Data on English article usage, based on a new classification system ("a," "the," and null article), were obtained from 20 adults and 20 two- and three-year-old children. An oral sentence completion technique was used with the child subjects, and the same items in written form were used with the adults. The results for the older children in the sample confirmed the results of previous studies in revealing a pattern of overuse of the definite article. However, the more complete usage data suggest that the overuse is a selective one that occurs predominately in one category and after a period of essentially correct usage. These findings argue against an explanation based on egocentrism and suggest that the incorrect usage of the more advanced children results from an overextension of a principle of shared knowledge found in adult article use. Overall, the data in this study allow description of an acquisition sequence for the English article system that extends earlier developmental findings and resolves some of their inconsistencies. (Author/GT)
The research reported herein was supported by the Department of Health, Education and Welfare, PHS Child Training Grant No. HD-002-44 to the first author and by the National Institute of Education under Contract No. US-NIE-C-400-76-011b. The authors wish to thank Ellen Brewer, Carolyn Mervis, M. Michael Akiyama, Anne Hay, and Linda Hunter for suggestions and comments on earlier versions of this paper. Thanks are also due to the teachers and children at the day care centers visited: The Learning Tree, Toddler's Campus, and Kiddie Kountry.
A new classification of English article (a, the, null) usage was developed. On the basis of this classification scheme, data on article usage were obtained from both adults and children. The sentence completion technique used in this experiment allowed the examination of children in the initial period of article acquisition (2-3 years). The results for older children in the sample confirmed previous studies in finding a pattern of overuse of the definite article. However, the more complete usage data and the age range used in this study suggest that the overuse is a selective one, which occurs predominately in one category, and after a period of essentially correct usage. These findings argue against an explanation based on egocentrism and suggest that the incorrect usage of the more advanced children results from an over-extension of a principle of shared knowledge found in adult article use. Overall, the data in this study allow description of an acquisition sequence for the English article system that both extends and resolves inconsistencies in earlier developmental findings.
Article Acquisition

Acquisition of the Article System in English

The article system has been a subject of inquiry for philosophers (Christophersen, 1939; Hewson, 1972; Kramsky, 1972; Russell, 1905) and linguists (Jespersen, 1933/1966; Perlmutter, 1970; Moravscik, Note 1), as well as for psychologists; it probably owes its wide appeal to the fact that the articles are important in a wide variety of discourse processes and in the interactions of linguistic and nonlinguistic knowledge. In psychology, there have been several recent studies concerned with the acquisition of the articles (Bresson, 1974; Brown, 1973; Maratsos, 1974, 1976; Warden, 1976), but these studies have not dealt with the full range of article usages, nor have they focused on the age period (2-3 years) that appears to be most crucial for article acquisition. Most of the fundamental questions about how children learn to use articles remain to be answered.

Previous Acquisition Studies

One of the earliest records of article use in children is that of Leopold (1949), who kept a diary of his daughter's bilingual speech that included all utterances and the history of their use. A and the appeared at 2-2 and 2-4 in Hildegard's speech, but their use was noted as being "still rare" at 2-6. By 2-9, she was using the articles regularly; however, since the brief contexts given leave the type of usage unclear, the accuracy of her article use cannot be determined. Brown (1973) also noted the acquisition of a/the in observations of three children. He was unable to distinguish the particular usage contexts, and thus could not give
Article Acquisition

separate accounts of the acquisition of each article. However, eliminating the doubtful cases, *a* and *the* were acquired at about the same time: 3-3, 3-5, and 3-0 for the three children (using a criterion of 90% correct usage). Brown concluded that *a* and *the* must be acquired as a system.

Since these earlier studies provided only general indications of early article use, Maratsos (1974, 1976) investigated the appropriateness of *a* and *the* usage in an experiment involving a comprehension task and two production tasks. In the comprehension task, 3- and 4-year-old subjects were asked to act out parts of stories in which the contrastive use of *a* and *the* as markers of nonspecificity and specificity was the main variable. He found that the children in both age groups responded differentially to the use of *a* and *the*. Article production was tested in a story-telling task and in a set of "game" tasks. Some of the older subjects gave adult-like responses on these tasks, but most of the younger children showed a pattern of errors resulting from overuse of the definite. Maratsos attributed this overuse of the definite to the children's egocentric (Piaget, 1926) point of view: i.e., they apparently failed to take into account the hearer's lack of knowledge about something already known to themselves.

Warden (1976) also carried out an experimental study of the articles, pointing out that the definite/indefinite contrast (specific-nonspecific) is only one function of the article system. He emphasized the function of the articles as "referring expressions" which enable a speaker to introduce and/or comment upon an item in a discourse. According to Warden, the context of the discourse is particularly important, since it determines whether the referent needs introduction and explanation (for the hearer's
benefit), or whether it is something that is already common knowledge for the speaker and the hearer. Warden's investigation was more concerned with the various usages of the indefinite article than with those of the definite article. He found that both adults and children used the indefinite correctly for naming objects. However, the children, unlike the adults, frequently used the definite rather than the indefinite article for introduction of a new referent. Warden attributed this incorrect use of the definite to the children's egocentric viewpoint.

A study by Bresson (1974) of article use found that French children exhibited the same pattern of frequent inappropriate use of the definite, thus lending additional support to Maratsos' and Warden's findings.

