UNIVERSITY OF CALIFORNIA

Los Angeles

A Cross-Linguistic Interlanguage Analysis

of the

Acquisition of the English Article System

A dissertation submitted in partial
satisfaction of the requirements for the
degree Doctor of Philosophy in Applied
Linguistics

by

Peter Antony Master

1987

Dedication

I dedicate this dissertation to Dr. John Schumann, whose philosophical spirit has inspired me throughout my career at UCLA to search for meaningful answers to meaningful questions, and to my friend and fellow classmate Bob Jacobs for his unflagging support.

Table of Contents

List of Illustrations

Abstract

Many ESL/EFL practitioners, especially writing teachers, realize that the article system (i.e., *a*, *the*, and Ø) is a very difficult aspect of English syntax to learn and acquire. A few methods of teaching the system have been published, but in order to develop an efficacious methodology, it must first be determined how the spoken use of the article system is acquired in a natural (i.e., non-classroom) setting.

This study provides a picture of article acquisition by analyzing the spoken English interlanguage of speakers of five different native languages, three with no article system (Chinese, Japanese, and Russian) and two with article systems (Spanish and German). Informal interviews of four speakers of each language at successive interlanguage levels provide the basis for a pseudolongitudinal analysis of article usage for each of the five languages represented. The interlanguage level is primarily determined by the negation criteria described by Cazden et al. (1975). The use of the article system was determined in two distinct ways. The first analyzed the accuracy of article usage in regard to the English target. The second analyzed the way in which articles were used without regard to target accuracy in order to provide a picture of what speakers actually do in their evolving interlangauges.

Analysis revealed that subjects whose first languages contained an article system differed markedly in English article acquisition from those whose first languages did not contain such a system, showing that English article usage, especially at the beginning levels, is clearly influenced by the first language. The most dramatic change in article usage appears to occur between the basilang and low mesolang levels for *the*- and Ø-usage. *A* appears to be acquired at a slower and more gradual rate, perhaps reflecting its linkage to the [±count] system.

Chapter 1 Introduction

The English article system contains two major classes: definite, represented by *the*, and indefinite, represented by *a(n)* and Ø, the zero article. The basic function of the three articles is to indicate that the object of which we have formed a conception is (or is not) marked off or defined, that is, thought of within certain physical or imaginary limits.

The article system has been of interest to a line of scholars beginning with Poutsma (1914) and Jespersen (1924), continuing with Kruisinga (1932) and Christophersen (1939), and proceeding up to the present with Postal (1966), Chafe (1969, 1972), Quirk et al. (1972), Kirsner (1979), DuBois (1980), and Celce-Murcia and Larsen-Freeman (1983), to name just a few. These scholars have largely been concerned with describing the complexities of the article system as it is used by native speakers of English.

In the field of second language acquisition, the article system is an interesting area of inquiry because its three members occur so often. Indeed, *a* and *the* constitute two of the ten most frequently used words in the English language and it is hard to find a spoken or written sentence that does not contain at least one of the three articles. Despite this frequency, the articles are given short shrift in many ESL textbooks and syllabuses and by many teachers and researchers in second language acquisition. The prevailing view from the teacher/ textbook standpoint seems to be that the articles will simply get learned in the process of acquisition. From the research standpoint, the articles are mere appendages to the noun phrase, and since they are often not essential to spoken communication, they are not high on the list of research topics.

Native speakers of English normally acquire the article system by the age of three. It is a very automatic system for a native speaker, so much so that the majority of native speakers are unable to formulate even rudimentary rules for article usage. Conversely, when nonnative speakers make errors with the article system, it is easy to identify them as such, and repeated misuse of the system can even lead to irritation on the part of the non-ESL oriented native speaker (interestingly, the English speech of native speakers of Russian is stereotypically identified with the absence of articles).

If misuse of the English article system can lead to a negative disposition on the part of the listener or reader, this is naturally of concern to nonnative speakers of English, especially university students, who are frequently required to express themselves in the written mode. For it is indeed in the written mode that article errors are most glaring. Students understandably want to know how to improve their article usage,

and this has lead to several pedagogical approaches to teaching the article system.

     In order to build a truly efficacious pedagogical method for teaching the article system, a materials developer must know how the system works, how it is used by native speakers, and how it is acquired. How the system works has been described in detail by the line of grammarians cited earlier. How it is used requires a thorough contextual analysis of spoken and written English and such a study is yet to be undertaken. Pica (1983) has begun the work in the spoken domain, but much remains to be done. How the article is acquired is the subject of the present study.

     With the word *interlanguage*, the title of the present study indicates that article usage will be investigated at different stages along the path that lies between the source or native language and the target or second language. This path has been appropriately called *interlanguage* (Selinker 1972) and a means for identifying certain stages along the interlanguage continuum will be described in Chapter 5. With the word *cross-linguistic*, the title also indicates that English speakers with different native languages will be studied. What these languages are and how the speakers are to be analyzed will be presented in Chapter 5.

     After the acquisition data have been analyzed in Chapters 6, 7, and 8, an attempt will be made to find some aspect of article usage that could possibly serve to help in identifying the stages along the interlanguage continuum that were mentioned above. One reason for doing so is to buttress earlier research (namely, the negation criteria to be discussed in Chapter 5). Another has to do with the relative frequency of article usage compared to that of most other morphemes. If some aspect of the article system worked as an interlanguage measure, it would require fewer data to make an accurate assessment than the current method does.

     Finally, the results of the acquisition analysis will be considered in Chapter 1O in light of a small number of published methods for teaching the English article system. This will fulfill the initial aim of the present study, which is to provide data from the natural spoken acquisition of the English article system by speakers of other languages for the purposes of improved pedagogical presentation. Since the article system is acquired naturally over time, pedagogy can have only two aims: 1) to inhibit or reverse the process of fossilization whereby learners cease to move along the interlanguage continuum and 2) to accelerate learning/acquisition. Whether reversal of the process of fossilization is possible is unlikely although little research has been done in this area. The second aim is a distinct possibility, and research investigating the notion of

accelerated learning/acquisition will be discussed in Chapter 10.

      The present study is an attempt to provide a global view of the acquisition of the article system. The results obtained will be used to paint a picture of the acquisition process but the global nature of the study will not allow for statistical analysis. It is hoped that this global picture along with the recommendations for further research in Chapter 10 will stimulate other researchers to flesh out the picture so that more reliable conclusions can be drawn as to the nature of the article acquisition process.

Chapter 2   First Language Article Acquisition

The earliest attempts to systematically describe first language acquisition focused on the acquisition of morphemes. Berko (1958) showed that morphemes could be elicited in a test situation by the use of nonsense words (e.g., "This is a wug, now there are two of them, there are two _____"). She asked pre-school and first grade children, ranging from four to seven years of age, to supply the English plurals, verb tenses, possessives, derivations, and compounds of her nonsense words. The results demonstrated that "children in this age range operate with clearly delimited morphological rules" (p. 164) and that there are no differences between boys and girls. The differences between the pre-schoolers and the first-graders was "in the direction of perfecting knowledge they already had" (p. 165). Berko concluded that the picture of the child's grasp of English morphological rules was one of consistency, regularity, and simplicity. Unfortunately, she did not consider the English article system in her analysis.

Brown, in his landmark work *A First Language* (1973), performed a longitudinal study on the acquisition of fourteen morphemes by three English-speaking children (Adam, Eve, and Sarah). He is known for setting the widely-used 90% acquisition threshold (i.e., where 90% of the morphemes appear correctly in obligatory contexts) and was the first to systematically investigate English article usage. Brown initially tallied *a(n)* and *the* separately, but he found their overall usage to be so similar and their classification into two distinct groups so difficult that he decided to collapse his results into a single tally. Brown determined that the article is acquired eighth, after the acquisition of the progressive, *in*, *on*, plural, past irregular, possessive, and uncontracted copula.

As Brown's work was based on only three children, de Villiers and de Villiers (1973) elicited spontaneous speech from 21 children in a cross-sectional study. They found an accuracy order highly correlated to Brown's. By the age of two or three, children use articles at the 90%-accuracy level.

The failure to include the notion of the zero article is a major problem in Brown's (1973) study. By collapsing *a(n)* and *the* into a single tally, he set the paradigm for considering the article as a single morpheme, which notion has persisted up to the present (cf. Tarone, 1985). It would be worthwhile, though beyond the scope of the present study, to reanalyze Brown's data with an eye to the acquisition of the zero article as well. I suspect that such an analysis would reveal a greater spread (over time) in the acquisition of the entire article system. To ignore the zero article is to ignore at least a third of the domain of article usage.

To determine the precise age at which children begin to use articles, Scollon (1976) undertook a longitudinal analysis of the speech of a native English speaking child (Brenda Wong) at four stages (Figure 2-1). He focused on what he calls "vertical constructions", by which he means not a single utterance (as typically measured by mean length of utterance or MLU) but a series of utterances whose function is to express a full speech act (e.g., /mama.mama.mama.mam.ʃ.ʃɪ.ʃɪs.ʃu.ʃu?.ʃuʃ/ = mama's shoe). He concludes that vertical constructions, which he considers an interactive phenomenon ("Brenda's ability to make vertical constructions depends crucially on the willingness of her interlocuters to perform their part", p. 224), prepare the child's linguistic system for development.

Brenda I (1;O-1;2)	Brenda II (1;7-1;8)	Brenda III (1;1O)	Brenda IV (2;O)
(B is holding a doll) ⌐pitibebii bebi s sIk sIk	⌐s...swim boat ⌐hiding balloon ⌐car ⌐pink car pink	cat sleeping monster go drink soup banana leaf dat one cat	I like ice cream this is lobster dis goldfish dis i a blue crab [Int: What's this?] Dat a--
[pretty baby is sick]			

Note: Large brackets show vertical constructions.

Figure 2-1   Examples from Brenda's Four Stages of Development

It is apparent from Figure 2-1 that the first article to emerge in Brenda's speech is indefinite *a*, and that this occurs at the age of 2;O (two years, zero months).

Weisenberger (1976) analyzed a 2;5 year old native English child named Jill. Her research question: How does the child choose from four or more possible words the few that he will lexicalize? Her answer: "The child relies on the situation to fill in the gaps caused by his limited lexicalization capacity" (p. 281). Although the focus of Weisenberger's research is, like Scollon (1976), not on the acquisition of the article system, Jill makes the following utterances (p. 278):

I no wa go in the water
Gimme the big toast
I wa see a banana (banana was inside a bag and not
    visible)

This supports Scollon's finding that *a* appears at around two years of age and suggests that *the* appears between 2;0 and 2;5.

Klima and Bellugi-Klima (1966:424) found that the complex relation of negative and indefinite (e.g., *This is not a toy*) is, however, not yet established by the third stage (3;0+ years) of child language development, suggesting that the relation is acquired later than other aspects of article usage. This research is mentioned because one of the classification systems to be used in the present study is concerned with the indefinite article associated with negation, questions, and irrealis mode.

Maratsos (1976) focused exclusively on definite and indefinite reference in his study of native English speaking children. He found that they used both linguistic and extralinguistic contexts in determining the functions of specific/nonspecific reference in article usage and that such reference was not connected in any clear way with external physical attributes or the relations between perceived objects. Maratsos concluded that the specific/nonspecific distinction is acquired by the age of four.

Zehler and Brewer (1982) developed a "new classification" of article usage that included *a*, *the* and the zero article, which they call *null*. They obtained data from children in the initial period of article acquisition (2-3 years), younger than those tested by Berko (1958). Using "shared narration" in a play session with several toys and objects, the researchers elicited article usage from the children by essentially leaving a blank where an article plus referent would occur. For example, a doll was put through the actions of walking as the researcher said the following:

> "And now this little boy is going to go to school. And he's walking to school, and as he's walking, he looks down [researcher makes doll look down] and he sees _____"
> (p. 1270)

Having tested adults with a written form of the same narration for comparison, Zehler and Brewer determined an early article acquisition sequence of no article use, *a* use only, essentially correct *a* and *the* patterns of use, and overextended *the* use.

Piaget (1926) also found overuse of the definite article, and he attributed this to the child's egocentrism, i.e., the inability to take the hearer's perspective into account. Zehler and Brewer's study disconfirms this widely-held notion and suggests instead that children err in presuming that the hearer knows that there are only a few alternatives possible, and this is sufficient to permit the definite article to be used. In the example sentence, "Mary got into the car and sat down on <u>the</u> seat" (p. 1274), *the* is correctly supplied on this basis.

