

# Morphosyntactic and Referential Cues to the Identification of Generic Statements

Phil Crone

pcrone@stanford.edu  
Department of Linguistics  
Stanford University

Michael C. Frank

mcf Frank@stanford.edu  
Department of Psychology  
Stanford University

## Abstract

Generic sentences (e.g. “Birds fly.”) express generalizations about kinds, as opposed to non-generic sentences that are about specific individuals or groups of individuals (e.g. “All birds fly.”). We investigate how language users use morphosyntactic and pragmatic cues to determine whether naturalistic sentences should receive generic interpretations. Experiment 1 demonstrates the effect of morphosyntactic features of a sentence’s subject noun phrase (NP) on generic interpretation. Experiments 2 and 3 reveal that when a sentence’s subject NP does not have an obvious reference in context, the sentence is more likely to receive a generic interpretation.

**Keywords:** pragmatics; generics

## Introduction

Generic sentences differ from non-generic sentences in that they express generalizations about kinds rather than properties of specific individuals or sets of individuals. For example, the sentence “Birds fly” express a general property of the kind *bird*, whereas the sentence “All birds fly” states that for every member  $x$  of the set consisting of all birds,  $x$  flies. A key difference between generic and non-generic statements is that generics allow for exceptions. “Birds fly” is true despite the fact that some birds do not fly. “All birds fly” is false in virtue of the fact that there are individuals that are birds and do not fly (Prasada, 2000).

We can identify two distinct puzzles that generics pose for the study of natural language semantics and pragmatics. The first is how to provide adequate truth conditions for generic sentences. These truth conditions must account for the fact that generics allow for exceptions and other peculiarities, such as the fact that generics may be judged true even when the generalization does not hold for most members of the kind. A second puzzle is how language users solve the problem of identifying whether a sentence should receive a generic or non-generic interpretation; this problem arises because sentences are often ambiguous between generic and non-generic interpretations. The current study is concerned with the second puzzle.

Individuals use three types of cues to guide their interpretation of sentences as generic or non-generic: morphosyntactic features, pragmatic cues, and world knowledge (Cimpian & Markman, 2008; Cimpian, Meltzer, & Markman, 2011; Gelman & Raman, 2003). In English, the subject NP of a generic sentence is often a bare plural (“Birds fly.”), but indefinite singular (“A bird has wings.”) and definite singular (“The bird is a warm-blooded animal.”) NPs can also serve as subjects of generic sentences. Definite plural NPs (“The birds have feathers.”) are generally thought to force non-generic interpretations. Tense and aspect also cue whether it is to

be interpreted generically. Generic sentences tend to use the simple present tense (“Birds fly.”), as opposed to the present progressive (“Birds are flying overhead.”), past tense (“Birds flew past my window.”), or tense/aspect categories (Carlson, 1977; Krifka et al., 1995; Lyons, 1977).

In addition to these morphosyntactic cues, the preceding discourse and nonlinguistic factors may influence whether a sentence is interpreted as generic or non-generic. For example, if a unique bird is present in the context of an utterance of a sentence with the subject NP “the bird,” a non-generic interpretation in which this NP refers to the bird in context may be more likely. Conversely, if no such bird exists in the context, a generic interpretation may be preferred. Finally, world knowledge about the properties shared by members of a kind will influence the interpretation of potentially generic sentences. The sentence “A bird does not fly” is interpreted as a non-generic sentence about some particular bird (e.g. a penguin), given world knowledge that, in general, birds fly.

Previous experimental work has demonstrated the relevance of these three factors to the identification of generic sentences. Gelman and Raman (2003) show that adults and children as young as 3 show a preference for interpreting bare plurals as generic, as compared to definite plurals, and are more likely to interpret sentences as generic when the subject NP has no available referent in context. Cimpian and Markman (2008) demonstrate that by age 3 children are less likely to assign a generic interpretation to a sentence when its subject NP has a possible referent in the preceding linguistic context. In addition, they show that children as young as 3 use knowledge about whether properties are generalizable to kinds as evidence about whether to interpret sentences as generic or not. Finally, Cimpian et al. (2011) show that 3-year-olds use definiteness of subject NPs as a cue to identifying generics and that adults and children as young as 4 use tense and aspect to identify generics.

The present study differs from previous work in several respects. The majority of previous work on the identification of generics has focused on children’s abilities, whereas the current study is primarily concerned with how adults identify generics. The focus on children’s identification of generics stems in part from the fact that children face an inductive problem not faced by adult language users regarding which types of NPs refer to kinds. However, recent work emphasizing the probabilistic nature of language comprehension (Frank & Goodman, 2012; Levy, 2008) suggests that adults face a similar problem. On this view, language users resolve uncertainty in language comprehension via probabilistic inference to the most likely interpretation. In the specific

case of identifying generics, we can view adults as reasoning about the likelihood of an utterance being generic given morphosyntactic features of the sentence, features of the context, and the listeners' world knowledge.