Overall, these studies of article acquisition suggest that young children demonstrate basic control of the distinction between a and the, but frequently use the definite article inappropriately in contexts where the indefinite should appear. This overuse of the has been attributed to the egocentrism of the child's viewpoint, which prevents consideration of the hearer's perspective. There are, however, important limitations in these studies. None of the studies examined the full range of article usages, and none employed procedures suited to obtaining reliable data from very young subjects. The present study is an attempt to resolve these difficulties by investigating the article system as a whole with a procedure that allows study of children as young as 2 years of age.
A FRAMEWORK FOR THE ARTICLE SYSTEM

Usage Categories

A framework for the article system will be developed by examining the function of the three article forms (a, the, and null) in three basic usage categories (Introduction/Anaphoric Reference, Context Frame, and Generic). First we will describe the three usage categories:

Introduction/Anaphoric Reference

In Introduction/Anaphoric Reference, the articles are used either to introduce a new topic into the discourse or to make reference back to a previously introduced referent (anaphora). Usually anaphoric reference is to a linguistically introduced item (e.g., "There was once a cow. The cow..."). However, it may also occur nonlinguistically, when the speaker has introduced a particular referent by pointing, gesturing, etc. (e.g., the speaker nods in the direction of another person present, saying: "The idiot just bought a huge new gas-guzzler.").

Context Frame

In the Context Frame usages, article selection is based on knowledge of typical objects and events ("context frame" knowledge), without previous specific linguistic or nonlinguistic introduction of the referent. This usage is related to Halliday and Hasan's (1976) discussion of "exophoric" definite reference or reference that is determined by predictability within the situation: for example, "The train is late," when both speaker and hearer are waiting for the same train; or when there is only one possible referent available, "the sun." The articles within this usage type
indicate what is or is not in the "consciousness" (Chafe, 1972, 1974, 1976) or in the "discourse registries" of both speaker and hearer (Kuno, 1972). Knowns and unknowns within a discourse depend not only on what is immediately apparent to the speaker and hearer, but also on their shared world-knowledge and experience that goes beyond the immediately present discourse situation. This knowledge takes the form of inferences which allow definite reference when no antecedent (linguistic or nonlinguistic) is present, as in: "John found a shop manual for his Fic, but the page specifying the dwell angle was missing" (Nash-Webber, in press). In this type of usage, the is used for knowns (or highly predictable elements) and a is used for unknowns (or less predictable elements) within the context frame involved.

Generic

In Generic usage the articles indicate reference to universal knowledge, knowledge of conceptual classes, and membership in these classes.

The Article Framework

The framework description includes the article forms a, the, and null. The article system has usually been discussed as composed of a and the only. We include here a third article form, null (noun without article), since the use of nouns without articles in English operates on principles similar to those governing a and the, is contrastive with both a and the for mass/count distinctions, and is involved in usages that parallel a/the usages. In this section an outline of the English article system is presented, along with examples of each category of use. The framework is organized by the three usage types: Introduction/Anaphoric Reference, Context Frame, Generic.
Within each usage division the various categories of use are discussed by article form: a, the, null. (Article use in geographical terms is a complex case and will not be covered.)

Introduction/Anaphoric Reference

1. The definite article marks Anaphoric Reference, the mention of an already introduced item: "I saw a man on the street. The man had a purple hat."

2. The indefinite article is Introductory either (a) as the first mention of a particular referent for later comment: "I saw a man on the street. The man . . ." or (b) nominatively, when an already focused referent is "introduced" as a member of a class: "That is a fountain pen."

3. The null article is Introductory when the speaker wants to introduce a particular group of like-referents in order to refer to that same group further: "There are horses running on that field. Those horses . . ."

Context Frame

1. The definite article marks (a) a Context-Unique referent which is either a Simple Context-Unique (only one possible referent): "The car is okay except for the steering wheel"; or a Determinative-Unique (several referents available, but modification makes the choice specific): "The woman with a blue hat on is leaving soon." (b) The can also mark a Context-Intermediate reference (one of only a few available like items within a familiar context frame): "The little girl ran to the car and opened the door." In these cases a traditional grammatical rule operating on a specificity/nonspecificity principle would give a; yet the is acceptable, and occurs frequently.
2. The indefinite article includes two categories of use within this usage type: (a) Context-Intermediate (as above): "The little girl ran to the car and opened a door" and (b) Context-Nonspecific (many like items are available, and an unspecified one of these is indicated): "The boy opened his bag of blocks and took out a block."

3. The null article seems to operate within a more general frame of the (relevant) world known to the individual. There are two categories of use: (a) In individualization, null occurs when the referent is specific and unique in and of itself and consequently needs no further limitation: "Mary is coming over to visit today." (b) The Abstraction use of null indicates concepts without boundaries or instances where the boundaries are vaguely defined. Mass/abstract nouns and plural count nouns (when no number is specified) fall within this category when they refer to less than the generic sense: "She drank milk for lunch"; "Their closet is filled with teacups." (This plural null use could be described as a plural nonspecific.)

Generic

1. The definite article appears to be used for generic statements when the underlying concept of a category is intended: "The dog was the first animal to be domesticated."

2. The indefinite article in this usage indicates a member of a class as a typical exemplar of that class: "A mouse eats cheese."

3. The null article as a Generic refers to universal knowns by indicating the class as a whole, including in its scope all possible exemplars: "Lions are noble creatures," "Food gives us energy."
A DEVELOPMENTAL STUDY OF THE ARTICLES

Focus of the Study

The article framework just presented guided our selection of article usage. Representative usage categories were selected from the full range of article uses to provide a more adequate description of children's early article system acquisition.

In addition to the greater scope of usages considered, the children studied were drawn from a younger age range than in previous experimental studies. Leopold's (1949) diary data indicated that a preceded the in acquisition; the more recent studies tested older children and found that at three years a basic control of the articles was evident, although the was often used inappropriately. The period from age 2 to 3 has not previously been studied experimentally, yet this is probably the period in which the acquisition of the article system is most actively taking place. The present study examined article use in children from 2-4 years to 3-5 years of age, to discover the steps by which children acquire competence in using the articles.