Children simply overuse this adult principle because they have not yet worked out precisely when a context nonspecific category provides no basis for familiarity (e.g., *She's wearing the blouse*.). Cziko (1986), in testing the component of Bickerton's Language Bioprogram Hypothesis (1984) that concerns the universal child language acquisition of the specific/nonspecific distinction, determined that a variety of studies support Bickerton's claims. Cziko found four stages in the acquisition of French and English articles:

> 1) the use of the definite and/or definite article(s)
>   for specific referents, and zero article for both
>   non-specific referents and naming
>
> 2) the use of the indefinite article for non-specific
>   referents, and the definite article for specific
>   referents whether or not they are presupposed
>
> 3) an increase in the correct use of the indefinite
>   article for specific, non-presupposed referents,
>   with a concomitant decrease in the correct use of
>   the definite article for specific, presupposed
>   referents
>
> 4) the correct use of the definite and indefinite
>   articles

Curiously, although Cziko cites many studies that are also mentioned in the present study, he seems unaware of Zehler and Brewer's work, which in some respects duplicates his own.
      Brown, Cazden and Bellugi-Klima (1968) is relevant to the present study only in so far as the child's phrase *that a* (as in *That a my boot* and *That a your car*) is described not as the misapplication of an article rule but rather as an unanalyzed single morpheme. Huebner's (1983a) study found the widespread use of *isa* as a topic marker in his subject's interlanguage, but Huebner never attributed this to article misuse. In analyzing the data of the subjects in the present study, especially at the basilang (i.e., lower) levels, such unanalyzed single morphemes may also arise. It would be pointless to count them as instances of the indefinite article.
      As far as first language acquisition of the English article system is concerned, the work of the researchers cited has led to the conception of the article system (i.e., the separate consideration of *a*, *the*, and Ø) that is the basis of the present study. Furthermore, the findings of Zehler and Brewer (1982) and those of Cziko (1986) as to the strategies that young children employ in article acquisition may be relevant and

perhaps comparable to the strategies second language learners of English resort to in acquiring the article system.

The next chapter reviews investigations concerned with the acquisition of the English article system by native speakers of other languages and is thus more directly relevant to the present study. The two chapters concerned with literature review do not include a survey of syntactic and discourse-related analyses of the English article system as the focus here is on the acquisition of the system rather than its description. However, a bibliography of works devoted to such description is included as part of the references section (pp.123-125).

Chapter 3 Second Language Article Acquisition

     The primary reason for including first language acquisition
research in this study is to provide a basis for comparison with
second language acquisition as, indeed, the earliest modern
research in this area attempted to do just that. This chapter
will describe a) morpheme acquisition studies, b) the
development of the negation criteria, and c) studies that
include research on the acquisition/ learning of the English
article system.

Morpheme Acquisition Studies
     Dulay, Burt and Hernandez (1973) devised the Bilingual
Syntax Measure to ascertain whether a morpheme acquisition
sequence could be identified for children learning a second
language that would parallel the findings of Brown (1973), de
Villiers and de Villier (1973), and others. Using Brown's
(1973) system of combining *a(n)* and *the* but not Ø under the
general category "Article", Dulay and Burt (1974a) found (in the
"Sacramento" group, the largest subgroup) that the mean correct
article score was the fifth highest of eight functors (see
Figure 3-1 below). Bailey, Madden and Krashen (1974) found
similar results for adult ESL learners. For Spanish speakers,
the article score was the highest of eight functors; for non-
Spanish speakers, including Japanese, it was the sixth of eight
functors (see Fig. 3-1 below). Both Dulay and Burt (1974) and
Bailey, Madden and Krashen (1974) found the majority of errors
made by their subjects to be developmental rather than
interference errors, i.e., interference from the first language
was not the primary source of error. Bertkau (1974), in her
analysis of article usage by Spanish and Japanese ESL students,
found that variants attributable to native language interference
did not appear consistently but were more likely to be the
product of simplification in the process of language learning
(e.g., the use of Ø in place of *a(n)* or *the*).
     Hakuta (1976), in his longitudinal study of a Japanese
child (Uguisu) learning English, found the articles (once again,
a(n) and *the*, but not Ø) at the 90% criterion level set by Brown
to be thirteenth out of 17 morphemes (see ranking out of eight
morphemes in Fig. 3-1 below). The child's early article usage
Hakuta described as "simply fragments retained in her speech due
to the salience and frequency of articles" and that it was not
until much later that his subject had "full control of the
semantics of the articles" (pp. 34O-341).
     Figure 3-1 shows the article ranking in several studies.
The eight functors common (more or less) to all studies are 1)
progress-sive, 2) plural, 3) irregular past, 4) possessive, 5)
third person singular, 6) contracted copula, 7) contracted

9

```
Researcher (Year)        L1*                                      .

Brown (1973)             E| ----------------------
de V. and de V. (1973)   E| ----------------
Hakuta (1976)            J| --------------------------------
Dulay et al. (1973)      S| -
Dulay et al. (1974a)     M| ---------------------
Bailey et al. (1974)     M| ----------------
Bailey et al. (1974)     S| -
Bailey et al. (1974)     J| --------------------------
Larsen-Freeman (1978)    M| -----------
                           └──────────────────────────────────
                           1    2    3    4    5    6    7
                                         Ranking

*L1: E = English, J = Japanese, S = Spanish, M = Mixed L1s
```
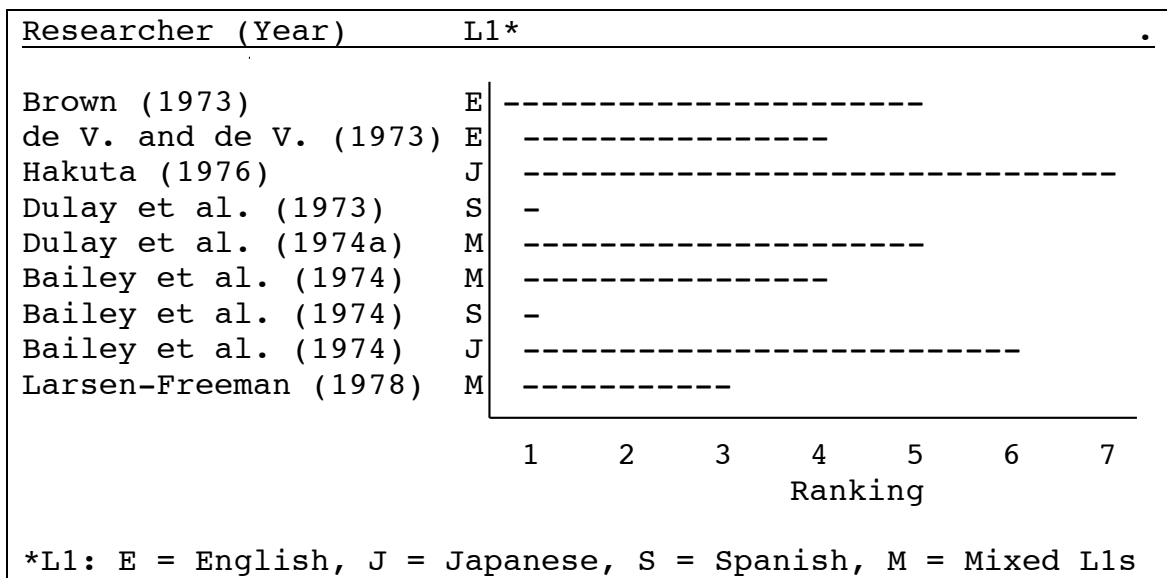
Figure 3-1. Article Ranking in Several Studies

auxiliary, and 8) articles. Contrary to the claims in most of these studies that the rankings overall correlate quite closely with other studies, the correlation is not at all obvious when seen from the point of view of a single morpheme (the article). Furthermore, despite claims that first language interference is not a significant factor, the figure shows that for the two studies with Japanese subjects (Hakuta and Bailey et al. [non-Spanish]), the ranking is much higher, i.e., later acquired. This would appear to indicate that first language is indeed a factor, at least in article acquisition.

Hatch (1978) reviewed the findings of the first and second language morpheme acquisition studies and concluded that morphemes, syntax, etc., arise out of the activity of communicating and interpreting the rules of conversation rather than as the sum total of the acquisition of individual items.

Larsen-Freeman (1978) also looked back at the earlier morpheme acquisition studies and found that "morpheme frequency of occurrence in native-speaker speech is the principle determinant for the oral production morpheme accuracy order of ESL learners" (1978:378-79). She suggested, in other words, that frequency of input determines morpheme acquisition order. However, given the inordinately high occurrence of the article in English speech, (e.g., Brown (1973) found the article to occur 552 times among the three sets of parents of his subjects whereas the regular past tense, for example, occurred only 44 times), one would expect, if Larsen-Freeman's claim is correct, greater article accuracy (or at least use) in beginning level learners than appears to be the case.

Andersen (1977) challenged the claim of earlier research that the majority of ESL errors are developmental rather than

10

interference errors. Furthermore, he did this with data concerning the article. First of all, he separated *a(n)* from *the* and is the first to do so according to Hatch (1978b:43). Secondly, he considered the Ø article as a full article, again a first. One of Andersen's several innovations was to plot the use of the two forms of Ø by Spanish learners of ESL, $Ø_1$ representing the use of Ø in English where Spanish requires *the*, $Ø_2$ representing the use of Ø in English where Spanish also requires Ø. From his data, Andersen was able to conclude that subjects perform well when the articles are the same in each language (i.e., $Ø_2$) but not nearly so well when the articles are different in the two languages (i.e., $Ø_1$). Furthermore, Andersen described the need to go beyond the notion of correct use of a morpheme in obligatory contexts to include the usage of morphemes when they are not required.

Huebner (1979) set forth Bickerton's (1975) notion of "dynamic paradigm", which seeks to describe exactly what learners do in their developing interlanguage systems rather than how closely they approximate the target. The same approach was suggested by Andersen (1977).

The work of Bailey (1973), Bickerton (1975) and Huebner (1983a) provides the core of the classification system that will be used in my study. Huebner's subject, Ge, moved through six stages in his use of *da*. "Through these changing hypotheses about the function of *da* or changing trajectories in the use of *da*, Ge arrives at a function comparable to that of SE [Standard English] *the*" (p. 146). Huebner points out the use of zero anaphora (the zero article) in environments where the existence of the referent is assumed known to the hearer, and he also hypothesizes the learner strategies of "flooding" and "trickling." Flooding is used when a linguistic form is generalized to all environments, presumably because of the form's salience or the speaker's lack of knowledge of specific rules. Trickling is the general reduction in the use of forms when a hypothesis is found to be untenable.

Kellerman (1984) analyzed the acquisition of certain structures by Dutch learners of English and German. He found that acquisition appeared to take a U-shaped trajectory in these cases. Such a trajectory is characterized by three stages: 1) target-like performance in some limited linguistic domain, 2) performance which deviates from the target, and 3) a return to target-like performance. Kellerman concluded that cross-linguistic influence plays a critical role in the manifestation of U-shaped behavior. Bowerman (1982) had earlier found similar behavior in the first language acquisition of causative and "reversative" verbs.

The Development of the Negation Criteria
     Cazden, Cancino, Rosansky, and Schumann (1975) established
the negation criteria for interlanguage level. Three
generalized levels are identified and named basilang, mesolang,
and acrolang to parallel the levels of basilect, mesolect, and
acrolect that occur in pidgin development. These interlanguage
stages are used in the present study to identify appropriate
subject level in approximating a longitudinal study. Stauble
(1977) expanded on the Cazden et al. (1975) negation analysis
and delineated the negation characteristics exhibited at each
stage of development by two native Spanish speakers over a 10-
month period. Stauble (1981) compared the negation
characteristics exhibited by six native Japanese speakers to
those she found for Spanish speakers. She concluded that "a
second language learner's negation characteristics can be
employed as a gross measure of his verb phrase morphology
development" (p. 351).

Studies Concerning the Acquisition/Learning of the English   !
!Article System
     Neuman (1977) analyzed the composition errors of 158
intermediate-level learners and found the greatest number (22%)
to be with the article system. She also found Japanese and
Korean students to have more article omission errors than other
learners. Negation errors, on the other hand, accounted for
only 0.5% of the errors found. This finding lends support to
one of the aims of the present study, which is to establish an
interlanguage measure based on article production rather than
negation.
     Yamada and Matsuura (1982) investigated article usage in
Japanese students. Although they looked for target-like usage
rather than a nontarget interlanguage system, the study is an
improvement on earlier work in that 1) the article is broken
down into separate tallies for *a(n)*, *the*, and Ø and 2) the full
range of article usage, rather than only usage at the 90%+
accuracy level established by Brown, (1973), is considered. In
general, Yamada and Matsuura found that the overall difficulty
order for intermediate level Japanese ESL students, from easiest
to hardest, was *the* > *a(n)* > Ø, whereas the difficulty order for
advanced level Japanese ESL students was *the* > Ø > *a(n)*. They
attributed most of the general difficulty to the specific/
nonspecific distinction, which does not occur in Japanese and
would therefore count as an interference error. In accounting
for the fact that advanced learners continue to make article
errors about 30% of the time in written English, the researchers
claim that "the articles had not received their (i.e., the
students') attention," which shows the need for a pedagogical
presentation of the article system.

