

ARTICLE INFO

Article history:
Received 3 September 2009
revision received 19 March 2010

Keywords:
Distant dependencies
Statistical learning
Morphosyntax
Language acquisition
Linguistic input

ABSTRACT

This study examines the link between distributional patterns in the input and infants' acquisition of non-adjacent dependencies. In two Headturn Preference experiments, Dutch-learning 24-month-olds (but not 17-month-olds) were found to track the remote dependency between the definite article *het* and the diminutive suffix *-je* while no such evidence was obtained for the remote dependency between the definite article *de* and the plural suffix *-en*. In a follow-up corpus analysis, the distributional statistics in children's input (i.e. frequency, forward and backward transitional probabilities, and average distance between the two elements) were found to elegantly align with the behavioral data; distributional properties of diminutive and plural dependencies differed substantially, with more advantageous patterns for diminutive than for plural dependencies. Our results thus support the notion that there is a strong link between input distributions and the ease with which children acquire sensitivity to remote dependencies. Potential implications are discussed.

Introduction

Theories of language development often focus on how children start acquiring the meaning of words. Comprehending language, however, entails more than simply accessing the dictionary definition of consecutive words in the mental lexicon. In order to understand sentences, it is also crucial to determine how words relate to one another. Relationships between elements in sentences are plentiful in natural languages. While determiners, for example, are typically followed by a noun, pronouns or auxiliaries tend to be followed by verbs (e.g., *a book* but not **a is reading*; *he is reading* but not **he book*). The question thus arises how listeners start processing these interdependent co-occurrences.

The skills to acquire dependencies are arguably in place from early on. Using artificial languages, both adults and infants have repeatedly been shown to possess the ability to track adjacent dependencies (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and, in the presence of sufficient convergent evidence in the input, form relationships between abstract linguistic categories (Frigo & McDonald, 1998; Gómez & Gerken, 1999; Gómez & Lakusta, 2004; Mintz, 2002; Saffran, 2001). At 12 months of age, for instance, infants presented with a training language consisting of *aX* and *bY* patterns (where *a*, *b*, *X*, and *Y* formed "grammatical categories", with *X* being bisyllabic and *Y* monosyllabic words) later generalized these patterns to novel instances of (bisyllabic) *X* following *a* and (monosyllabic) *Y* following *b* (Gómez & Lakusta, 2004). The ability to track such abstract categories could be of great value for the acquisition of natural language dependencies, in that this could be the mechanism that enables children in their second year of life to start grouping words into different word categories based on the immediately preceding element (Gerken & McIntosh, 1993; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Johnson, 2005; Kedar, Casasola, & Lust, 2006; Lew-Williams & Fernald, 2007;

* Corresponding author. Fax: +1 905 569 4326.
E-mail address: elizabeth.johnson@utoronto.ca (E.K. Johnson).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.04.001

Shi & Melançon, in press; Shipley, Smith, & Gleitman, 1969; Van Heugten & Johnson, in press; Van Heugten & Shi, 2009; Zangl & Fernald, 2007).

Children thus readily acquire the relationship between frequently co-occurring adjacent elements in natural languages. These elements of a dependency, however, need not be (and are often not) adjacent. A determiner and noun, for example, may very well be separated by an adjective (e.g., *the colorful ball*). Moreover, some dependencies are never adjacent. Consider the present continuous tense, in which a form of *to be* is paired with the suffix *-ing* (e.g., *is singing*) or the English plural dependency consisting of a plural determiner and the plural suffix *-s* (e.g., *these balls*). These co-occurrences never occur adjacently, but can nonetheless potentially be greatly informative. In fact, corpus studies of child-directed speech have shown that frequent frames, combinations of two non-adjacent frequently co-occurring function words, are of considerable predictive value for determining the word category of the intervening syllables (Chemla, Mintz, Bernal, & Christophe, 2009; Mintz, 2003). The (remote) co-presence of the function words *the* and *in*, for example, is almost always interceded by a noun, while the function words *you* and *it* are typically interceded by a verb. Basic distributional co-occurrence patterns in the input could thus be used to categorically group intervening words into separate categories. Needless to say, this categorization cue can only be used once infants gain sensitivity to these frequent remote dependencies.