The experimental tasks used in the previous experimental studies (Maratsos, 1974, 1976; Warden 1976) were difficult for 3-year-old children, and so would be even less appropriate for younger children. A possible solution would be to use data from recordings of naturally occurring speech. However, using transcripts of this type limits the researcher to what the child happens to produce. The range of uses and the number of utterances per usage type may then be too small to be of use. The ideal procedure would be a task that is naturalistic for the child, yet allows the
experimental control needed to ensure a broad range of data. The task used in this study was developed in an attempt to fulfill these requirements. Taking advantage of the fact that children while playing often spontaneously narrate what their toys "are doing," the experimenter used this type of narration as the base for presentation of the test items within a free-play session. In this way the task was natural for the child, yet provided an adequate range of data.

Warden tested adults as well as children on all his tasks and noted that even in adult usage the discourse context may interact with the references made to produce "incorrect" article usage. Obviously one needs a clear picture of adult usage patterns before drawing conclusions concerning correct or incorrect usage in children. Therefore, in this study adult subjects were tested on the same items that were presented to the children, and the children's data was interpreted in light of the adult usage.

In summary, the purposes of the study were: (a) to observe how articles are acquired from their first use; (b) to observe how, in the course of acquisition, children differ from adults in their usage; and (c) to attempt to provide a theoretical account of these differences if found.

The study was designed to accomplish these purposes by: (a) examination of the full range of article usages; (b) inclusion of younger subjects; (c) use of a new, more naturalistic task; and (d) comparison of the children's data with empirically determined adult article use.
Method

Subjects

Twenty children from day care centers in the Champaign-Urbana area were used as subjects. The younger age group included 10 children (3 male, 7 female) with an age range of 2-4 to 2-11, and a mean age of 2-8. The older age group also included 10 children (2 male, 8 female) with an age range of 3-0 to 3-5, and a mean age of 3-2. Mean Length of Utterance (MLU) scores based on spontaneous speech data (Brown, 1973) were computed for each subject. In the 2-year-old group, the MLU range was 2.93 to 4.77, with a mean of 4.04; for the 3-year-old group, the MLU range was 4.12 to 5.73, with a mean of 4.72.

The adult sample consisted of 20 college freshmen and sophomores who participated in the experiment as part of the course requirements for Introductory Psychology.

Materials

The experimental items for the study were developed from the article framework presented above, using six categories representative of the range of article usages. Each item was constructed by using one or more sentences to create an appropriate context for the usage to be tested. Each ended with a slot for the response (article and referent) and was presented as a sentence completion item. For example, one of the Anaphoric items was: "This little boy was swinging on a swing . . . And he was swinging and swinging . . . (Experimenter swings doll in air) . . . And then he got off of (the swing)." The items were written on 3 x 5 index cards which were shuffled in order to randomize the order of item presentation for
Article Acquisition

12

each subject. The cards were used with a board game and several small toys in a play session.

There were 55 items. The specific categories and the distribution of items per category were as follows: Anaphoric, 8; (Simple) Context-Unique, 10; Context-Intermediate, 6; Introductory, 7; Context-Nonspecific, 6; Generic, 18. (Since Generic statements can occur in a number of different forms, it was necessary to include more items in this category.) All items (except some Generic items) were constructed to elicit singular nouns. No items were constructed for the Individualization and Abstraction categories of use. In order to study null use, each child's responses were classified as mass/abstract vs. singular count nouns vs. plural count nouns, and a comparison of article use before each class was made. A transcript of each child's spontaneous speech was used to provide some indication of Individualization usage.

Procedure

The researcher testing the children made initial visits to each of the day care centers to become acquainted with the children that would be included in the study. On later visits, each subject went to a separate room with the researcher for the experimental session.

Sentence-completion responses were elicited from the children within the context of spontaneous play narration. At the beginning of the session, the child was shown a board game and several toys. Play with the game soon evolved into a play session with the toys in general. During this play, the researcher presented the items by working each into the play, narrating the relevant scene as she acted it out with a toy character. The sentence-
completion response was elicited by the researcher's use of a raised intonation and pause; this was a sufficient cue for the child to complete the sentence. The method was successful with even the youngest subjects; the children enjoyed the play sessions, and enthusiastically completed the sentence items. They frequently narrated on their own or continued the researcher's story lines beyond the "required" response.

Each subject was given all 55 items in two or three sessions of approximately 20-25 min each. At the end of the last session, the child's spontaneous speech was recorded to provide data for calculating the MLU score. All sessions were tape recorded with a Superscope cassette recorder and transcribed by the experimenter the same day, or as soon as possible thereafter. After the transcriptions had been completed, the tapes were reviewed by an assistant experienced in distinguishing phonological forms in children's speech. There was 93% agreement between this second transcription of the article forms and the original transcription.

The adult subjects were given the same sentence-completion items in booklet form and asked to respond to these as if they were speaking them in a natural context.

Results

Scoring

Responses were scored as follows: (a) A response was classified as either "appropriate" or "other." A response to an item was "appropriate" if it fulfilled the sense of the category use being tested by that item. For example, given the Context-Nonspecific item: "She reached into her bag of blocks and took out ____," "block" would be an appropriate referent,
but "bag of blocks" would not be and would be classified as "other."

(b) Each "appropriate" response was also classified as a, the, or null.

(c) Each "other" response was classified as either: inappropriate for the category use being tested, as an instance of no-response, or as a "wild" response (completely irrelevant). All mass/abstract responses to the items were classified as "other."

Correct article responses for evaluating the children's data were obtained by examining the adult responses for each item: Any response that occurred in 15% or more of the total adult responses to an item was considered to be a correct response for that item. This criterion level was chosen to exclude wild or irrelevant responses while including acceptable minority responses. The average of the individual item response scores in each category of use produced the following pattern of adult article usage: Anaphoric, 94% the; Context-Unique, 97% the; Context-Intermediate, 92% the; Introductory, 100% a; Context-Nonspecific, 92% a. These responses were all as predicted, except that in the Context-Intermediate category a had been expected as well as the. In the Generic category, 90% of the responses were correct, and these were distributed across items as follows: a only, 4 items; null only, 6 items; two or more possible responses, 8 items; (no the-only items were found). These data provided the criterion of correct performance for scoring the children's data.