Lamotte, Pearson-Joseph, and Zupko (1982) challenge the Cazden et al. (1975) negation criteria later delineated by Stauble (1978 and 1981) as an effective interlanguage measure. Using the TLU (target-like utterance) measure devised by Stauble (1981) to take into account the use of a morpheme outside its obligatory context, Lamotte et al. investigated the use of *a(n)* and *the* by Spanish, Japanese, and Vietnamese subjects who were identified as to interlanguage level by the Cazden et al. negation criteria, which were based exclusively on Spanish-speaking subjects. They found a "smooth progression" (i.e., continuous increase) in the use of the articles for the Spanish subjects, but not for the Japanese and Vietnamese ones, who evinced "variations within stages, across stages and across languages" (p. 8). This led them to conclude that the negation criteria may work for Spanish speakers but "may not be the most accurate representation of second language acquisition for all language groups" (p. 8).

Since the identification of the interlanguage level of the subjects in the present study is dependent on the negation criteria proposed by Cazden et al. (1975), Lamotte et al. (1982) is a direct challenge to the methodology on which the present investigation is based. However, three factors employed by Lamotte et al. have led me to discount their rejection of the negation criteria as an effective measure: 1) they do not consider the use of the zero article, 2) they include the use of articles with proper names, and 3) their rejection is based on the fact that the TLU score does not increase consistently across interlanguage levels. The zero article has already been described as an essential element of article usage, but Lamotte et al. avoid it because "with more advanced speakers, it is impossible to distinguish between informed, target-like use of [the] zero-article and non-systematic or arbitrary non-uses of any article which turn[s] out to be correct" (p. 4). In spite of this accurate characterization of the zero article (to be discussed at length in Chapter 7), its "use", in my opinion, cannot be ignored. Secondly, proper nouns should not be included in an analysis of article usage because learning the arbitrary rules that apply may be a function of individual experience (this notion will be discussed in Chapter 5). Finally, the TLU measure has certain problems (see Chapter 7), but the expectation of a "smooth progression" in article acquisition is not really justified, especially in light of Kellerman's (1984) U-shaped learning hypothesis. For these reasons, the methodological basis of the present study has not been changed.

Parish, Tarone, and Taghavi (1986) investigated the effect of task on the production of certain morphemes, including the article. In the three tasks investigated (a grammar test, free

conversation, and an oral narrative), the morpheme production accuracy varied. A criticism of an earlier paper (Tarone 1985) given at the Los Angeles Second Language Research Forum (SLRF) was that the article should not be considered a single morpheme a là Brown (1973), Dulay and Burt (1974), and others, especially as article usage in Tarone (1985) showed a different pattern than the other morphemes studied. Parish, Tarone, and Taghavi (1986) then retroactively applied Huebner's "semantic wheel for article usage" (1983b) to Tarone's (1985) data but continued to link production to target accuracy, thus going against the spirit of Bickerton's dynamic paradigm. In the present study, the data will be looked at both in terms of dynamic paradigm and, separately, in terms of accuracy vis-à-vis the target. The "task" in all cases will be participation in an informal interview. The findings of Parish et al. (1986)suggest that an oral narrative (i.e., giving and following instructions) would provide the highest accuracy, but since such data are not available for the present study, the informal interview (corresponding to Tarone's "free conversation" task) will have to suffice. However, since the "task" is the same for all 20 subjects in this study, the relative characteristics of article usage among the subjects should not be affected.

Acquisition of the article system by learners of English as a second language has not been studied in the degree of detail proposed in the present study, nor with such a wide cross-linguistic scope. What the study gains in breadth of outlook, however, it loses in statistical significance and hence generalizability. For this reason, the study should be seen as providing an overall picture of article acquisition that allows some speculation as to tendencies and apparent strategies but which cannot make concrete claims. The ability to make such claims requires a replication of the study using not one but 5-10 subjects for each cell, i.e., 100-200 subjects. It is hoped that such a replication will one day be undertaken.

Chapter 4 Introduction to the Present Study

A true picture of article acquisition should be based on longitudinal studies like that of Huebner (1983b), which analyzes the English acquisition of a Hmong speaker. Unfortunately, such a study requires a considerable period of data gathering, an impossibility under the present circumstances. The longitudinal approach is approximated in the present study by using four speakers of the same native language at four different stages of English interlanguage development. Cazden, Cancino, Rosansky, and Schumann's (1975) negation criteria are used to identify subjects at the appropriate level.

The English interlanguage of a basilang (BA) speaker, a low-mesolang (LM) speaker, a mid-mesolang (MM) speaker, and a high-mesolang (HM) speaker of each language are analyzed for article usage, resulting in the generation of 20 cells (4 stages X 5 languages) to represent five pseudolongitudinal pictures of article acquisition. The native languages of the subjects selected for analysis are three that do not have an article system (Chinese, Japanese, and Russian) and two that do (Spanish and German).

This investigation of the acquisition of the English article system takes two directions. The first looks at article aquisition in terms of its correctness or approximation to the target; the second views article acquisition in terms of learner usage, that is, without regard to accuracy or target. As mentioned in Chapter 3, the morpheme acquisition studies of the 1970's were based entirely on the former approach. Morphemes were considered acquired after a certain threshold of accuracy had been crossed, usually 90% (Brown, 1973). But as Andersen (1977) pointed out, the consideration of morphemes only above a certain threshold ignored a sizeable quantity-- indeed, the bulk-- of useful data. For this reason, in the present study no absolute "threshold of acquisition" is postulated. Instead, development in terms of accuracy is considered across all interlanguage levels. The notion of U-shaped learning behavior (Kellerman, 1984) will also be noted in seeking confirmation of the first hypothesis of this study.

Hypothesis I: Accuracy in the use of the English
              article system reflects overall
              linguistic competence.

Bickerton (1975), Andersen (1977), and Huebner (1979) challenged morpheme studies that were based entirely on accuracy. They argued that this view ignored the actual strategies that a nonnative speaker employed in acquiring syntax, among other aspects, in a second language. Huebner's

(1983b) longitudinal analysis of a Hmong speaker describes what
learners do in their developing interlanguage systems, leading
him to postulate the notions of "flooding" and "trickling",
described in Chapter 3, as a strategy for acquiring English *the*.
This forms the basis of the second hypothesis of this study.

       Hypothesis II: The use of the English article system
                     reflects certain strategies of
                     interlanguage development, some of
                     which are universal, others language
                     (or perhaps learner) specific.

Chapter 7 is concerned with the acquisition of the English
article system in terms of Hypothesis I. Chapter 8 is concerned
with the acquisition of the system in terms of Hypothesis II.

Chapter 5   Method

        As described in Chapter 4, the subjects of this study
comprise twenty nonnative speakers of English representing five
language groups. The data are taken from tape-recorded informal
interviews that were conducted between 198O and 1986. The
majority (15/20) of the interviews were conducted by students
from Dr. John Schumann's "Contrastive Analysis" classes (English
241K) at UCLA during the years mentioned above. The remaining
five interviews were conducted by the author.
        Students in the Contrastive Analysis class are required to
record an informal interview with a nonnative speaker of
English, to transcribe the interview, and then to perform an
interlanguage analysis of the data. This consists of making a
list of "pulled utterances" reflecting the use of different
morphemes, primarily those concerned with verb phrase
morphology. One of the most important aspects of this process
is the determination of the subject's use of negation. Cazden,
et al. (1975) established that interlanguage level could be
ascertained, at least for Spanish speakers, by verbal negation
structures in the following pattern, as introduced in Chapter 3.

     1. Basilang: predominance of *no* or *not* + verb

     2. Low Mesolang: predominance of unanalyzed *don't* + verb

     3. Mid-Mesolang: predominance of auxiliary + negation +
             verb

     4. High Mesolang: predominance of analyzed *don't* + verb

Each of the subject's interviews was analyzed by the original
interviewer in terms of negation as well as certain aspects of
noun phrase morphology, namely plural and possessive markers,
but not for article use. A prediction was then made as to the
subject's interlanguage (IL) level.

Selection of Subjects
        In selecting subjects for the present study, I looked at
the subject's negation characteristics, the interviewer's stated
IL level, and the subject's morpheme acquisition at the 20%,
50%, and 70% accuracy levels in order to make sure that I had
subjects with the desired native language who were at the
appropriate IL level. In all cases, the predominance of
negation form (Table 5-1) was the strongest criterion for
selection, as morpheme acquisition at various percentages of
accuracy is at best an approximate indication of IL level.
Indeed, negation also has its problems as an IL measure

Table 5-1. Verbal Negation in the Twenty Subjects

# SUBJ	L1	*No(t)* +V		Unan *don't*		Aux-neg		An *don't*		Interviewer
Basilang										
1. Ah Chun	CH	31/32	97%	1/32	3%	0/32	0%	0/32	0%	Lin
2. Elisabeth	GE	39/44	89%	0/44	0%	5/44	11%	0/44	0%	Ediger
3. Amy	JA	71/84	85%	6/84	7%	3/84	4%	4/84	5%	Moore
4. Gersona	SP	51/51	100%	0/51	0%	0/51	0%	0/51	0%	Bechet
5. Kinga	PO	55/75	73%	10/75	13%	10/75	13%	0/75	0%	Cannon
Low Mesolang										
6. Guican	CH	1/50	2%	24/50	48%	21/50	42%	4/50	8%	Li
7. Blanca	SP	7/83	8%	48/83	58%	24/83	29%	4/83	5%	Bernstein
8. Yoko	JA	2/56	3%	25/56	45%	24/56	43%	5/56	9%	Walters
9. Repin	RU	5/52	9%	31/52	60%	15/52	29%	1/52	2%	Newmark
10. Tina	GE	2/32	6%	17/32	53%	13/32	41%	0/32	0%	Master
Mid Mesolang										
11. Dr. X	CH	1/37	3%	8/37	22%	28/37	75%	0/37	0%	Taylor
12. Mako	JA	0/24	0%	11/24	46%	12/24	50%	1/24	4%	Jacobs
13. Ruben	SP	1/64	2%	18/64	28%	32/64	50%	13/64	20%	Gregory
14. Nina	RU	1/27	4%	12/27	44%	14/27	52%	0/27	0%	Flashner
15. Reto	GE	0/60	0%	27/60	45%	33/60	55%	0/60	0%	Master
High Mesolang										
16. Igor	RU	0/82	0%	0/82	0%	38/82	46%	44/82	54%	Master
17. Hitomi	JA	0/135	0%	0/135	0%	32/135	39%	83/135	61%	Master
18. Joan	GE	0/104	0%	0/104	0%	35/104	34%	69/104	66%	Billings
19. Juan	SP	1/37	3%	0/37	0%	14/37	38%	22/37	59%	Borkowski
20. Mingte	CH	0/56	0%	6/56	11%	10/56	18%	40/56	71%	Master

(e.g., negation in certain fossilized English speakers can reflect a higher IL level than is appropriate). One of the purposes of the present study is to determine if article usage or some aspect thereof can function as a more reliable interlanguage measure. This will be discussed in Chapter 9.

As a result of this selection process, 20 subjects were selected. Table 5-2 provides background information on each of the subjects. In the group as a whole, there are nine males and eleven females. The age range is 13 - 93 years. The length of time in the U.S. ranges from three weeks to 32 years. The level of education ranges from elementary to university. The subjects in each of the five language groups share the same first language, a dialect thereof, or a genetically-related language. Thus the Chinese group includes speakers of both Mandarin and Cantonese; the German group includes speakers of both Standard German and Swiss German; the Spanish group includes speakers from Central and South America. The one exception is the

Table 5-2 Background Data on the Twenty Subjects

SUBJECT	L1	Age	Sex	Birthplace	Occupation	Other Langs.	Age ar US	Time in US	Level of Education	English in Land	English in US	Date of Interview
Ah Chun	CH	63	F	China	housekeeper	none	47	16 yrs	elementary	1 yr	none	Winter 82
Elisabeth	GE	21	F	Switzerland	sport train	FR	21	3 wks	tech schl.	9 mos	3 wks	3-4-85
Amy	JA	54	F	Japan	housewife	none	32	18 yrs	little	none	none	Winter 82
Gersona	SP	48	F	Honduras	housekeeper	none	43	5 yrs	junior HS	none	3 yrs	3-19-85
Kinga	PO	93	F	Poland	grandmother	none	61	32 yrs	high schl?	none	none	1982
Guican	CH	43	M	China(PR)	professor	none	43	6 mos	university	little	6 mos	3-24-86
Blanca	SP	30	F	El Salvador	housekeeper	none	23	7 yrs	high schl	none	7 mos	Winter 84
Yoko	JA	30	F	Japan	fashion des	none	25	5 yrs	high schl	8 yrs	2 yrs	3-19-84
Repin	RU	60	M	Russia	professor	GE,IT	58	1.5yrs	university	none	none	1982
Tina	GE	47	F	Germany	med. asst.	FR	46	.75yrs	prof.schl.	none	4 mos	11-18-86
Dr. X	CH	60	M	China(PR)	professor	RU,JA	60	2 mos	university	little	none	3-26-86
Mako	JA	13	M	Japan	JHS student	none	13	3 mos	8th grade	1.5 yrs	3 mos	Winter 85
Ruben	SP	26	M	Argentina	activist	IT	19	7 yrs	high schl.	none	4 yrs	3-16-84
Nina	RU	66	F	Russia	engineer	none	59	7 yrs	university	little	2 yrs	1982
Reto	GE	23	M	Switzerland	mechanic	FR	23	9 wks	high schl.	5 yrs	10 wks	6-5-86
Igor	RU	25	M	Russia	pre-med	none	20	6 yrs	university	6 yrs	1 yr	10-30-86
Hitomi	JA	34	F	Japan	TESL stud.	none	21	15 yrs	university	6 yrs	10 yrs	10-25-86
Joan	GE	45	F	Germany	housewife	RU	24	21 yrs	high schl.	none	none	Winter 80
Juan	SP	26	M	Argentina	student	none	26	1 mo	high schl.	3 yrs	5 wks	3-19-85
Mingte	CH	25	M	Taiwan	student	none	25	7 mos	university	?	7 mos	6-4-86

representative of the Russian group. Her first language is not Russian but Polish. However, the two languages are both of Slavic origin, and neither has an article system. Furthermore, I used Kinga (the Polish speaker) in a pilot study that included Russian speakers (Master 1984b) and found her not to diverge in any significant way from Russian speakers in her use of the article system.

