# Distributional signatures of superordinate nouns

**Introduction** Taxonomies of object categories (and their labels) often include the superordinate (e.g., *animal*), basic (e.g., *dog*), and subordinate (e.g., *dalmatian*) levels of specificity [1,2]. Of these, superordinate nouns are notoriously difficult to learn [3,4], perhaps because their members are intractably diverse. This has led to proposals that young children's conceptual representation of superordinate categories may be underdeveloped [5,6]. Here, setting aside such conceptual issues, we posit that superordinate nouns are also challenging for acquisition at least in part due to the difficulty of mapping their meanings to referents in the world.

**Hypothesis** We hypothesize that the choice of a superordinate label encodes a particular pragmatic level of informativeness [7,8], which lends itself to certain functions. These functions should, in turn, lead to specific linguistic contexts of use. Thus, we expect distributional signatures of superordinate meanings in child-directed input. We focus on two such functions leading to five distributional signatures. For one, superordinate labels can be used to communicate a lack of specialized knowledge [9], and so should be frequent in wh-phrases within information-seeking questions. Superordinates also serve the function of generalizing over sets and instances of many types [10], and so should occur often with universal quantifiers and plurals [11,12]. Additionally, because of their role in generalizations, superordinates likely involve "anchoring" cues suggesting kind-inclusion (e.g., *This is a wug. A wug is a kind of terval. This is another terval.*) [13,14]. We probe whether these five distributional features indeed support the superordinate over basic-level meanings in child-directed input. We also consider two control features not expected to facilitate superordinate conjectures (labelling and co-occurrence with definites).

**Method** Child-directed English data for children between 1 and 6 years of age ($n_{children}$=597) were extracted from CHILDES [15]. Seven superordinate-level nouns and their closest frequency-matched basic-level counterparts entered the analysis ($n_{tokens}$=17,514; Table 1). Each utterance containing a token was tagged for the five target and two control linguistic features above (Table 2) using a combination of regex and the spaCy parser [16].

**Results** Our broad predictions were confirmed (Figure 1). Superordinate nouns more frequently headed wh-phrases (3.9%) and participated in structures involving "type/kind of" (3.2%) and "(an)other" (4.9%) compared to close basic-level nouns (under 1% for all). Furthermore, superordinate nouns combined more often with universal quantifiers (66.8% vs. 12.1%) and plurals (12.6% vs. 1.3%). However, no clear advantage for superordinates was observed in contexts with ostensive labelling or definite articles.

**Conclusion** Superordinate and basic-level nouns differ in pragmatic levels of informativity in ways that affect contexts of use. Strikingly, these nouns show distinct distributional signatures, despite sharing a more abstract kind of semantic organization compared to other, traditionally studied areas that benefit from distributional learning [17,18]. Conceptual issues aside, the mapping challenge for superordinates may be solved by attending to distributional cues, consistent with pragmatic models of acquisition where language input guides the organization of semantic domains [8,19,20]. We predict that these cues, if experimentally manipulated, should support the acquisition of superordinate nouns in both adults and children.

Word Count: 500

| Type | Word | Tokens |
|---|---|---|
| Superordinate | toy | 3,317 |
| Basic | ball | 5,160 |
| Superordinate | animal | 2,829 |
| Basic | cat | 2,843 |
| Superordinate | tool | 427 |
| Basic | fork | 419 |
| Superordinate | building | 327 |
| Basic | hospital | 423 |
| Superordinate | fruit | 373 |
| Basic | strawberry | 319 |
| Superordinate | vegetable | 313 |
| Basic | pepper | 274 |
| Superordinate | dessert | 237 |
| Basic | waffle | 253 |

**Table 1.** The frequency-matched basic and superordinate noun pairs used in the analysis and their number of tokens.

| Feature | Examples |
|---|---|
| "type/kind of" | "What kind of building is it?" "That's the name of the kind of toy that you're playing with." |
| wh-phrase | "What animals do you want to play with?" "Which toy do you like the best?" |
| "(an)other" | "Carrots, beets, what other vegetable do you like?" "Let Mommy get you another toy." |
| Quantifiers | "Wanna go in the garden and pick some vegetables?" "That's where all the tools go." |
| Plural | "Oh, look at these animals here." "Go pick up your toys please." |
| Labelling | "Yes, it is a little stuffed animal." "This is a ball." |
| Definites | "Are we all done with the tools?" "The cat jumped out right behind the tree." |

