Zero anaphora and object reference in Japanese child-directed speech

Cybelle Smith
cybelle@umd.edu
Department of Linguistics
University of Maryland

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract
To learn the meanings of words, children must connect referents in the world around them with the sounds they hear. One proposed mechanism for this process is cross-situational word learning: tracking associations between words and objects across time. We consider the problem of anaphora for a cross-situational word learner: after an object has been introduced it is unlikely to be named in every succeeding reference to it. This problem is particularly pronounced in Japanese, which uses “zero anaphora,” where pronouns can be omitted from utterances. We analyze a corpus of Japanese mothers talking to children about sets of objects, originally recorded by Fernald and Morikawa (1993). Overall rates of anaphora were much higher for Japanese mothers compared with English mothers. Zero anaphora was primarily used when the discourse topic was already established, suggesting that a discourse-finding strategy may be important for word learning in Japanese. In addition, unexpectedly, due to the existence of zero anaphora as a common referential strategy, pronouns were more likely to be used when the topic was new than when it was given (reversing common results for English).

Keywords: Child-directed speech; Japanese; zero anaphora; discourse analysis; language acquisition.

Introduction
From the perspective of a scientist, early word learning seems a difficult problem. Although infants quickly learn to pair sound sequences in their caregivers’ speech with concepts and entities in their environment, it is still uncertain how they associate words and concepts. One proposed mechanism for this process is cross-situational word learning: tracking associations between words and objects across time (Siskind, 1996; Yu & Smith, 2007). Though possible for even very young children in simple contexts (Smith & Yu, 2008) and in principle feasible for large lexicons (Blythe, Smith, & Smith, 2010), the effectiveness of pure cross-situational learning in natural contexts is still unknown.

One issue for cross-situational learning in the natural learning environment is anaphora: the use of shortened—or even absent—expressions that refer back to a previously-named entity. If learners are keeping track of associations between words and objects, the tendency towards anaphoric reference should cut down considerably on these associations. Although anaphora is pervasive in language use, object names are typically repeated frequently in the speech of English-speaking mothers to their children (Fernald & Morikawa, 1993). This repetition has allowed models of cross-situational word learning to succeed in establishing word-object mappings even in small natural datasets (Yu & Ballard, 2007; Frank, Goodman, & Tenenbaum, 2009).

Repetition of object labels is not nearly as prevalent in some other languages, however. In the same study that established the presence of repetition in English mothers’ speech, Fernald and Morikawa (1993) noted that Japanese mothers used far fewer noun labels and the labels they used were generally more diverse, including onomatopoeia and diminutive forms as well as the prototypical labels used by American mothers. In addition, Japanese, unlike English, is a pro-drop language, meaning that the subject and object of verbs may be omitted. This omission is known as “zero anaphora.” For example, when an English speaker might say “the dog barks,” a Japanese speaker might say only “barks.” One study suggests that zero anaphora in Japanese may lead to increased difficulty in early verb learning (Rispoli, 1995), but to our knowledge, no work has examined the direct relationship between zero anaphora and object reference, in Japanese or any other language.

What effect does the varied use of noun-labels and anaphora (especially zero anaphora) have for Japanese word learners? Under a pure cross-situational analysis, the sparse mappings between words and objects in this language might be very difficult to overcome. If objects that are being talked about often go unnamed in Japanese, a pure cross-situational learner might be more likely to learn, for example, “bark” rather than “dog” for the concept of a dog. If, however, word learners do not treat utterances as independent entities, but instead resolve reference within a topical discourse—a set of utterances about a particular topic—the problems posed by changing object labels and zero anaphora might be mitigated. A learner could figure out what topic was being talked about and then assume that future utterances refer to this topic, even if it was not named.

Recent work has taken up the suggestion that children could potentially aggregate information about word meanings—as well as other knowledge about a particular object referent—not just across sentences but also across these topical discourse units (Rohde & Frank, under review; Frank, Tenenbaum, & Fernald, in press). On this kind of view, a first utterance establishes the topic (in the cases we examine, often a simple object referent), and then future utterances contribute new information (Clark, 1996). For such a learner, zero anaphora might not be as problematic if the discourse topic were already known.

The current study examined zero anaphora in Japanese from this perspective. We conducted a reanalysis and annotation of Japanese infant- and child-directed speech from the Fernald and Morikawa (1993) study, focusing on anaphora. We asked when zero anaphora was used within topical discourses, in comparison with object naming and the use of other pronouns. We found that although overall rates of anaphora were much higher for Japanese mothers compared
with English mothers, zero anaphora was primarily used when the discourse topic is already established. Also, we found a trade-off between zero anaphora and pronominal anaphora that caused pronouns to be more likely when the discourse topic was newly established. This sensitivity of zero and pronominal anaphora to the discourse topic suggests that a discourse-finding strategy may be even more important for Japanese-learning children than it is for English learners.