For the analysis of the children's data, the subjects were classified into four developmental groups. Since age is not a reliable index of linguistic ability (Brown, 1973), this variable alone would not be an appropriate basis for the classification. However, a measure based on linguistic
ability alone, such as Mean Length of Utterance (MLU), does not reflect differences in conceptual maturity that would be likely to affect accuracy in article usage. Examination of the data suggested that a composite index based on both MLU and age should be used: A plot of the individual subjects' appropriate use by age showed several children with marked deviations from the expected pattern of gradually increasing usage. These children were characterized by high or low MLU's for their age range. When the use was plotted by MLU, another group of subjects showing deviating scores was found; this group of children tended to be particularly young or old for their MLU range. Thus, an index was developed that took both age and MLU into account in equal proportions. This index of 'maturity' was calculated for each subject by averaging the age in years and (to adjust for range differences) half of the MLU. The resulting 20 index scores were divided into four equal intervals to give four subject groups with the following range of scores in each (from least mature to most mature): Group 1, 1.9 - 2.2; Group 2, 2.3 - 2.5; Group 3, 2.6 - 2.8; Group 4, 2.9 - 3.1. The distribution of subjects in these groups was: Group 1, four subjects; Group 2, eight subjects; Group 3, five subjects; Group 4, three subjects. The analysis was based on these four subject groups.

Article usage scores were calculated as the percentage of each subject's use of each article form (a, the, null) for each category (Anaphoric, Context-Unique, Context-Intermediate, Introductory, Context-Nonspecific, and Generic), yielding 18 scores for each subject.
Analysis Based on Percent Correct Responses

An ANOVA with Group and Category as factors was carried out using percent correct responses for each subject in each category (see Table 1). The main effect for Group was significant, $F(3,16) = 9.73, p < .001$; with the developmentally mature subjects performing better than the less mature. The main effect of Category was also significant, $F(5,80) = 18.93, p < .001$. There was an interaction of Group and Category, $F(15,80) = 1.86, p < .05$. The a-categories (Introductory and Context-Nonspecific) for Groups 3 and 4 seemed to be exceptions to the general increase in performance with developmental maturity. The apparent decline in performance from Group 1 to Group 4 on the Context-Nonspecific category was nonsignificant (Mann-Whitney U). The interaction effect may be due to the overall pattern of general increase in the-usage occurring in conjunction with an apparent decline in a-usage.

A follow-up comparison (Tukey Test [b], Winer, 1962) of the category means showed that the Introductory category was significantly ($p < .05$) higher in performance than any of the other categories. Performance on the Context-Nonspecific category was significantly better than that for the Unique and Generic categories. The ordering of the various categories from easiest to most difficult was: Introductory, Context-Nonspecific, Anaphoric, Context-Intermediate, Unique, Generic. In general, a-categories were easier than the-categories.

Overall, these data show a preceding the in acquisition and an overall improvement in article use with increasing developmental maturity. For the
more mature subjects, there was a consistent but nonsignificant tendency toward decreased accuracy in the Context-Nonspecific category and a suggestion of a drop in performance in the Introductory category.

Analysis Based on Appropriate Responses

The scores in the preceding analysis were based on the total percent correct and so were influenced by any type of error made by the children. In order to understand how well children were able to use the articles once they had made an appropriate response to an item, percentage scores based on appropriate responses only were calculated. These scores, which reduce the variability due to irrelevant responses, have been used in the following analyses.

For each Category and Group, the mean percent correct of appropriate responses is given in Table 2. As would be expected, the scores improve with this index of performance. The pattern of means is generally similar to that for percent correct total responses. Perhaps the most striking difference is the increase in a-category performance for the least mature subjects. This early accuracy with a usage in the appropriate categories cannot be attributed to an undifferentiated use of a wherever an article is required. Production of a in the appropriate categories (Introductory and Context-Nonspecific) is consistently higher than in the the-categories, with Groups 1 and 2 producing 96% and 88% a, respectively, for combined a-category usage, but producing 58% and 60% a responses for the-category instances. The more mature subjects show overall improvement in performance, but the previously noted decrease in accuracy for Group 4 subjects on Context-
Nonspecific category items is still evident, especially when compared with Group 1 performance.

An ANOVA by Category and by Subject Group found significant effects for Subject Group, $F(3,16) = 7.69, p < .01$; for Category, $F(5,80) = 17.25, p < .001$; and for the interaction of Group by Category, $F(15,80) = 3.90, p < .001$. Examination of the means for the groups suggested that the interaction could be attributed to the better performance of the less mature subjects over the more mature subjects in the a-categories, contrasted with the opposite pattern for the the-categories.

These data showed a decrease in performance (particularly for Group 4 in the Context-Nonspecific category), so Groups 1 and 4 were compared for accuracy of response in this category. The difference between these two groups approached significance, Mann-Whitney $U = 1, N = 3, 4, p = .057$. The two groups were each compared with the adult sample. Group 1 did not differ from the adults, $U = 36, N = 4, 20, NS$; Group 4, however, was significantly different from the adult sample, $U = 5, N = 3, 20, p < .01$. This comparison gives a rather striking finding—the least mature group of children was more adult-like in use of a in the Context-Nonspecific category than was the most mature group of children. The Introductory category showed a small (nonsignificant) decrease for Groups 3 and 4. Thus, the decrease in a-performance is apparently a selective one, occurring in the Context-Nonspecific category, and is not simply an overall drop in accuracy.
Tukey Test (b) (Winer, 1962) comparisons (p < .05) were performed to determine which categories were significantly different from others. No significant difference was found between the Introductory and Context-Nonspecific categories. Each of these categories, however, showed significantly better performance than each of the remaining four categories. The categories ranged from least difficult to most difficult as follows: Introductory, Context-Nonspecific, Context-Intermediate, Generic, Context-Unique, Anaphoric. As in the previous analysis, the performance on a-categories is better than on the the-categories.

