Introduction

- Syntactic categorization (like noun and verb) are the building blocks of grammar. Lexical categorization is thus a crucial part of language learning.
- In child-directed speech, distributional patterns contained in linear word sequences have been informative categorization cues (Mintz, Newport, & Bever, 2002; Mintz, 2003; Redding, Crater, & Finch, 1998).
- Some early criticisms of distributional analyses were that linear sequence information would be ambiguous or misleading if not constrained by structural information (Pinker, 1984).
- Mintz (2003) demonstrated that frequent frames are extremely accurate at categorizing words.
- A frame is defined as two jointly occurring words with one word intervening. Frequent frames are frames occurring frequently in a corpus.
- The procedure was repeated on and compared between each frame type, the proportions of the four most frequent GR patterns were computed.
- Two similar environments - F_F1 and F_F2 (where the most frequent co-occurring two adjacent words categorize the words directly before or after them).

Results

- For GR, GR, FF of FFs, on average 92% of tokens in each frame have the same pair of syntactic relations (F1 & F2).
- Comparing FFs to other environments.
- The proportion of tokens accounted for by the four most frequent GR, GR, & FF, is significantly higher for FFs (65% & 44%) than for bigrams (75% & 59%), and GR, GR, GR, GR, (81% & 64%, all p<.001).
- The proportion of tokens accounted for by the four most frequent GR, GR, & FF for FF (92%) account for the largest proportion of tokens.
- For F_F1 and F_F2, GR, GR, GR, and GR, are the highest (86% and 88%, respectively) compared to the other three GR patterns.
- Suggests that the context words of frequent frames have the most coherent and consistent relations.
- The exceptionally high proportions of the three GR patterns confirm our prediction that the grammatical structures involved in FFs are dominated by a restricted set of structures, which consequently constrain the possible word categories occurring in the target position of a FF.

References


Questions

- Why are frequent frames (FFs) so good at categorization?
- Grammars developed over the past 50 years account for structural regularities within and across languages by proposing hierarchical organization of words.
- Yet frequent frames are defined over simple linear sequences. Why are they such robust category cues when grammatical categories are abstract concepts involved in hierarchically organized syntactic structure?

Hypotheses

- A frequent frame arises from a restricted set of syntactic structures that constrain the category in target position; the grammatical relations between words in a FF are consistent and stable across instances of FFs.