Methods

Corpus Materials

Our data consisted of a set of transcribed videos of object-centered play between mothers and children in their homes, from a study by Fernald and Morikawa (1993). While the original corpus contained both American and Japanese mother-child pairs, the current analysis focuses primarily on the Japanese mothers (American data was analyzed in Frank et al., in press). Discourse from 29 Japanese mother-child pairs with audio and video data was analyzed. The infants were divided into three age groups: 5-6 months (N=9, 5 males), 11-14 months (N=10, 5 males), and 18-21 months (N=10, 5 males).

Prior to recording, mother and child played comfortably together with the child’s toys. Next, the child’s own toys were removed and the video recording began. During the video the mothers were asked to play with the child using three standardized pairs of toys: dog and pig, car and truck, and brush and box. The mother was asked to play with the toys “as she normally would.” The toys were introduced one pair at a time and removed before introduction of the following pair. The ordering of whether the dog and pig were introduced first or the car and truck were introduced first was counterbalanced across trials, but the brush and box were always introduced last (and only for the older two groups).

Towards the end of the play session, mothers of children from the two older age groups were asked to hide the toys and get the child to retrieve them using words alone. Because this scenario might affect how the mother referred to the objects (indeed it was inserted in order to elicit object names), we only considered utterances prior to this “hiding game.” We also excluded utterances with sound and audio issues (167), and those spoken by the mother to the experimenter. In total, 8852 utterances taken from 6 hours and 51 minutes of video were analyzed in the current study.

Conventions for Annotating Object Reference

A native Japanese speaker first divided the mothers’ speech into “utterances,” or segments of speech separated by pauses, on the basis of prosodic and syntactic cues. Most utterances ranged from a single word to a complete sentence; complete sentences were usually not counted as multiple utterances unless there was a pause or interruption of the speaker’s turn.

Next, using the video and transcript data, a native Japanese speaker annotated, for each utterance spoken by the mother to the infant, what object or objects (if any) were being referred to. An object was considered to be referred to if A) the mother said the name of the object or B) the mother used a pronoun that the annotator judged to refer to the object. Although other toys and objects were occasionally referred to in the corpus, in the following analysis we will only examine references made to the six toys that were standardized across participating dyads (dog, pig, car, truck, brush, and box).

In Japanese, baby words for toys are often derived from onomatopoeia. We counted misnomers and onomatopoeia as references to the toy under clearly referential circumstances. For misnomers, this was when e.g. “moomoo” (“moo-moo,” meaning “cow”) was used in reference to the pig. We also counted alternative labels, such as “omocha” (“toy”) and “nuigurumi” (“stuffed animal”), as references. However, we did not count misnomers and alternative labels as cases of “object naming,” which we defined in a more restricted sense, described below. We counted onomatopoeia-derived noun phrases as references to objects, but not onomatopoeia that was used to describe sounds or actions. Our annotator made judgments about when the mother was using phrases such as “wanwan” as a noun and when she was using them to indicate sounds or actions that the toy was making. Thus, “wan-chan” (“Mr. woof”) and, in some cases, “wanwan” (“woof-woof”), were coded as referring to nouns.

In Japanese, objects are frequently referred to without use of an explicit noun or pronoun. As noted above, grammat-
Results

Mothers talked about each toy in alternating bouts of utterances. They frequently used onomatopoeia, and engaged in “social routines” (Fernald & Morikawa, 1993), such as requesting objects from the child and saying thank you. A plot of references to the six toys over a single video is presented in Figure 2, as a representative discourse structure.

We report three main sets of analyses. First, simple univariate counts of pronominal and zero anaphora. Second, analyses of transitions between different kinds of reference. Third, changes in use of zero anaphora across the corpus.

Anaphora Use Across Languages

Our first analysis counted pronominal and zero anaphora proportions in the corpus at each age. We normalized these counts by the total number of object references that were identified (including all anaphoric references). These data are shown in Figure 1, along with English data on pronoun use in object references from Rohde and Frank (under review).

Several trends are apparent in these data. First, pronoun use is approximately equivalent between English and Japanese speakers. Speakers of both languages use pronouns approximately one-third of the time. Second, Japanese use of zero anaphora constitutes an additional third of all object references. Unlike the English-speaking mothers, Japanese-speaking mothers were using anaphora more often than not to refer to the toys that they were playing with. Finally, pronoun use declined with the age of the children in our sample. This trend was somewhat modest compared with the large cross-linguistic differences, but was nevertheless significant in a simple linear regression predicting pronoun use by age in months (β = -.005, p <.001). However, this age difference likely reflects the fact that videos were shorter for the 6 month old group than the other two groups. As will be discussed below, pronoun use declined substantially over the discourse for all three age groups, and so proportion of pronoun use in the 6 month old group might have been inflated by the shorter video—and hence shorter discourse—length.

Discourse Continuity in Japanese

We next examined how likely each mother was to refer to the same object in two consecutive utterances. This analysis was used by Frank et al. (in press) as a first-pass indicator that discourse references to objects were relatively continuous. A high probability of repeated reference to an object suggests that a learner who “smoothed” their guesses about reference across time would be relatively successful. If they didn’t know what a particular utterance was referring to, they could just guess that the referent was the same as in the previous utterances in the discourse.