The Interviews

The interviews are informal in so far as they are simply elicitations of the subject's speech without any preplanned questions. As many of the subjects are immigrants to the United States, the topics discussed commonly concerned why the subject came to the U.S., what life was like in the native country, what the subject's occupation was, what the subject felt about the U.S., what the subject planned to do in the future, or recountings of stories or film plots. The interviews average 60 minutes in length, with a range of 40 to 75 minutes.

Transcriptions

The interviews were all transcribed by the original interviewer. However, the assignment did not require an analysis of article usage. Since the article is generally an unstressed morpheme in English, I could not be sure that the interviewer had correctly transcribed all instances of article usage. For this reason, I obtained a copy of the original tape

recording of each interview and listened to it carefully while
following the transcript, making any corrections necessary. For
the most part, the transcriptions were quite accurate, although
it was apparent from this phase of the study that transcriptions
in general should be as detailed as possible (i.e., including
every repetition, false start, "um", etc.) in order to catch
unstressed morphological or syntactic features.

Pulled Utterances
     Once the transcript had been double checked against the
original tape-recorded interview, a list of pulled utterances
was created for each subject. Each entry indicated the number
of the utterance, the page and line number from where it came in
the transcript, the noun phrase being considered plus minimal
attendant context to justify its classification, the article
used, the article required, and the classification code. An
example from the MM German subject illustrates the format.

#	Line	Noun Phrase	Used	Required	Class
216	8-41	the winner from this group	the	the	2d

<div align="center">Figure 5-1   Format of Pulled Utterances</div>

In some cases, especially at the basilang (BA) and low mesolang
(LM) levels, a decision had to be made from context as to
whether the speaker had used an epenthetic vowel (e.g., *have*
pronounced as /hævə/) instead of the indefinite article. This
decision was based on the occurrence, if any, of epenthetic
vowels in other structures in the same way that Huebner (1983b)
counted *isa* as a single morpheme. Similarly, context was used
to decide whether the speaker had really used the Ø article or
had simply neglected to add the final *-s* morpheme. For example,
the Chinese BA speaker Ah Chun uttered the sentence "Do
cabbagee" [= we had to plant cabbage(s)]. *Cabbage* can be either
a [+count] or a [-count] noun and the classification system used
in this study requires the distinction of the two. In this
case, I reasoned that the [-count] form *cabbage* would be more
appropriate for cooking or harvesting, but that for planting,
the [+count] form *cabbages* is probably more appropriate. Since
Ah Chun never uses the plural *-s* form, I counted "cabbagee" in
her transcript as Ø + *cabbage* + deleted *-s*, i.e., a correct use
of the Ø article. Whenever such decisions were made based on
context, the pulled utterance was flagged with a question mark.
The number of question marks is highest for the basilang
speakers and diminishes to zero by the mid-mesolang level. Even
so, question marks in Ah Chun's pulled utterances account for
only 2.6% (8/389) of the total.

Classification Codes
     The classification codes used in this study utilize
Bickerton's "semantic wheel for noun phrase reference" as cited
by Huebner (1983a). This paradigm divides the article system
into four major categories, as shown in Figure 5-2. These

```
                    ┌─────────────────────────┐
                    │  SPECIFIC     GENERIC    │
                    └─────────────────────────┘

┌───┐
│ D │
│ E │
│ F │         CATEGORY II    CATEGORY I
│ I │
│ N │         Specific definite   Generic
│ I │
│ T │         Article: the    Articles: a,Ø,the
│ E │
│   │         Features:[+SR][+HK]  Features:[-SR][+HK]
│   │
│ I │
│ N │         CATEGORY III   CATEGORY IV
│ D │
│ E │         Specific indefinite  Indefinite generic
│ F │
│ I │         Articles: a,Ø   Articles: a,Ø
│ N │
│ I │         Features:[+SR][-HK]  Features:[-SR][-HK]
│ T │
│ E │
└───┘
```
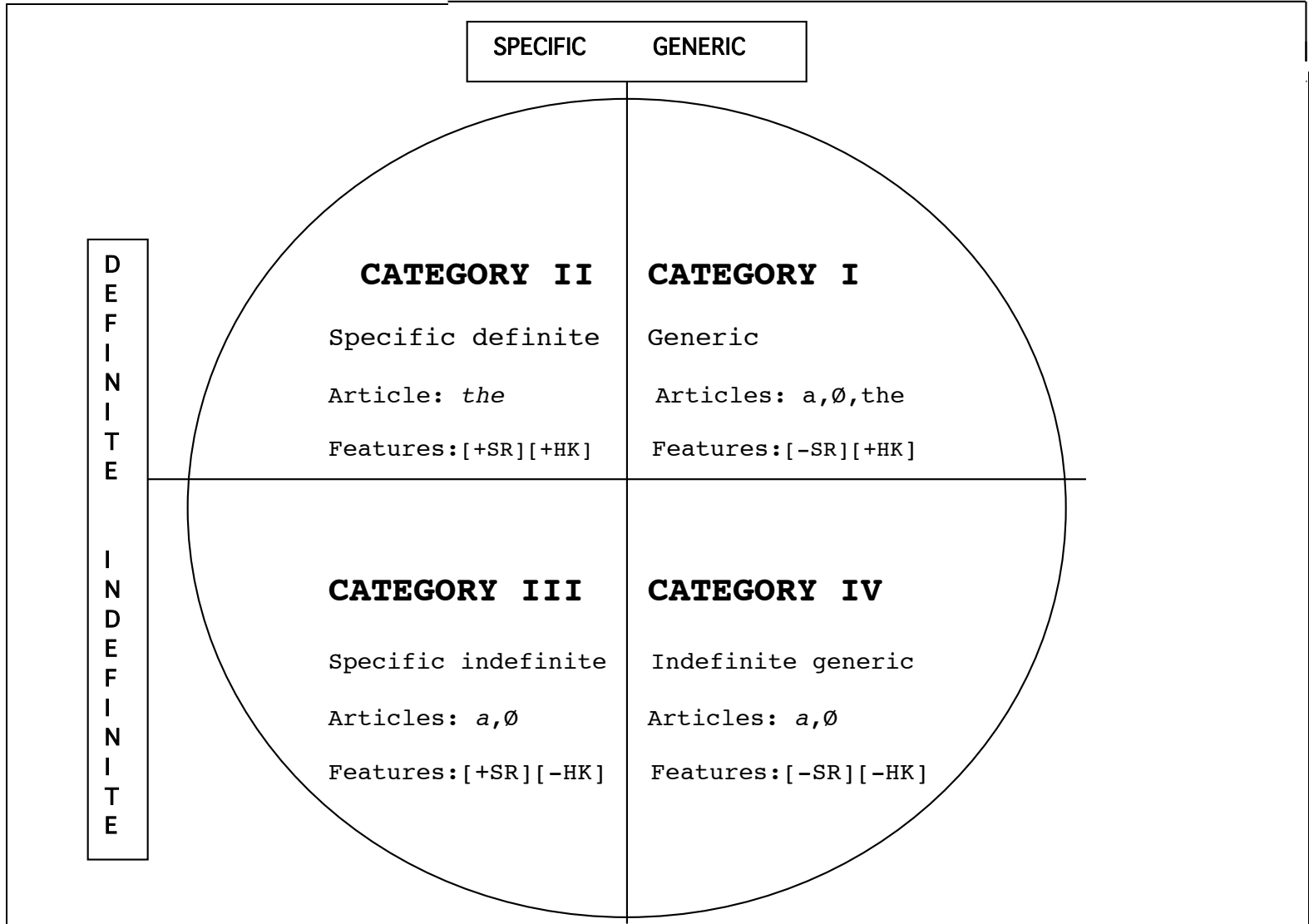
Fig. 5-2. Bickerton's "Semantic Wheel of Noun Phrase Reference"

categories are a considerable improvement on Brown (1973) and
the later SLA morpheme studies (which consider the and a as a
single morpheme and do not count Ø as an article), but they
still place several different aspects of article usage within a
single category. Bickerton's Category II, for example,
describes the class of environments requiring specific (as
opposed to generic) the in English. However, subsumed under
[+SR,+HK] are several, sometimes overlapping, uses of the:

1. subsequent mention [i.e., with an identifiable first referent] (e.g., I bought a sandwich. <u>The sandwich</u> was stale.)

2. shared knowledge (e.g., <u>The sun</u> is shining today.)

3. ranking adjectives (e.g., <u>The smallest</u> atom is hydrogen.)

4. postmodified specific nouns (e.g., <u>The man whom I met</u> was a doctor.)

5. specific nouns with implied postmodification (e.g., <u>The driver</u> didn't see us.)

6. specific nouns of which knowledge is shared that there is only one or that all possible instances are meant (e.g., He drove to <u>the capital</u>. I admire <u>the Russian people</u>.)

Bickerton's semantic wheel essentially separates definite (Category II) from indefinite article usage (Category III) and clear generic (Category I) from indefinite generic usages (Category IV). Huebner's adaptation of the semantic wheel does not count idiomatic usages, set phrases (e.g., *What's the matter?*), and proper nouns.

In a textbook for foreign students (Master 1986a), I set forth a classification of article usage based for the most part on the work of grammarians ranging from Poutsma, Jespersen, and Christophersen to Quirk, Greenbaum, Leech and Svartvik. That schema is as follows:

      I. Specific
          A. Indefinite (*a*, Ø)
              1. First mention
              2. General characteristics
              3. *a = one*
          B. Definite (*the*)
              1. Subsequent mention
              2. Ranking adjectives
              3. Shared knowledge
              4. Postmodified nouns
     II. Generic
          A. Abstract generic (*the*)
          B. Concrete generic (*a*, Ø)
    III. Idiomatic structures
          A. Idiomatic structures with Ø
          B. Idiomatic structures with *a*
          C. Idiomatic structures with *the*

In the present study, I have combined the two classification
schemas above into 19 classifications, as presented in Table
5-3. This model retains the four categories Bickerton proposed
with certain additions. For example, Bickerton's Category I
(Generic [-SR,+HK]) is expanded to include certain kinds of
idiomatic structures, namely preposition + Ø + singular count
noun (e.g., *by car*). His Category III (Specific [+SR,-HK]) is
expanded to include instances of *a = one* (e.g., *He makes $30,000
a year*) and the relatively infrequent Ø + rank, post, etc.
(e.g., *Mr. Jones, director of the program*). The resulting 19
classifications are described and exemplified in Table 5-3.