Headturn Preference studies have shown that the ability to track such non-adjacent co-occurrences develops early in life. Santelmann and Jusczyk (1998) tested English learners on their sensitivity to the non-adjacent present continuous *is-ing* dependency. Infants in this study were alternately presented with some passages containing grammatical (e.g., *is baking bread*) dependencies and some passages containing ungrammatical dependencies (e.g., **can baking bread*). To create ungrammatical dependencies, the grammatical auxiliary *is* was replaced by the ungrammatical auxiliary *can*. Eighteen-month-olds listened significantly longer to grammatical as compared to ungrammatical passages, indicating that they have already acquired some sensitivity to the remote *is-ing* dependency. Similar results have been obtained for comparable verbal dependencies in German and Dutch (Höhle, Schmitz, Santelmann, & Weissenborn, 2006; Wilsenach, 2006 respectively).

While children thus possess the ability to track at least some dependencies from early on, artificial language studies have suggested that the ease with which they are learned depends on various factors. Perceptual cues such as the phonological similarity between the elements of the non-adjacent dependencies (Onnis, Monaghan, Richmond, & Chater, 2005), the use of presegmented units (Peña, Bonnatti, Nespor, & Mehler, 2002), and the units the dependency consists of (Bonatti, Peña, Nespor, & Mehler, 2005; Newport & Aslin, 2004) have all been argued to play an important role in non-adjacent dependency acquisition.

In addition to the nature of the non-adjacent elements, a continuously growing body of research demonstrates that the exact distributional properties of the materials determine co-occurence acquisition. The frequency of the dependency, its strength (both expressed in forward and backward transitional probabilities (TPs), the distance between elements in the dependency, and the variability in intervening syllables, for example, all affect how readily dependencies are acquired. The frequency of co-occurrence (e.g., Braine et al., 1990; as used by Mintz (2003)) may be the most intuitive measure of all and forms a simple count of the number of times both elements of a dependency co-occur. In general, the more frequently the two items co-occur, the easier the dependency is to learn. Related to frequency, but quantified as a relative number are the forward (e.g., Aslin et al., 1998; Morgan, Meier, & Newport, 1987; Saffran, 2001; Saffran, 2002; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996) and backward (Gervain, 2008; Pelucchi, Hay, & Saffran, 2009; Perruchet & Desaulty, 2008) TPs. Consider, for example, the co-occurrence of *the* and *dog*. Although the phrase *the dog* may occur relatively frequently in English, *the* can also be followed by numerous other nouns. Similarly, *dog* is not solely preceded by *the*, but also by other determiners such as *a*, *every*, or *that*. Taking into account the instances in which the individual forms occur with other words may thus prevent listeners from erroneously accepting *the dog* as one unit. Forward and backward TPs take into account this base frequency of the individual words. More specifically, forward TPs are defined as the probability of encountering the second element of a dependency (i.e. *dog* in the above example) given the first (*the*) and backward TPs are defined as the probability of encountering the first element (*the*) given the second (*dog*). The higher the TPs, the more strongly related the two elements are, and, ceteris paribus, the easier they are to track. Both frequency and TPs, as well as the nature of listeners' sensitivity to these distributional cues (Aslin et al., 1998; Mirman, Graf Estes, & Magnuson, in press; Perruchet & Peereman, 2004), have been discussed at length in the literature examining the acquisition of adjacent dependencies. Because the acquisition of adjacent and non-adjacent dependencies are strongly related (e.g., Lany & Gómez, 2008), mechanisms employed in acquiring one dependency are likely also used in the acquisition of the other dependency and should hence be taken into account. A further distributional property that may affect dependency acquisition is the average distance between the two non-adjacent elements. Infants only have a limited processing window and tracking dependencies over long distances is more demanding than tracking near-by dependencies (Santelmann & Jusczyk, 1998), causing shorter-distance relationships to be learned more easily than longer-distance relationships. A final factor considered here is the variability in intervening material. Artificial language studies have shown that non-adjacent dependencies are learned with greater ease when the intervening material is more diverse (Gómez, 2002; Gómez & Maye, 2005). It has been argued that the less likely the intervening material will form an adjacent dependency, the more likely listeners' focus will be drawn towards longer-distance dependencies. More variability in the interceding syllables may thus facilitate non-adjacent dependency learning.

Artificial language studies have thus generated important hypothesis concerning what dependencies should be easy and what dependencies should be challenging to