Error Data

Table 3 gives the mean percentage of incorrect responses (within appropriate responses) for each group for each category (omitting the Generic category). The decrease in performance for the Context-Nonspecific category is reflected in the mean percentage of incorrect the responses for Context-Nonspecific items. The number of errors by Group 4 children (38%) in this category is more than twice the number of the responses for any of the other developmental groups. The was rarely given as an incorrect response for Introductory items by any group. The other article alternative, null, is rarely given as an error in the Introductory and Context-Nonspecific categories. Thus, it appears the decrease in correct performance in the Context-Nonspecific category is due to an intrusion of the responses. Note that this decrease in accuracy occurs in the children who have passed beyond 50% accuracy in the the-categories. The mean percentage of incorrect

Insert Table 3 about here.
null responses is significantly higher for the the-categories than for the a-categories, Wilcoxon matched-pairs signed-rank test, \( T = 48.5, p < .05 \), suggesting some systematic use of null in the the-categories. (Generics were excluded in this comparison since the-response items were few and the only appeared as an alternative response.)

In the Generic category, a complex category for which a variety of different forms is possible, a number of response errors reflected a bias toward singular statements; for example, "a/the tree" or "null tree," instead of "the trees" or "null trees." This bias was especially evident in Group 1 subjects; all of their incorrect Generic responses and all of their correct responses were singular expressions. For Groups 2, 3, and 4, singular responses accounted for 67%, 77%, and 65%, respectively, of the appropriate (incorrect and correct) responses made. This singular bias together with the early advantage of a-Generics suggests that the children may be using the Generic items as descriptions of typical events rather than as true Generic statements. For instance, in responding to the item: "Birds build their nests in _____," the preferred response given by the adults was "trees," but the preferred response given by the children was "a tree." This use of the singular form might well occur if the children were recalling and describing the familiar event of seeing a bird's nest in a tree/the tree near their home. In other words, they may be referring to a typical experiential event, rather than making a general statement about birds' habits in nest-building. If children are able to use a general sense, the singular bias might reflect the fact that they are better able to make general statements by speaking on the level of one typical object rather than on the level
of "the set of objects." The explanation of the singular bias as based on descriptions of typical events seems the more plausible explanation when the types of "other" responses for this category are examined. The more mature subjects performed better than the less mature subjects on the Generic items in that they often used the appropriate referent, but their scores remained low due to incorrect number or article use (singulars and definites predominated). For Groups 1, 2, and Group 3 to a lesser extent, many "I don't know" responses and irrelevant responses occurred, and it seemed that the basic competence underlying Generic responses was lacking in these children. Response difficulty with Generic items appeared to vary with the form of the particular item and with the type of referent presented in the item. This complex category obviously deserves additional experimental study.

Null Usage

General data. Since the only items constructed to elicit null article use were the null Generic items, most of the information on null usage was tabulated from responses to the other types of items. Article use before singular count nouns, mass/abstract nouns, and plural count nouns was tallied for both correct and incorrect responses in a representative portion of the response data. Table 4 gives the mean percentage of article use for each noun type (singular, plural, mass/abstract). Clearly, a distinction was made by all subjects: For singular nouns the mean percentage of a/the responses ranged from 77% for Group 1 up to 90% for Group 4. For mass/abstract nouns, the distribution of article use reversed; the mean percentages of null responses were 100%, 88%, 91%, and 83% for Groups 1 through 4,
respectively. No plural count nouns were present in the data for Group 1 (perhaps these children avoid plural count nouns because they would complicate a simple \(\text{mass} = \text{null} \ vs. \ \text{count} = \text{a/the}\) distinction), but for the other groups the majority of the responses were properly null (81%, 100%, and 90%). In order to check the reliability of the Group 1 subjects' omission of plural nouns, the remainder of their responses were examined for plural use. Six instances of plurals were found (out of 102 responses): "a cards," "a candies," "in a rocks," "kitchen doors," "clouds and stars," "stars" (also three plurals modified by numerals thus needing and receiving no articles).

Insert Table 4 about here.

Overall, it appears that Group 2, 3, and 4 children used null correctly to distinguish mass/abstract nouns and plural count nouns from singular count nouns. Group 1 children also used null correctly to distinguish mass/abstract nouns from singular count nouns, but showed some confusion about the use of articles before plurals. They may be avoiding the use of plural count nouns in order to reduce this confusion. From the few occurrences of individualization usages of null in the spontaneous speech data, it appears that children in all groups used null correctly before names of persons and places (individualization usage).

Generic item data. The null-Generic items were the only constructed items involving null article use, and the correct responses for these items according to the adult model were most frequently null + plural count noun. When children did respond with the plural referent to a null-Generic item,
they almost always correctly used null. This probably indicates mastery of article usage with plural count nouns rather than a mastery of Generic null usage.

"Other" Responses

A number of the responses classified as "other" consisted of an appropriate referent noun preceded by a modifier such as a demonstrative, possessive, or numeral. These forms, which do not require an article, were used more extensively by the less developmentally mature children than by the more mature children. The percentages of these forms (compared to all "other" responses) were: Group 1, 42%; Group 2, 25%; Group 3, 33%; Group 4, 18%; Adults; 3%. This pattern suggests that one strategy of young children who have not completely mastered the article system is to use modifiers that make articles unnecessary.