Table 5-3. Classification Codes Used in This Study

CODE	DESCRIPTION	EXAMPLE Source	Noun Phrase + Context
I. Generic (*a*, *the*, Ø)			
1a.	generic *the*	Igor 5-38	how the AIDS virus affects the nervous system
1b.	generic *a*	Igor 6-42	to call a virus a living thing
1c.	generic Ø	Igor 4-2	they're behind…in medicine
1d.	PREP+ Ø + sing count noun	Igor 7-3	he saw his brothers going to school
II. Specific Definite (*the*)			
2a.	subsequent mention	Hitomi 5-8	[the magazine he was holding] my purse between the magazine
2b.	shared knowledge	Hitomi 5-31	I was waiting for the bus
2c.	ranking adjectives*	Hitomi 8-46	that was the second time
2d.	postmodification	Hitomi 9-2	the guy who corrected my test
2e.	implied postmod	Hitomi 5-29	I didn't like the students
2f.	shared knowledge of uniqueness	Hitomi 4-5	they don't care [about] the outside world
III. Specific Indefinite (*a*,Ø)			
3a.	count singular (*a*)	Joan 3-26	she got a shot
3b.	count plural (Ø)	Joan 3-25	I took antibiotics
3c.	noncount (Ø)	Joan 4-6	[he] was making money
3d.	*a = one*	Joan 1-30	once a week they have a test
3e.	Ø + rank or post	Nina 13-32	my husband is president from the old people
IV. Ambiguous Generic (*a*, Ø)			
4a.	equative	Reto 2-30	it's a small city
4b.	negated	Reto 12-40	that's not a good idea
4c.	interrogative	Nina 17-34	you have a room?
4d.	irrealis mode	Ruben 17-9	if a policeman give you a ticket

* Ranking adjectives contain three subclassifications: superlatives (e.g., *the largest city*), sequence adjectives (e.g, *the first time*, *the nex day*, *the last chapter*), and unique adjectives (e.g., *the same story*, *the only exception*, *the main argument*). The exceptions to this rule (e.g., *next week*, *last year*) were tallied separately as Classification 2x.

The Tally

In order to facilitate both the speed and accuracy of the final count, a data file was created on an Apple II Plus computer using a data management program called Datadex (Information Unlimited Software). The data from the lists of pulled utterances, specifically the number of the noun phrase, the article used, the article required, and the classification code, were entered into the program. This second entry phase allowed a double check of the original classification made on the pulled utterance lists for each subject. It also allowed a reconsideration of the noun phrases marked with a question mark, which in a few cases lead to a reclassification of that noun phrase.   When the data for a subject had been entered and checked (by means of an accumulator function which confirmed that the supposed number of noun phrases had been entered), various data sorts were performed, allowing the production of data tables from which graphs could be constructed. Due to the presence of both the article used (including the words *one*, *this*, and *that*, which some subjects occasionally substituted for an article) and the article required in each entry, the data could be tallied in terms of both accuracy (article used equals/does not equal article required) and usage (article required ignored) and the appropriate data tables assembled. For the indefinite articles, both *a* and *an* were counted as correct, even if *a* was incorrectly used before a vowel sound (e.g., **a apple*). This decision was made because I was interested in the selection of *a* over Ø and *the* and not whether the speaker had learned the phonetic requirement of using *an* before a vowel sound. However, *one* or *ein* in place of *a* was counted as an incorrect usage. Similarly, *de*, *ze*, *der*, and *die* (*der* and *die* because they sound so similar to English *the* in rapid speech) were counted as correct forms of the definite article, whereas *that* and *this* were counted as incorrect.

In my original proposal, I intended to investigate the articles with proper nouns. After a thorough analysis, however, I found that there seemed to be no clear pattern of acquisition short of a generalized tendency to improved accuracy with increasing interlanguage level and that a subject's knowledge and usage of the somewhat idiosyncratic rules for articles with proper nouns tended to depend on his or her experience in the world. Thus, the houseworkers, for example, knew and used street names, the travelers knew and used the names of rivers and mountains, and the students knew and used the names of languages. However, not every subject used every type of proper noun, and the N for some types of proper nouns was quite low. For this reason, like Huebner (1983b), I decided to confine my investigation to the use of articles with common nouns.

The data were analyzed in light of the two hypotheses set forth in Chapter 4. Hypothesis I concerns accuracy or degree of approximation to the target. Hypothesis II concerns usage or a description of the speaker's evolving interlanguage without regard to target. The goal of this bidirectional approach is to "paint a picture" of article acquisition.
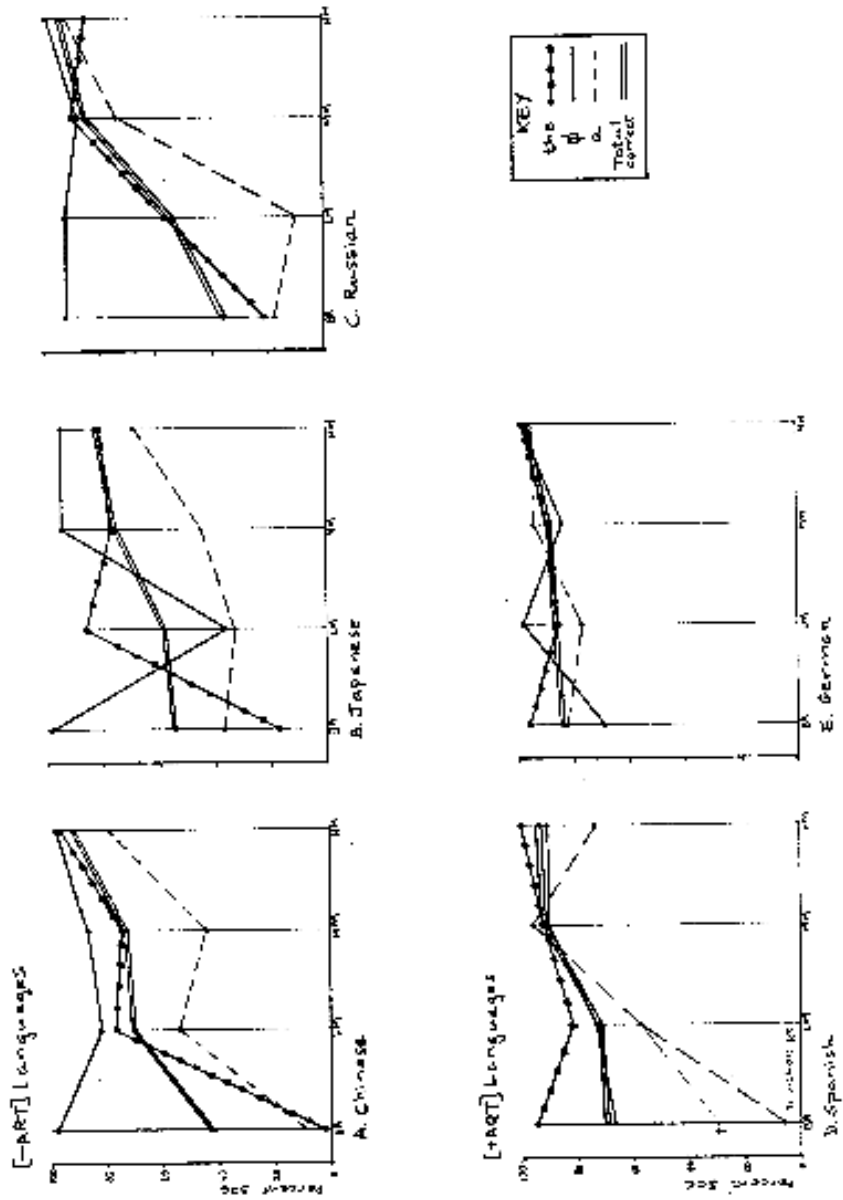
ACCURACY
    Accuracy in this study is based on the standard notion of supplied in obligatory context, SOC. SOC indicates the number of correct items divided by the number of environments or obligatory contexts in which the article should be used. Since we are interested in patterns of English article acquisition by speakers of different native languages, the data are first presented as graphs of article accuracy by the four speakers of each of the five language groups.

Article Accuracy (SOC) for Each Language Group
    Article accuracy (SOC) for each language group is shown in Figure 6-1. The [-ART] languages (Chinese, Japanese, and Russian) are at the top, the [+ART] languages (Spanish and German) at the bottom. The double line indicates the accuracy of the three articles combined. A comparison of the [-ART] graphs shows them to have four elements in common:

    1) Ø is the most accurate, starting near the 100% accuracy
        level for all three [-ART] BA subjects. There is a
        slight drop in accuracy at the LM stage for the
        Chinese subject, at the MM stage for the Russian
        subject. The Japanese LM subject shows the most
        marked decrease in Ø-accuracy, which is partially a
        product of her overuse of *the* (*the*-"flooding", to use
        Huebner's 1983a term) at the expense of Ø-accuracy.

    2) *the* shows the second highest accuracy for all three
        [-ART] subjects. There is a sharp increase in
        accuracy between the BA and LM stage for the Chinese
        and Japanese subjects, the Japanese subject being
        higher because of her extreme *the*-flooding. The
        Russian subjects increase less dramatically but
        consistently to the MM level, and then drop off.

    3) *a* is lowest in accuracy for all three [-ART] subjects.
        Each of the subjects shows a peak followed by a
        trough, although the IL level at which this happens is
        different: LM for the Japanese and Russian subjects,

25

Fig 6-1 Accuracy (Soc) for Each Language Group

[-AQT] Languages

A. Chinese

B. Japanese

C. Russian

KEY

[+AQT] Languages

D. Spanish

E. German

26

MM for the Chinese subjects.

4) the total accuracy plots show a roughly similar pattern for all the [-ART]subjects, and each shows continuous increase in accuracy, with no troughs, across all IL levels and is hence a potential candidate for an IL measure (to be discussed in Chapter 9).

A comparison of the [+ART] group shows them to have three elements in common:

1) *the* is the most accurate, starting near the 100% accuracy level for both [+ART] subjects. There is a slight drop in accuracy at the LM level for both the Spanish and the German subjects.

2) even though they start at different levels, *a*-accuracy for both groups exceeds *the*-accuracy at the MM stage (*a*-accuracy would have been higher for the Spanish BA subject if her 21 uses of *one* in place of *a* were included).

3) the total accuracy plots show a roughly similar pattern for both the [+ART] subjects. As in the [-ART] group, each shows a continuous increase in accuracy, with no troughs, across all IL levels and is thus a candidate for an IL measure.

Overall Accuracy of the Individual Articles
     Since Fig. 6-1 suggests that that [-ART] language group shows generally similar patterns of accuracy for each article and the [+ART] language group also has generally similar patterns of accuracy, the SOC for each article is plotted in the same graph for the sake of comparison.[1] The overall accuracy of the individual articles for all five language groups together is shown in Figure 6-2. Ø-accuracy (Graph A) shows considerable confluence between the MM and HM levels for all five groups. At the BA level, the [-ART] group clusters above the 90%-accuracy level whereas the [+ART] group lies below the 70%-accuracy level. If it were not for the extreme *the*-flooding of the Japanese subject, clustering at the LM level among the [-ART] group might have been greater (the notion of individual variation will be taken up in Chapter 7).

_____
[1]Despite the fact that the tenets of the Columbia form/content school only apply to native speaker usage (Garcia 1975), the consideration of the form (i.e., Ø, *a*, and *the*) over the function adheres to the spirit of this approach.

Fig 6-2 Overall Accuracy(SOC) of the Individual Articles
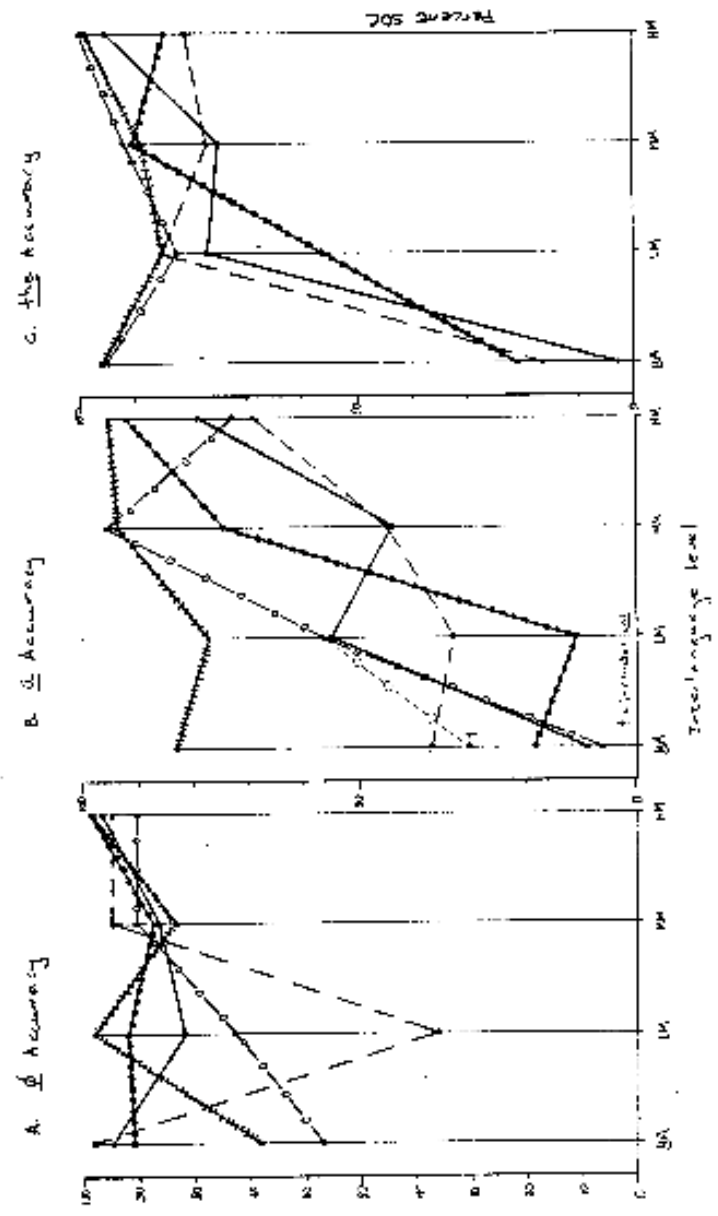
A-accuracy for the [+ART] group at the BA level shows the widest range of all overall article usage, but this is basically because the Spanish BA subject performed at such a low accuracy level (even with *one* included)-- or else because the German BA subject performed at such a high level. Individual differences may account for this lack of clustering. Otherwise, it may reflect the linkage of *a* to the whole [±count] system in English, a system that requires considerable lexical competence for high accuracy-- to be discussed at length in Chapter 7. Even at the HM level, *a*-accuracy is the lowest for the three articles, suggesting that it is the most difficult to acquire.