The current study also differs from previous work by collecting naturalistic examples of generic and non-generic sentences generated by study participants. This approach allows for a more realistic representation of how genericity is used in natural language. The approach also allows for consideration of examples that are more ambiguous between generic and non-generic interpretations than examples created by experimenters.

## Experiment 1: Morphosyntactic Cues

As discussed above, the number and definiteness of a sentence's subject NP influence its interpretation as generic or non-generic (Carlson, 1977; Krifka et al., 1995; Lyons, 1977). Previous work investigating morphosyntactic cues to genericity have fixed the number of the subject NP as either singular (Cimpian et al., 2011) or plural (Gelman & Raman, 2003) and only manipulated definiteness. In Experiment 1, we considered number, definiteness, and their interaction, on the interpretation of generics. We asked participants to perform a sentence completion task in which the subject NP was provided. Participants were then asked to indicate whether the sentences they produced were about specific individuals or kinds.

## Method

**Participants** We recruited 100 participants to participate through Amazon's Mechanical Turk website. Participants were restricted to individuals within the United States and were paid 50 cents to complete the study. The study took approximately 14 minutes to complete.

**Stimuli** Forty-eight nouns were chosen to use as the bases for subject NPs. To ensure diversity among these subject NPs, twenty-four nouns were animate and twenty-four were inanimate. For each study participant, a random set of twelve nouns were assigned morphosyntactic features using a  $2 \times 2$  factorial design crossing number (singular, plural) with definiteness (definite, indefinite). Exactly half of nouns assigned to each factorial point were animate and half were inanimate. Base nouns were then programmatically edited to reflect the assigned number and definiteness values and create full NPs. For example, if the noun "panda" were assigned values *plural* and *definite*, the full NP would be "the pandas."

In the first part of the experiment, participants saw a single NP followed by a single-line text box. They were instructed to "write a sentence starting with the phrase below." In the second part of the experiment, participants were shown the sentences they had written in the first part of the experiment. They were shown one sentence at a time and asked whether the sentence was about a specific *noun* (for singular NPs), a specific group of *nouns* (for plural NPs), or about *nouns* in general. Participants indicated their response using

a 5-point Likert scale with the following values: "Definitely about a specific *noun*/group of *nouns* (=1)," "Probably about a specific *noun*/group of *nouns*," "Not sure," "Probably about *nouns* in general," "Definitely about *nouns* in general (=5)."

**Procedure** We first presented participants with four example noun phrases. After these first four items, participants were shown an example sentence that they could have used for the example noun phrase. These sentences were constructed to favor non-generic interpretations for all NP types. After seeing the examples, participants were informed that they would not receive any feedback for the rest of the experiment.

Participants then began the first part of the experiment. All forty-eight subject NPs were presented in pseudorandom order counterbalanced so that no two consecutive NPs matched in both number and definiteness. We required participants to provide a sentence completion for each item and required that sentence completions be a minimum of six characters long. After completing part one, participants were told that they were entering the second part of the experiment and were informed that they would be evaluating sentences' genericity. They then evaluated the sentences they had produced in part one in the manner described above. Once again, sentences were presented in a pseudorandom order counterbalanced so that no two consecutive NPs matched in both number and definiteness. After judging all forty-eight sentences, they were then required to provide their native language. We measured reaction times for each item, measured from the time the item was presented until the time a response was submitted.

**Data Analysis** We excluded 4 participants for indicating that their native language was not English. Thus, responses from 96 participants were analyzed. In addition, responses from items whose reaction times were greater than 2 standard deviations from the mean for the second part of the experiment.

## Results & Discussion

Paragraph about figure.

To analyze these findings, we fit a linear mixed-effects model to model participants' ratings of their sentences for genericity. We examined the interaction between animacy, definiteness, and number of subject NPs.<sup>1</sup> The model shows main effects of animacy, definiteness, and number, with animate NPs ( $\beta = 0.428, t = 4.396, p < 0.001$ ), plural NPs ( $\beta = 0.227, t = 2.596, p < 0.001$ ), and indefinite NPs ( $\beta = 1.293, t = 10.497, p < 0.001$ ) being rated significantly more generic. In addition, the model revealed interaction effects between animacy and definiteness such that sentences with animate, indefinite NPs were rated significantly less generic ( $\beta = -0.571, t = -4.756, p < 0.001$ ) and between definite-

<sup>1</sup>All mixed-effects models were fit using the lme4 package in R version 3.1.2. The model specification was as follows: `response ~ animacy * definiteness * number + (animacy + definiteness + number | WorkerId) + (definiteness + number | subject)`.

ness and number such that bare plurals were rated significantly more generic ( $\beta = 0.488, t = 4.725, p < 0.001$ ). Finally, there was a significant three-way interaction such that animate, bare plurals were rated significantly more generic ( $\beta = 0.536, t = 3.669, p < 0.001$ ).