Summary

The preceding analyses suggest the following acquisition sequence:

1. Initial use of a and null only. In the least developmentally mature subjects, a was the predominant article form; null was also used, but less frequently. Group 1 children rarely produced any the's. Several children were found who used only a or predominately a with some few the's, but the reverse of this pattern was not found. The initial use of a was not an undifferentiated use across all categories since it was consistently used more often in the appropriate categories. Null was correctly used before mass/abstract nouns, but there was some confusion in use before plurals, possibly leading to the avoidance of plurals by children at this level (Group 1).
2. Beginning use of the. In Groups 2 and 3, a-category usage remained high in accuracy (although some decrease was noted); at the same time the responses appeared in all of the appropriate categories (Anaphoric, Context-Intermediate, Context-Unique). Null was consistently used before plurals, as well as before mass/abstract nouns.

3. Overuse of the. Once the usage was firmly established, incorrect use of the definite article began to occur selectively in the Context-Nonspecific category, which requires a. Thus, the most mature subjects (Group 4), compared to the other, less mature subjects, showed better performance on the-category items but were poorer on a-category items.

A longitudinal study would be required to confirm this interpretation of our cross-sectional data. However, the consistency and strength of these findings make it seem very likely that an individual child learning to use the articles would follow the sequence of acquisition outlined above.

Other findings were: (a) Incorrect use of null occurs significantly more often in the-categories than in a-categories, suggesting that null may be associated, with a definite sense for young children. (b) Examination of responses classified as "other" showed that a subset of these responses containing demonstratives, possessives, and numerals (precluding article use) were used far more frequently by the children than by the adults. Children may be strategically avoiding article use where an alternative is readily available. (c) The predominant article in the adult responses for each category fit the predictions of the article framework for all but the Context-Intermediate category, where although both a and the were expected, the was strongly preferred.
Discussion

Comparison with Previous Studies

To the extent that one assumes that first-acquired forms are "easier" than later-acquired forms, the data provided by Leopold (1949) are inconsistent with data reported by Bresson (1974), Maratsos (1974, 1976), and Warden (1976). Leopold found a appearing before the at a very young age, yet in the three experimental studies of older children the was used more frequently and more accurately in the appropriate categories. The data from the present study suggest a resolution for this inconsistency. Both the article usage observed by Leopold and the usage reported by the experimental studies are accounted for by the sequence of article acquisition described above.

Leopold's data reflect the initial phase of article use where a is shown to be prior to the, a sequence which might be attributed to the salience of new information or perhaps to phonological differences in the two forms. The data from the experimental studies are consistent with our third phase of acquisition. The implication of these findings is that the observed overuse of the should not be interpreted as a lack of sophistication occurring during the earliest period of acquisition, since it follows a period where there is not any significant incorrect the usage. The reported weakness of performance with a in the previous studies must also be reinterpreted. Rather than attributing the errors to a lack of ability that marks a period of initial a usage, these errors should be viewed as a reversal of earlier correct performance.

The sequence of article acquisition developed in the present study has replicated the findings of the previous observational and experimental studies.
and has resolved the apparent inconsistencies in these studies through the description of an intervening period of acquisition. By this linkage of the diary and experimental studies as part of a developmental sequence, a more complex but lawful progression in learning to use the articles has been suggested.

Incorrect "the" Usage: Evaluation of the Egocentrism Hypothesis

Several earlier studies (Maratsos, 1976; Warden, 1976) interpreted the child's overuse of the in terms of "egocentrism" (Piaget, 1926). However, a number of arguments can be made against this explanation. First, the overuse of the that leads to poorer performance in the Context-Nonspecific category here has been shown to follow a period of correct a usage. If egocentrism were the cause of the incorrect the use in the a-categories, this would suggest that the prior period exemplified non-egocentric use. Such a view would not reflect a logical course of development, or at least, a very parsimonious view of development.

One might reformulate the egocentric position and hypothesize that egocentrism begins operating only after a principle of the-usage is established. This modified position would escape the difficulty outlined above; however, a second problem would still exist for the egocentrism position—the overuse of the was a selective one, occurring predominately in one usage category. If an egocentric viewpoint is the source of the overuse, its operation should be seen across both a-categories equally and differential intrusion of the in one category (Context-Nonspecific) should not be found.

Finally, research on the young child's social-cognitive skills has shown that the hypothesized egocentrism is not always found. On many tasks
children can and do take their hearer's perspective into account (Krauss & Glucksberg, 1969; Shatz & Gelman, 1973).

Examining the patterns of article use over development, it is apparent that systematic changes must be going on, changes that have to do with the nature of the article system itself. Although one might hypothesize that the incorrect the usage is part of a random overflow of the's occurring simply because the is now a viable response for the child, evidence from the children in Groups 2 and 3 argues against this type of explanation: The subjects in Group 2 began acquiring the, yet generally maintained accuracy for a usage. The subjects in Group 3 showed a strong increase in the usage over Group 2 but the Context-Nonspecific category performance remained the same for the two groups. The decrease in performance that occurred for Group 4 subjects is thus not simply the result of more the's being used in speech.

An Alternative Explanation

The inadequacy of the egocentrism hypothesis for explaining the acquisition data leaves open the question of what is causing the incorrect the usage found so consistently in this and other studies. The selectivity of the incorrect usage in our data and the fact that it is found after correct the usage has become stable leads us to suggest that overgeneralization of some principle guiding the usage underlies the phenomenon. The article framework may provide a suggestion for such a principle.