The-accuracy shows a strong [+ART]/[-ART] split at the BA level, the [+ART] subjects clustering above the 90%-accuracy level much as did the [-ART] subjects with Ø. The BA [-ART] subjects cluster below the 20%-accuracy level. The Chinese and Japanese LM subjects then shoot up above the 75%-accuracy level, the Russian subjects achieving a similar height at the MM level. The [+ART] group forms a trough at the LM level and then rises steadily to near 100%-accuracy, forming a typical shallow U-shaped curve, as does the [-ART] group with Ø.

Overall Article Accuracy: [+ART] versus [-ART]

Overall article accuracy for the [+ART] group versus the [-ART] group is shown in Figure 6-3. Combining the groups in this way was motivated by the fact that the different language groups show similar patterns of development in Figures 6-1 and 6-2. In Figure 6-3, [+ART] *the*-accuracy again starts high at the BA level, as does [-ART] Ø-accuracy. [-ART] Ø-accuracy goes on to form the now familiar U-shaped curve. [-ART] *the*-accuracy shows the meteoric rise at LM (LM Japanese *the*-flooding compensating for the lower accuracy of the Russian LM subject), which then continues to climb to higher accuracy at the HM level. [-ART] *a*-accuracy develops in a virtual straight positive slope across all four IL levels.

Conclusions

The conclusions that can be made at this point are the following:

1. the [-ART] accuracy (SOC) pattern differs markedly from the [+ART] accuracy (SOC) pattern

2. the [-ART] group appears to attain ~90%-accuracy for the whole article system at the high mesolang (HM) level, whereas the [+ART] group appears to attain ~90%-accuracy at the mid-mesolang (MM) level
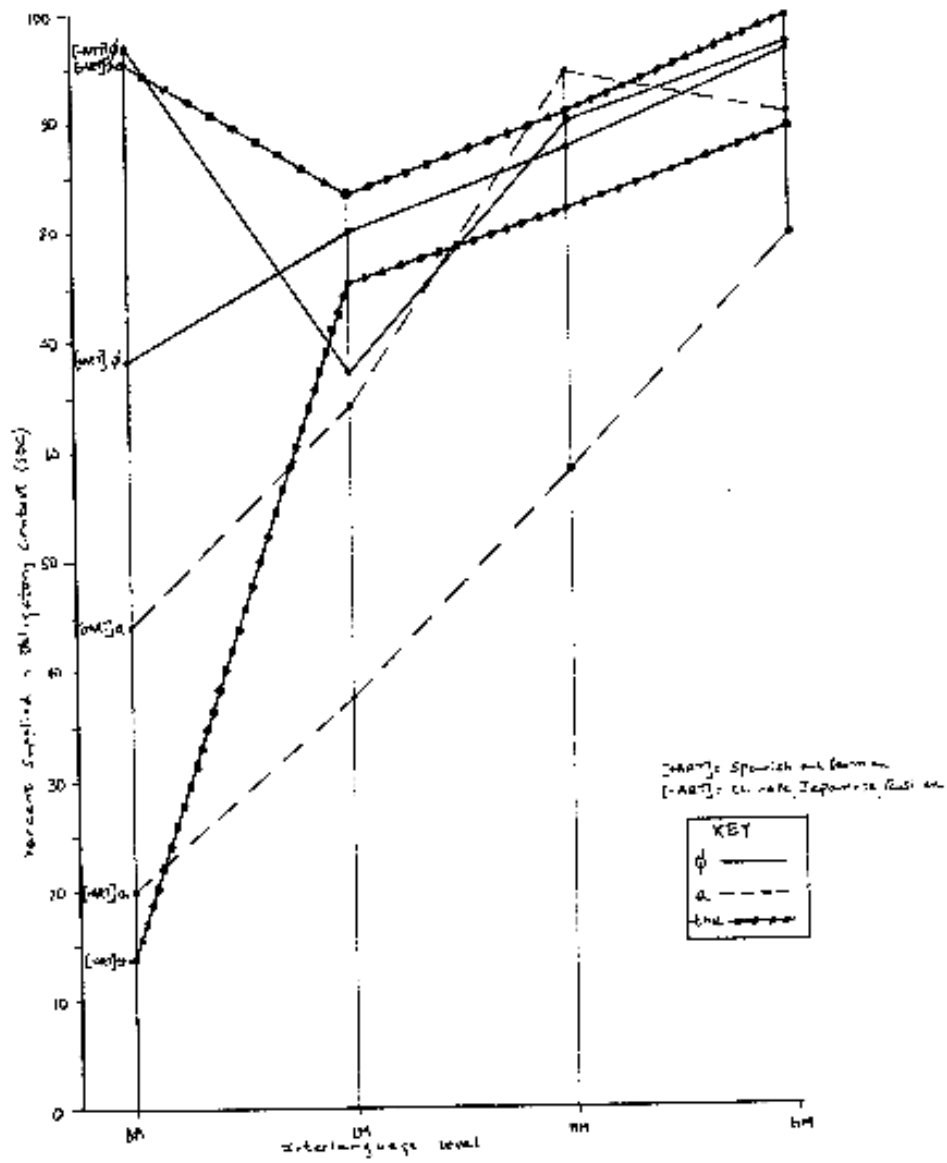
Figure 6-3 Overall Article Accuracy

30

3. the [-ART] acquisition sequence appears to be Ø > *the* >
      a (which, if difficulty reflects acquisition order,
      agrees with the findings of Yamada et al. (1982) for
      advanced Japanese ESL students), whereas the [+ART]
      acquisition sequence appears to be *the/0* > *a* (the
      three are actually so close that there is no readily
      discernible acquisition sequence; however, *a-*
      acquisition clearly lags behind Ø- and *the-*
      acquisition)

Accuracy in Specific Areas
      Accuracy (SOC) was measured not only for overall article
usage but also in terms of Bickerton's four categories.
Although this was undertaken primarily in the hope of finding a
single element of article usage that might function as an
interlanguage measure (to be pursued in Chapter 9), these
details provide a more precise idea of specific areas within the
article system that contribute to the overall accuracy figures.

Accuracy in Bickerton's Four Categories
      The accuracy data are first examined within each of
Bickerton's four major categories. Category I concerns the use
of Ø, *a*, and *the* with generic noun phrases. Category II
concerns the use of the definite article *the* with specific noun
phrases. Category III concerns the use of the indefinite
articles Ø and *a* with specific nouns. Category IV concerns the
use of the indefinite articles Ø and *a* with ambiguously generic
noun phrases in equative, negated, interrogative, and irrealis
contexts. Figure 6-4 shows the combined accuracy of Ø, *a*, and
the within the four categories for each of the five language
groups. The three leftmost elements of each column depict the
subjects whose first languages belong to the [-ART] group) while
the two rightmost elements of each column depict the subjects
whose first languages belong to the [+ART] group).
      Overall, the bar graphs in Figure 6-4 show a general
tendency to increased accuracy in all four categories. In terms
of increasing accuracy with increasing IL level, the least
consistent category appears to be Category I. This probably
reflects the fact that generic usage, particularly in the
sometimes sparse language of the basilang (BA) subjects, is
often rather difficult to ascertain as generic usage can only be
determined by context. For example, Ah Chun (BA Chinese) makes
the statement:

      74-5 Houseekeepah workee har'

This was interpreted to mean, "A housekeeper has to work hard"
and classified as 1b (generic *a*). However, it could possibly

31

Fig. 6-4 Accuracy in Categories I–IV
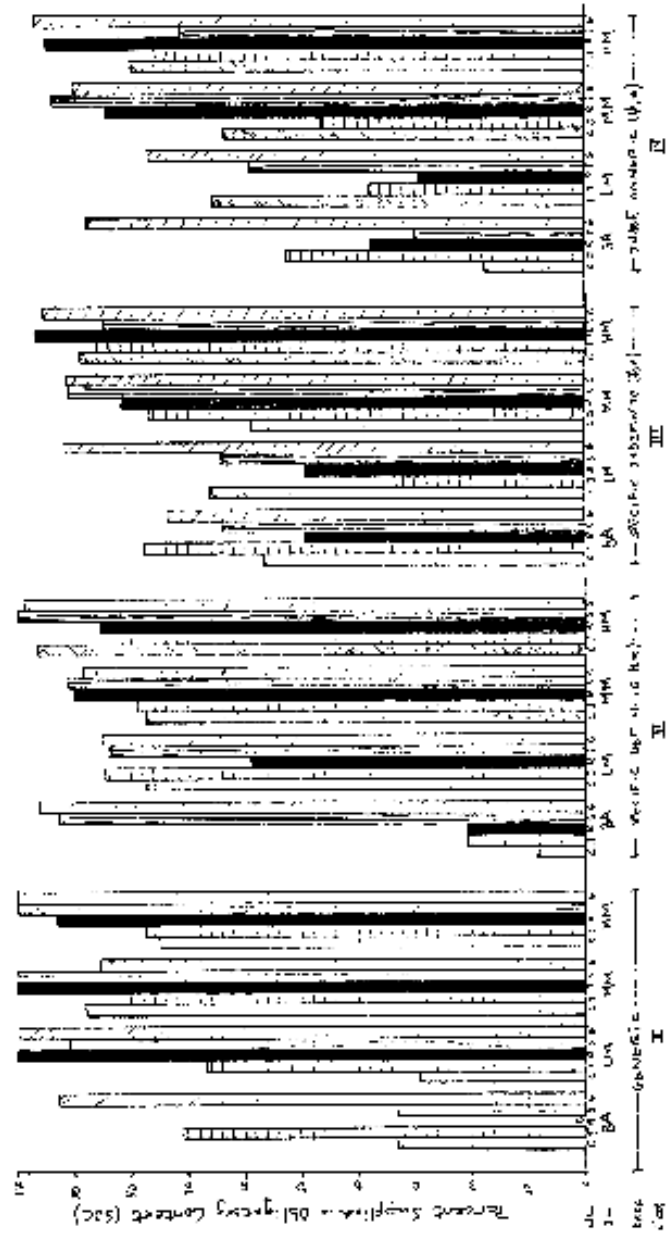
have been interpreted as, "I, the housekeeper, work hard", which
would have been classified as 2e (implied postmodification).
Perhaps the inconsistent patterns of generic usage reflect this
difficulty of classification, especially at the lower
interlanguage (IL) levels. Furthermore, the informal day-to-day
nature of topics discussed offered comparatively few
opportunities for generic usage. These tended to increase with
IL level, the high mesolang (HM) subjects producing the highest
numbers of clear generic phrases.

    To the extent that Kellerman's (1984) U-shaped learning
curve can be applied to a pseudolongitudinal study, there are
several U-shaped trajectories in Figure 6-4, especially for the
Spanish and German subjects in Category II, the Japanese
subjects in Category II, and the Russian subjects in the
equative classification of Category IV. This characteristic
will become clearer in later figures.

    As far as "the most difficult area" can be judged, Category
IV shows the lowest overall accuracy. This is probably due to
problems with the [±count] feature. However, since this category
also includes negation with the indefinite article, it is an
interesting result in light of Klima and Bellugi-Klima's (1966)
finding that negative indefinites are still not established by
the third stage of child language development. It suggests that
adults learning a second language also find it difficult to
establish this relationship, and that L1 and L2 pose similar
difficulties for their respective learners in certain respects.