Continuity of Reference We calculated transition probabilities between referential and non-referential utterances of
various types. As in our previous work, transition probabilities were first calculated for each of the six toys, then averaged together, weighted by the number of times each toy was referred to in the video (or referred to using a particular reference type, as appropriate).

Mothers varied considerably in how likely they were to refer to an object in two consecutive utterances. The probability of referring to an object in the next utterance given it was referred to in the current utterance ranged between .20 and .57 for each mother (mean = .43). Probability of repeated reference remained relatively stable when calculated for the utterances in each age group, although it tended to rise with the age of the child, from .39 at 6 mos. to .44 at 12 mos, and .47 at 18 mos. Overall these levels were somewhat lower than those for English-speaking dyads.

When we removed utterances that consisted of backchanneling—checking for a response from the child—such as “mn,” “n?,” “a,” and “hai hai hai” (“yes, yes, yes”), this slightly increased the probability of consecutive reference (mean = .49), and decreased variance between mothers. This analysis suggests that the higher tendency of Japanese mothers to use backchanneling, combined with the lower overall frequency of object reference, may account for the difference between Japanese and English-speaking dyads.

Transitions Between Reference Types We next examined transitions between referential and non-referential utterances of various types, including zero anaphora, pronouns, and object naming. The goal of this analysis was to understand the directionality of zero anaphora use in discourse. If zero anaphora is used more after naming, then transition probabilities should be asymmetric: $p(\text{zero}|\text{name})$ should be higher than $p(\text{name}|\text{zero})$.

Results are summarized in Table 2. The largest trend was for referential utterances to be followed by non-referential utterances, indicating (as above) that Japanese discourses contained more non-referential speech overall than in English. Nevertheless, there were still distinct trends in which referential strategies were used earlier in discourses. Zero anaphora after object naming was three times as likely as object naming after zero anaphora. Zero anaphora after pronominal anaphora was three times as likely as pronominal anaphora after zero anaphora. Surprisingly, pronouns were more likely to be used before object naming than after it, unlike in English (Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993).

Changes in Anaphora Use Across the Discourse

Zero Anaphora Over Time In our third analysis, we examined how often mothers used zero anaphora at different points in the discourse. For each utterance that referred to one of the six main toys, we calculated the number of times the object had been mentioned prior to that utterance. We call this the “number of previous references” (NPR) for the utterance. If an utterance refers to the dog for the fifth time in a video, that utterance has an NPR of 4. Next, we collapsed all references with a particular NPR across mothers and toy referents, and calculated the proportion of the time the references used zero anaphora.

The proportion of time that mothers referred to an object using zero anaphora increased with the log of the NPR at all ages (Figure 3). We created a mixed effects model of zero anaphora usage, with age group and NPR as fixed effects and mother as a random effect with random intercept and slope with respect to NPR. Because of the very large number of observations, we used the $z$ approximation to esti-

Table 2: Transition probabilities between referential and non-referential utterances of various types. $P(Y|X)$ refers to the probability that in the current utterance an object is referred to using $Y$ given that in the previous utterance it was referred to using $X$. zero = zero anaphora, name = object naming, nonref = nonreferential utterance.

| X       | P(name|X) | P(zero|X) | P(pronoun|X) | P(other|X) | P(nonref|X) |
|---------|-------|---------|-----------|-----------|-----------|
| name    | .25   | .12     | .06       | .02       | .55       |
| zero    | .04   | .28     | .05       | .01       | .62       |
| pronoun | .10   | .15     | .19       | .03       | .53       |
| other   | .05   | .06     | .11       | .08       | .70       |
null
sonal pronouns and zero anaphora are used only when the referent is in focus, but demonstratives can also be used when it is out of focus but in working memory. By Grice’s maxims, this leads to demonstrative pronouns being used more frequently when there is a topic shift, or when the referent has been introduced non-linguistically (Gundel et al., 1993; Gundel, Hedberg, & Zacharski, 2004). In our data, Japanese pronouns were more likely to be used in reference to an object the first time it was referred to than at any other point in the discourse. This suggests that in Japanese child directed speech pronoun use may signal topic change (e.g. “look at this!”), rather than topic continuity.

As suggested originally by Fernald and Morikawa (1993), pure word-object mapping in Japanese might be a very hard problem: Only about a third of references name the object, while the other two-thirds make use of pronouns and zero anaphora. Moreover, references to objects in Japanese (as in English) are not evenly distributed in discourse. But if children have some sense of what the current topic of discourse is—what is given, and what is new—this problem might be somewhat alleviated. A Japanese-learning infant could infer the topic of conversation and then assume that future comments, whether using names or anaphora, referred to that topic. Thus, this study underscores the importance of topic continuity in early word learning, suggesting that tracking the topic of conversation across utterances may be even more crucial to word learning success in pro-drop languages.

Acknowledgments

Special thanks to Akiko Knott for her invaluable annotation work, and to Stephan Meylan for advice on using the mechanical turk interface to rekey our transcripts.

References