The adult Context Frame usage pattern of the for Context-Uniques and a for Context-Nonspecifics follows the traditional distinction of the for knowns and a for unknowns. However, the was the predominant response for
adults (as well as for children) in the Context-Intermediate category, although grammatical accounts predict a. In the Context-Intermediate category, what might be called "quasi-knowns" occur, where shared world-knowledge and conversational postulates (Grice, 1975) take precedence over a simple specific/nonspecific contrast. A quasi-known instance is one in which (a) the referent is one of a few like items available; (b) the specification of the item is not particularly relevant for discourse continuity; and (c) the item is an intrinsic, highly predictable, element of the discourse context frame. In sentences such as "John got hit on the leg by a bat," and "Mary got into the car and sat down on the seat," the speaker assumes that the hearer knows that there are only a few alternatives available, and this is sufficient to permit the definite article to be used, both as a "shorthand" for conversation and as a confirmation of speaker/hearer "solidarity" (Hinds, 1976). In fact, the grammatically "correct" marking of specificity/nonspecificity ("a leg," "a seat") produces sentences of marginal acceptability at best. When the set of available items becomes larger, however, as in the Context-Nonspecific category, the use of the definite for one member of a larger set becomes counterproductive.

In the acquisition data, Group 4 children used the definite article for Context-Uniques and Context-Intermediates, as the adults did; but, unlike the adults, the children frequently used the for Context-Nonspecifiers as well. These children may have overgeneralized from the adult usage pattern and formed a rule that anything which is predictable within a typical, often-experienced context frame is "known" and so can be definite. In other words, they use the definite article for uniques and quasi-known
instances; but then also extend it to cases where many possible like items exist. This account also explains why incorrect the's do not occur in the Introductory category. Referents that are introduced with a are items not previously referred to or focused upon that introduce a new aspect or change of focus into the discourse and thus are not predictable or known in any way. One apparent problem with the overgeneralization hypothesis is Warden's (1976) data reporting incorrect the usage within Introductory items. The subjects in his study were given a three-picture sequence and asked to describe it to another subject behind a screen. Incorrect definites occurred in cases such as the following: For an event including a cat and a dog, the child might correctly introduce the cat and then continue with an incorrect introduction of "the dog," as in: "A cat runs up a tree and the dog comes along." However, for young children, introducing "a cat running up a tree" may provide a sufficiently familiar event context frame to allow definite reference to "the dog." Typical context frames for the child may not always match the adult set of typical, commonly known, context frames.

Overall, the data in this study provide a description of the acquisition of the English article system which extends earlier developmental findings and leads to an explanation in terms of an extension of usage principles apparent in adult data, rather than due to an egocentric failure to consider the hearer's perspective.
Reference Note

References


Nash-Webber, B. L. *Syntax beyond the sentence: Anaphora.* In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education.* Hillsdale, N.J.: Erlbaum, in press.


### Table 1

Mean Percentage Correct of Total Responses

<table>
<thead>
<tr>
<th>Article usage categories</th>
<th>1(^a)</th>
<th>2(^b)</th>
<th>3(^c)</th>
<th>4(^d)</th>
<th>Adult(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The-categories:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A-categories:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A, the, null:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean across categories: 15\(^a\) 38\(^b\) 48\(^c\) 60\(^d\) 80\(^e\)

**Note.** Numbers in parentheses indicate the number of items in each category; numbers in brackets indicate the number of responses in each subject group by category cell.

\(^a\) = 4 subjects
\(^b\) = 8 subjects
\(^c\) = 5 subjects
\(^d\) = 3 subjects
\(^e\) = 20 subjects
Table 2
Mean Percentage Correct of Appropriate Responses

<table>
<thead>
<tr>
<th>Article usage categories</th>
<th>Subject groups by composite index</th>
<th>Adult</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>The-categories:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-categories:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, the, null:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean across categories:</td>
<td>36</td>
<td>58</td>
</tr>
</tbody>
</table>

Note. Numbers in parentheses indicate the number of items in each category; numbers in brackets indicate the number of responses in each subject group by category cell.

<sup>a</sup><sub>n</sub> = 4 subjects
<sup>b</sup><sub>n</sub> = 8 subjects
<sup>c</sup><sub>n</sub> = 5 subjects
<sup>d</sup><sub>n</sub> = 3 subjects
<sup>e</sup><sub>n</sub> = 20 subjects
Table 3

Mean Incorrect Response Percentages by Developmental Maturity Group and by Category

<table>
<thead>
<tr>
<th>Article usage categories</th>
<th>Subject groups by composite index</th>
<th>1a</th>
<th>2b</th>
<th>3c</th>
<th>4d</th>
</tr>
</thead>
<tbody>
<tr>
<td>The-categories:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content-Intermediate (6)</td>
<td>a null</td>
<td>71</td>
<td>14 [14]</td>
<td>54</td>
<td>0 [49]</td>
</tr>
<tr>
<td>A-categories:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introductory (7)</td>
<td>the null</td>
<td>0</td>
<td>0 [16]</td>
<td>0</td>
<td>0 [45]</td>
</tr>
</tbody>
</table>

Note. Numbers in parentheses indicate the number of items per category; numbers in brackets indicate the number of responses on which the percentages (for each category by subject group cell) are based.

\[ a_n = 4 \text{ subjects} \]
\[ b_n = 8 \text{ subjects} \]
\[ c_n = 5 \text{ subjects} \]
\[ d_n = 3 \text{ subjects} \]
<table>
<thead>
<tr>
<th>Subject groups by composite index</th>
<th>Singular noun (412)</th>
<th>Plural noun (303)</th>
<th>Mass/Abstract noun (216)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a/the</td>
<td>null</td>
<td>a/the</td>
</tr>
<tr>
<td>Group 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>77</td>
<td>23</td>
<td>[26]</td>
</tr>
<tr>
<td>Group 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>89</td>
<td>11</td>
<td>[97]</td>
</tr>
<tr>
<td>Group 3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>78</td>
<td>22</td>
<td>[55]</td>
</tr>
<tr>
<td>Group 4&lt;sup&gt;d&lt;/sup&gt;</td>
<td>90</td>
<td>10</td>
<td>[30]</td>
</tr>
<tr>
<td>Adults&lt;sup&gt;e&lt;/sup&gt;</td>
<td>92</td>
<td>8</td>
<td>[194]</td>
</tr>
</tbody>
</table>

Note. Numbers in parentheses indicate the total number of occurrences of each noun type; numbers in brackets indicate the number of noun occurrences for each subject group by noun type cell.