    To provide a more detailed picture of the accuracy data, it
is next analyzed in terms of the 19 categories into which
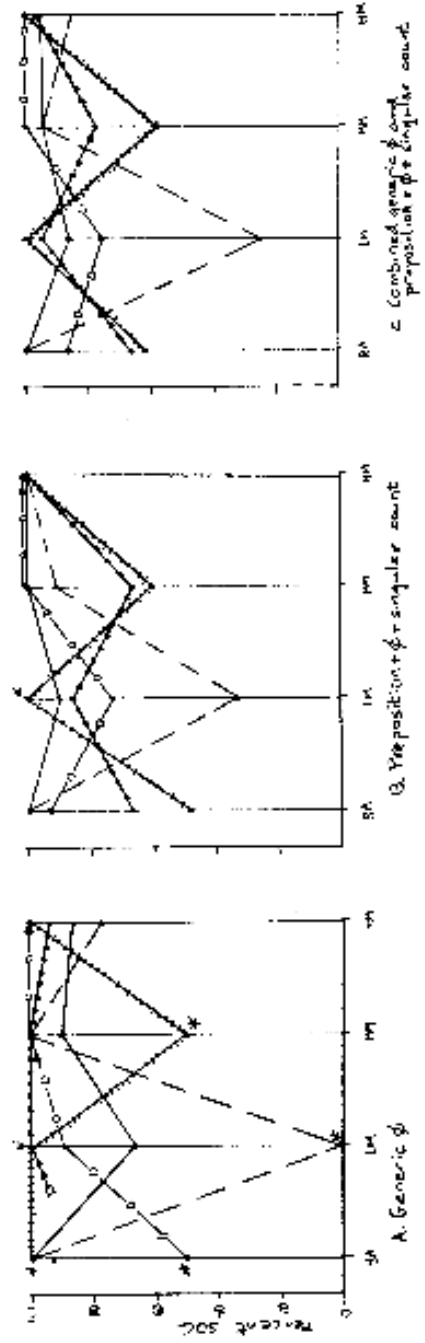Bickerton's four categories were divided for the purposes of
this study.

Accuracy in the Classifications within Category I
    Category I was tallied in terms of the individual articles
the, *a*, and Ø. The total number of environments requiring
generic *the* in the entire corpus of 2O subjects was 22/7236.
The total number of environments for generic *a* was 25/7236.
This produces averages of just over one instance per subject,
which is much too few to be able to draw any conclusions. For
this reason, only generic Ø was considered here. Generic Ø was
tallied for 1) generic Ø environments and 2) prepositions
followed by Ø plus a singular countable noun. Figure 6-5
contains three graphs showing the accuracy of the five language
groups in:

        1. generic Ø
        2. preposition + Ø + singular countable noun
        3. combined generic Ø and preposition + Ø + singular
                 countable noun

Fig. 6-5 Accuracy in Subclassifications of Category 1 (φ only): Generic

KEY
CH
JA
Ru
SP
GE
* < 5 indexed
++ 1 indexes
□ no instances

A. Generic φ

b. Preposition φ + singular count

c. Combined generic φ and preposition φ + singular count

The generic Ø classification (Graph A) in Figure 6-5 has several starred coordinates indicating fewer than five instances, particularly the lowest accuracy coordinates for the Japanese, Spanish, and German subjects. These latter three are based on only two instances each. The extreme Japanese LM trough can be explained by the extensive *the*-flooding this subject resorted to, using *the* even in environments outside noun phrase morphology. For example,

> Yoko 3:2   Then the she applied the the radio company. Then
>   the she had a very funny examination over there;
>   uh the person com- coming to see her, then the
>   uh, "Please wait here." Then the never come back
>   the the person ask her.

The [-ART] group is not so accurate with generic Ø as it will later be with other uses of Ø. Nevertheless, overall accuracy is considerably higher for Ø than for *a* and *the* by this group. In the [+ART] group, the subjects move gradually to full accuracy at the MM level in a manner analogous to [-ART] subsequent mention *the*-accuracy, to be discussed in the next section.

The preposition + Ø + singular count noun classification (Graph B) differs from Graph A. In the [-ART] group, the Chinese subjects show the same pattern with greater accuracy, the Japanese subjects show a similar pattern with slightly lower accuracy, and the Russian pattern is quite different and reflects lower accuracy. In the [+ART] group, the Spanish subjects show the same pattern with the exception of the BA subject, who evinces much greater accuracy in Graph B than in Graph A.

The combined generic Ø and preposition + Ø + singular count noun classification (Graph C) shows a pattern almost identical to Graph B, showing the comparatively weak influence of generic Ø (Graph A) on the combination due to the relative infrequency of generic Ø in the corpus.

Accuracy in the Classifications within Category II

One of the most dramatic changes in Figure 6-4 is the [-ART] group's marked increase in accuracy in Category II between the BA and LM stages, especially in contrast with the [+ART] group's performance. For this reason, the six classifications within Category II are looked at in detail. Figure 6-6 contains six graphs, showing the accuracy of the five language groups in:
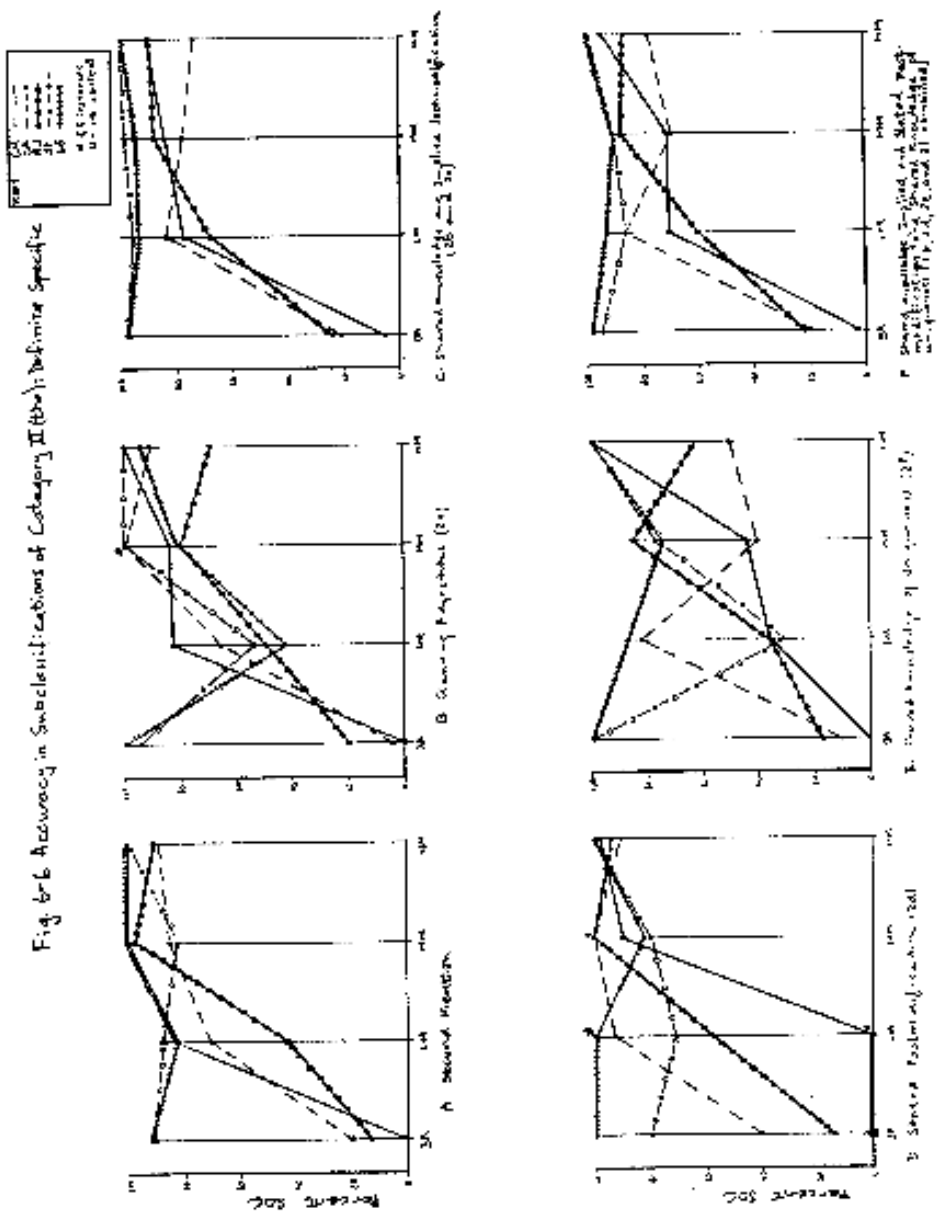
Fig 6.6 Accuracy in Subclassifications of Category II (b): Definite Specific

1. subsequent mention
2. ranking adjectives
3. shared knowledge and implied postmodification
4. stated postmodification
5. shared knowledge of uniqueness
6. shared knowledge, implied and stated postmod-
   ification, and shared knowledge of uniqueness

Shared knowledge and implied postmodification are considered
together in the third graph because in the process of coding the
data it became quite obvious that these two groups overlapped.
In fact, all shared knowledge noun phrases can be considered to
be cases of postmodification. For example, *the moon*, a clear
example of shared knowledge in that any human being would
presumably know which moon was meant (and would never ask "Which
moon?") could be rendered as *the moon which inhabits the night
sky* or *the moon that orbits the earth*. Indeed, the only
distinction between a proper noun and a shared knowledge/implied
postmodification noun phrase is that the former requires no such
postmodification, e.g., *Everyone on Earth has the same DNA* vs.
Everyone on the planet has the same DNA.
    Shared knowledge of uniqueness is also a type of implied
postmodification. A typical example of this classification is
the Russian people, i.e., the unique group that is identified by
this phrase. It would obviously be redundant to add *who live in
Russia*, although we understand that this is meant. The noun
phrase *the government* caused some difficulty in coding, as it
seemed to reflect shared knowledge, implied postmodification
(e.g., *the government of Iran*), and shared knowledge of
uniqueness (there is usually only one government in any given
country at any one time). For consistency's sake, *the
government* was always classified as implied postmodification in
this study. Nevertheless, because of potential miscodings, the
four classifications of shared knowledge, implied and stated
postmodification, and shared knowledge of uniqueness are also
collapsed into a single classification in the sixth graph to
provide a composite picture.
    The second mention classification (Graph A) for the [-ART]
subjects is the only one that appears to increase consistently
with IL level. The accuracy scores for the Russian MM and HM
subjects in this category both contain a single error, the
difference in number of instances causing the HM subject to have
a lower accuracy than the MM one. The high accuracy for the
[+ART] subjects reflects the presence of this function in their
native languages.
    The ranking adjective classification (Graph B) shows
consistent improvement to the MM level for the [-ART] group.
This improvement persists, however, only in the Chinese subject,

the Japanese and Russian subjects falling below the MM level.
In the [+ART] group, the Spanish subject shows a considerable
drop at the LM level, suggesting that this notion is not always
present in Spanish. It must be relearned (or adjusted, since
the words meaning *same*, *only*, and other ranking adjectives
requiring the article *the* in English require the equivalent to
the in Spanish); however, this appears to happen by the next IL
level, though the fact that there were fewer than five instances
(indicated by the asterisk) undermines its significance.

   The combined shared knowledge and implied postmodification
classifications (Graph C) shows sharply improved performance
between the BA and LM levels for the [-ART] group. While the
Japanese and Russian subjects continue to greater accuracy at
the MM and HM levels, the Chinese subjects continue to drop
consistently in accuracy at these levels for no apparent reason.
The [+ART] group hovers near the 100%-accuracy line for all IL
levels, suggesting that this feature is present in Spanish and
German.

   The stated postmodification classification (Graph D) shows
the familiar sharp increase for the [-ART] group, although in
this case, the Chinese subjects exhibit a delay in acquiring
this feature. However, the BA Chinese subject had no instances
of stated postmodification and the LM Chinese subject had 0/2
instances correct, well below the minimum of 5 instances
generally required in morpheme acquisition studies. In other
words, this "delay" may simply have been the result of too few
instances.

   The shared knowledge of uniqueness classification (Graph E)
shows the greatest difference from the other five graphs in
Figure 6-6. In the [-ART] group, the Russian subjects show the
greatest consistency with respect to their performance in other
aspects of Category II accuracy. The Chinese subjects show the
slowest improvement of any subarea in Category II, suggesting
that this feature is very difficult for them to acquire. The
Japanese LM subject shows the highest accuracy for any of the LM
subjects in this category. This subject, however, is the "*the*-
flooder" of the [-ART] group, and the strategy serves her
especially well in this instance.

   The combined shared knowledge, implied and stated
postmodification and shared knowledge of uniqueness
classification (Graph F) shows for the [-ART] group the
characteristic meteoric rise in accuracy between the BA and LM
levels, although this is once again most evident for the
Japanese LM subject (the *the*-flooder) and least for the Russian
subjects, which show a characteristic gradual rise to the MM
level. The [+ART] group shows a characteristic "U", with the
trough at the LM level.