**Table 2.** The seven linguistic features that each utterance containing a noun token was tagged with. An utterance could be tagged for multiple features. Examples are from the extracted sample of CHILDES.
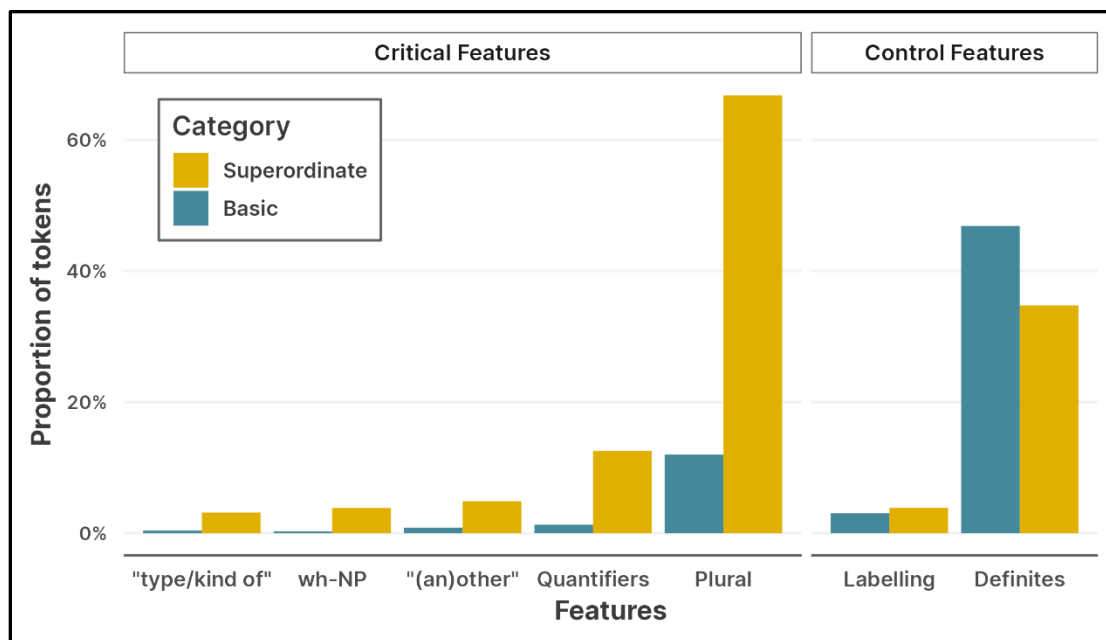


**Figure 2.** The distribution and frequency of linguistic features for superordinate and basic-level nouns. For example, 66.8% of all superordinate nouns in the sample of tokens occur in the plural form. All pairwise comparisons between superordinate and basic-level categories are statistically significant, which is trivially true given the large number of tokens – this is omitted from the graph.

**References**: [1] Rosch et al., 1976. *Cognitive Psychology*. [2] Markman, 1990. *Cognitive Science*. [3] Mervis & Crisafi, 1982 *Child Development* [4] Horton & Markman, 1980, *Child Development* [5] Golinkoff et al., 1995. *Development Psychology*. [6] Xu & Tanenbaum, 2007. *Psychological Review*. [7] Benelli, 1988, *Human Development*. [8] Choe & Papafragou, 2023, *Journal of Memory and Language*. [9] Levinson, 2001. *Presumptive Meanings. Cog. Psy.* [10] Nelson, 1978. *Children's language*. [11] White, 1982. *Journal of Experimental Child Psychology* [12] Wisniewski & Murphy, 1989. *Discourse Processes*. [13] Callanan, 1989. *Developmental Psychology*. [14] Shipley, Kuhn, & Madden, 1983. *Journal of Child Language*. [15] MacWhinney, 2000. *The CHILDES project*. [16] Honnibal et al., 2020. *spaCy (v3.5)*. [17] Gleitman, 1990. *Language Acquisition*. [18] Gleitman, et al., 2015. *Language Learning and Development*. [19] Clark, 1997. *Cognition* [20] Clark, 2017. *Language Learning and Development*.