<sup>a</sup>n = 4 subjects; total of 32 noun occurrences.
<sup>b</sup>n = 8 subjects; total of 170 noun occurrences.
<sup>c</sup>n = 5 subjects; total of 108 noun occurrences.
<sup>d</sup>n = 3 subjects; total of 74 noun occurrences.
<sup>e</sup>n = 20 subjects; total of 528 noun occurrences.
CENTER FOR THE STUDY OF READING

READING EDUCATION REPORTS

No. 1: Durkin, D. *Comprehension Instruction—Where are You?*, October 1977. (ERIC Document Reproduction Service No. ED 146 566, 14p., PC-$1.82, MF-$0.83)


No. 4: Jenkins, J. R., & Pany, D. *Teaching Reading Comprehension in the Middle Grades*, January 1978. (ERIC Document Reproduction Service No. ED 151 756, 36p., PC-$3.32, MF-$0.83)

No. 5: Bruce, B. *What Makes a Good Story?*, June 1978 (ERIC Document Reproduction Service No. ED 158 222, 16p., PC-$1.82, MF-$0.83)

No. 6: Anderson, T. H. *Another Look at the Self-Questioning Study Technique*, September 1978. (ERIC Document Reproduction Service No. ED 163 441, 19p., PC-$1.82, MF-$0.83)


No. 8: Collins, A., & Haviland, S. E. *Children's Reading Problems*, June 1979. (ERIC Document Reproduction Service No. ED 172 188, 19p., PC-$1.82, MF-$0.83)

No. 9: Schallert, D. L., & Kleiman, G. M. *Some Reasons Why Teachers are Easier to Understand than Textbooks*, June 1979. (ERIC Document Reproduction Service No. ED 172 189, 17p., PC-$1.82, MF-$0.83)

No. 10: Baker, L. *Do I Understand or Do I not Understand: That is the Question*, July 1979. (ERIC Document Reproduction Service No. ED 174 948, 27p., PC-$3.32, MF-$0.83)

No. 11: Anderson, R. C., & Freebody, P. *Vocabulary Knowledge and Reading*, August 1979. (ERIC Document Reproduction Service No. ED 177 470, 52p., PC-$4.82, MF-$0.83)


No. 13: Adams, M., & Bruce, B. *Background Knowledge and Reading Comprehension*, January 1980.


CENTER FOR THE STUDY OF READING

TECHNICAL REPORTS

No. 1: Halff, H. M. Graphical Evaluation of Hierarchical Clustering Schemes, October 1975. (ERIC Document Reproduction Service No. ED 134 926, 11p., PC-$1.82, MF-$0.83)


No. 3: Goetz, E. T. Sentences in Lists and in Connected Discourse, November 1975. (ERIC Document Reproduction Service No. ED 134 927, 75p., PC-$4.82, MF-$0.83)

No. 4: Alessi, S. M., Anderson, T. H., & Biddle, W. B. Hardware and Software Considerations in Computer Based Course Management, November 1975. (ERIC Document Reproduction Service No. ED 134 928, 21p., PC-$1.82, MF-$0.83)


No. 7: Ortony, A. Names, Descriptions, and Pragmatics, February 1976. (ERIC Document Reproduction Service No. ED 134 931, 25p., PC-$1.82, MF-$0.83)

No. 8: Mason, J. M. Questioning the Notion of Independent Processing Stages in Reading, February 1976. (Journal of Educational Psychology, 1977, 69, 288-297)


No. 15: Schwartz, R. M. Strategic Processes in Beginning Reading, November 1976. (ERIC Document Reproduction Service No. ED 134 937, 19p., PC-$1.82, MF-$0.83)

No. 16: Jenkins, J. R., & Pany, D. Curriculum Biases in Reading Achievement Tests, November 1976. (ERIC Document Reproduction Service No. ED 134 938, 24p., PC-$1.82, MF-$0.83)


No. 20: Kleiman, G. M. The Effect of Previous Context on Reading Individual Words, February 1977. (ERIC Document Reproduction Service No. ED 134 941, 76p., PC-$6.32, MF-$0.83)


No. 59: Mason, J. M. *Reading Readiness: A Definition and Skills Hierarchy from Preschoolers' Developing Conceptions of Print*, September 1977. (ERIC Document Reproduction Service No. ED 145 403, 57p., PC $4.82, MF $0.83)


No. 70: Shoben, E. J. *Choosing a Model of Sentence Picture Comparisons: A Reply to Catlin and Jones*, February 1978. (ERIC Document Reproduction Service No. ED 150 577, 30p., PC $3.32, MF $0.83)


No. 138. Stein, N. L., & Goldman, S. Children's Knowledge about Social Situations: From Causes to Consequences, October 1979. (ERIC Document Reproduction Service No. ED 177 524. 54p., PC:$4.82, MF-$0.83)


No. 144. Pichert, J. W. Sensitivity to What is Important in Prose, November 1979. (ERIC Document Reproduction Service No. ED 179 946. 64p., PC:$4.82, MF-$0.83)
No. 157: Trabasso, T. On the Making of Inferences During Reading and Their Assessment, January 1980.
No. 166: Hansen, J., & Pearson, P. D. The Effects of Inference Training and Practice on Young Children's Comprehension, April 1980.