The results are consistent with previous findings that indefinite singulars and bare plurals facilitate generic interpretations compared to definite singulars and definite plurals, respectively (Cimpian et al., 2011; Gelman & Raman, 2003). However, these results also show that plurality is independently associated with genericity, which has not previously been acknowledged.

**Experiment 2A: Contextual Cues in Production**

First level headings should be in 12 point, initial caps, bold and centered. Leave one line space above the heading and 1/4 line space below the heading.

**Method**

**Participants**

**Stimuli**

**Procedure**

**Data Analysis**

**Results & Discussion**

**Experiment 2B: Contextual Cues in Comprehension**

First level headings should be in 12 point, initial caps, bold and centered. Leave one line space above the heading and 1/4 line space below the heading.

**Method**

**Participants**

**Stimuli**

**Procedure**

**Data Analysis**

**Results & Discussion**

**Experiment 3: Ambiguous Sentences**

First level headings should be in 12 point, initial caps, bold and centered. Leave one line space above the heading and 1/4 line space below the heading.

**Method**

**Participants**

**Stimuli**

**Procedure**

**Data Analysis**

**Results & Discussion**

**General Discussion**

Use standard APA citation format. Citations within the text should include the author’s last name and year. If the authors’ names are included in the sentence, place only the year in parentheses, as in ? (?), but otherwise place the entire reference in parentheses with the authors and year separated by a comma (?, ?). List multiple references alphabetically and separate them by semicolons . Use the “et al.” construction only after listing all the authors to a publication in an earlier reference and for citations with four or more authors.

**Footnotes**

Indicate footnotes with a number<sup>2</sup> in the text. Place the footnotes in 9 point type at the bottom of the column on which they appear. Precede the footnote block with a horizontal rule.<sup>3</sup>

**Tables**

Number tables consecutively. Place the table number and title (in 10 point) above the table with one line space above the caption and one line space below it, as in Table 1. You may float tables to the top or bottom of a column, or set wide tables across both columns.

Table 1: Sample table title.

Error type	Example
Take smaller	63 - 44 = 21
Always borrow	96 - 42 = 34
0 - N = N	70 - 47 = 37
0 - N = 0	70 - 47 = 30

**Figures**

All artwork must be very dark for purposes of reproduction and should not be hand drawn. Number figures sequentially, placing the figure number and caption, in 10 point, after the figure with one line space above the caption and one line space below it, as in Figure 1. If necessary, leave extra white space at the bottom of the page to avoid splitting the figure and figure caption. You may float figures to the top or bottom of a column, or set wide figures across both columns.

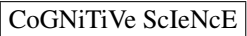


Figure 1: This is a figure.

**Acknowledgments**

Place acknowledgments (including funding information) in a section at the end of the paper.

<sup>2</sup>Sample of the first footnote.  
<sup>3</sup>Sample of the second footnote.

## References Instructions

Follow the APA Publication Manual for citation format, both within the text and in the reference list, with the following exceptions: (a) do not cite the page numbers of any book, including chapters in edited volumes; (b) use the same format for unpublished references as for published ones. Alphabetize references by the surnames of the authors, with single author entries preceding multiple author entries. Order references by the same authors by the year of publication, with the earliest first.

Use a first level section heading, “**References**”, as shown below. Use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 1/8 inch. Below are example references for a conference paper, book chapter, journal article, dissertation, book, technical report, and edited volume, respectively.

## References

- Carlson, G. N. (1977). *Reference to Kinds in English*. Ph.D. dissertation. University of Massachusetts, Amherst.
- Cimpian, A., & Markman, E. (2008). Preschool children's use of cues to generic meaning. *Cognition*, 107(1), 19–53.
- Cimpian, A., Meltzer, T. J., & Markman, E. (2011). Preschoolers' use of morphosyntactic cues to identify generic sentences: Indefinite singular noun phrases, tense, and aspect. *Child Development*, 82(5), 1561–1578.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998.
- Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, 74(1), 308–325.
- Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., & Chierchia, G. (1995). Genericity: An introduction. In *The Generic Book* (pp. 1–124). Chicago: The University of Chicago Press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Lyons, J. (1977). *Semantics: Volume 1*. New York: Cambridge University Press.
- Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Science*, 4(2), 66–72.