   As stated at the beginning of this section, it is only the

second mention classification (Graph A) that shows increasing
accuracy for the [-ART] group, the exception noted for the
Russian HM subject notwithstanding. This distinction would have
been lost if Bickerton's Category II had only been analyzed as a
single measure. Similarly, the apparent learner confusion over
shared knowledge of uniqueness (Graph E) would also have been
lost.

Accuracy in the Classifications within Category III
   Accuracy in Category III (indefinite specific or first
mention), as depicted in Figure 6-1, shows a sharp increase
between the LM and MM levels. Like Category II in the last
section, Category III also masks several features in the single
measure. For this reason, the four classifications within
Category III were also looked at in detail. Figure 6-7 contains
six graphs, showing the accuracy of the five language groups in:

     1. count singular (*a*)
     2. count plural (Ø)
     3. noncount [i.e., mass] (Ø)
     4. *a = one*
     5. combined count singular *a* and *a = one*
     6. combined count plural (Ø) and noncount (Ø)

Count singular and *a = one* are considered together in the fifth
graph to show the overall accuracy of *a* in Category III.
Similarly, plural and noncount are considered together in the
sixth graph to show the overall accuracy of Ø in Category III.
     The count singular classification (Graph A) shows a varied
pattern of accuracy except at the HM level. For the first time,
the members of the [+ART] group show a distinct difference at
the BA level, the German BA subject achieving an accuracy of 83%
whereas the Spanish BA speaker shows an accuracy of 2% (one
instance correct). However, the Spanish BA speaker used *one* in
place of *a* in 17 instances, and if this is considered, the
accuracy rises to 41%, reducing the observed distance between
the hitherto similarly-performing groups. Graph A, as indeed
all the graphs in Figure 6-7, is otherwise inconsistent.   The
Japanese and Russian LM subjects show troughs whereas the
Japanese LM subject shows a peak. At the MM level, the Chinese
and Japanese cluster, as do the Spanish and German subjects, but
the Russian subject lies midway between. The reason for this
lack of a pattern of usage probably stems from individual
difficulties with the [±count] system, which will be discussed in
Chapter 7.
     The count plural classification (Graph B) shows the
advantage that the [-ART] group has in having no article system:

Fig. 6-9 Accuracy in Subclassifications of Category III (φ and): Indefinite Specific
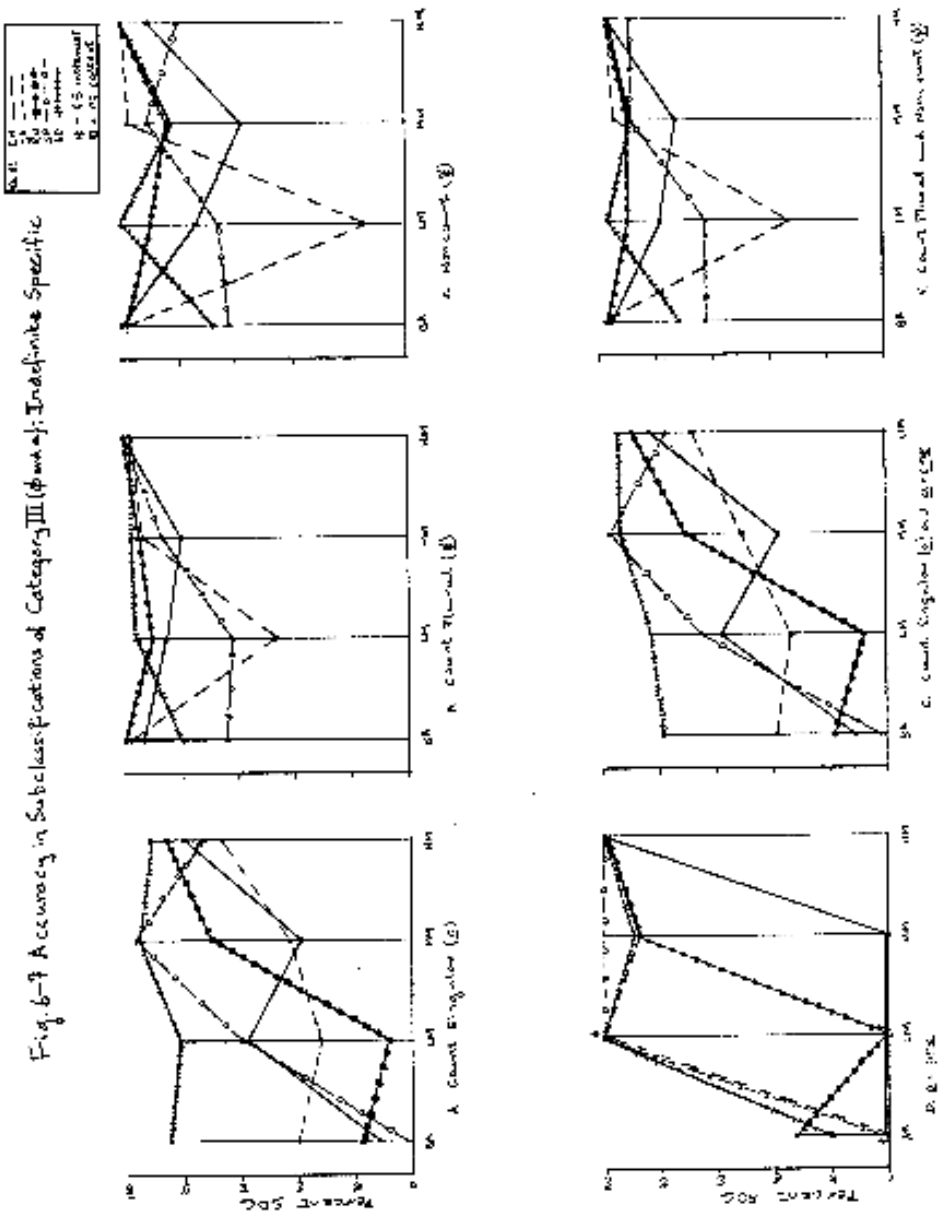
Ø-accuracy is very high at the BA level, just as was *the*-accuracy for the [+ART] group in Category II (Fig. 6-6). Ø does occur, of course, in [+ART] languages, but it must be meted out in distribution with *a*. The deep trough of the Japanese LM subject is a consequence of her extreme *the*-flooding; she used *the* in so many contexts that her Ø-accuracy suffered as a result. In general, the Ø-accuracy of the [-ART] group parallels the *the*-accuracy of the [+ART] group in Category II, generating a shallow U-shaped curve. Ø-accuracy for the [+ART] group looks more like *the*-accuracy for the [-ART] group, except that BA-accuracy is initially higher for the former.

The noncount classification (Graph C) parallels the count plural classification (Graph B) in practically every respect (a suggestion that article form, at least in same cases, has more significance than article function, i.e., Ø is used in the same pattern in both cases irrespective of the functional differences between plural and mass nouns). The overall difference is that accuracy is somewhat lower and the troughs of the [-ART] group are all deeper. The one noticeable individual difference is that the HM Spanish speaker showed higher accuracy than the MM Spanish subject for count plural (Graph B), whereas the same Spanish subject shows lower accuracy than the MM Spanish subject for noncount Ø. This suggests that accurately determining the [-count] feature is more difficult than pluralizing a [+count] noun.

The *a = one* classification (Graph D) is too extreme to be of much use. This stems from the low numbers of instances for all subjects-- in fact, only seven of the 20 subjects had more than five instances. The Russian subjects' performance parallels their performance with singular count *a*, but this is not true for any of the other subjects.

The combined count singular (Graph A) and *a = one* (Graph D) classifications (Graph E) produce a graph almost identical to Graph A, showing the weak impact of the lower figures for *a = one*.

The combined count plural and noncount classifications (Graph F) produce a parallel pattern to both Graph B and Graph C and offers no new insights other than support for form over function.

Accuracy in the Classifications within Category IV
Accuracy in Category IV, as depicted in Figure 6-8, shows the sharpest increase between LM and MM, as in Category III. Category IV was tallied in four subcategories: equative, negated, interrogative, and irrealis mode. In the entire corpus of 7236 noun phrases, only three were interrogative and twelve
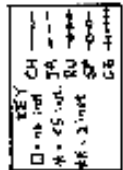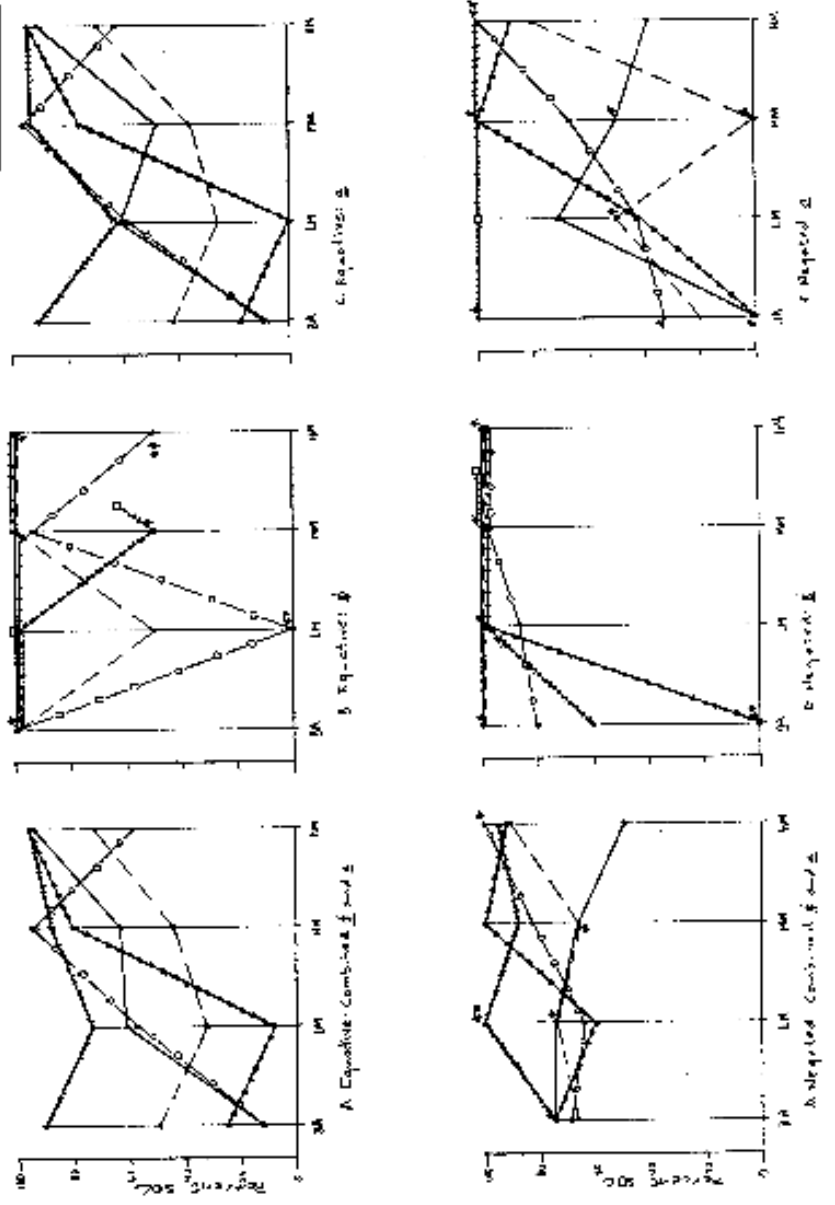
41

Fig. 6-8 Accuracy in Subclassifications of Category IV (Manual); Indefinite Generic

were irrealis mode.[2]   These latter categories were therefore
ignored.   The equative and negative noun phrases, which require
either Ø or a, were looked at in detail. Figure 6-8 contains
six graphs, showing the accuracy of the five language groups in:

1. combined equative Ø and *a*
2. equative Ø
3. equative *a*
4. combined negated Ø and *a*
5. negated Ø
6. negated *a*

The number of equative noun phrases in the corpus was quite
small, with the result that several coordinates on Graph D are
starred (i.e., fewer than five instances occurred). The number
of negated noun phrases was even smaller (which is a pity really
because they might have shown some linkage to Cazden's et al.
negation criteria), with the result that the majority of
coordinates in Graph E are starred, producing a disappointingly
inconclusive picture of negated Ø-usage. The same applies to
negated *a*, although Graph F contains fewer starred coordinates.

The combined equative Ø and *a* classification (Graph A)
reveals a pattern very close to that of Graph A in Category III
(Fig. 6-7). The marked difference between the [+ART] BA
subjects occurs again, which is exclusively due to *a*-usage as
Graph C will show.

The equative Ø classification (Graph B) has too many
starred coordinates to be conclusive. The high Ø-accuracy for
the [-ART] subjects has been seen before. The plunging Ø-
accuracy for the Japanese LM subject is once again the product
of her excessive *the*-flooding. The plunging Ø-accuracy of the
Spanish LM subject is the product of a single instance and
should be discounted. The HM German subject had no occurrence
of equative Ø, the HM Spanish subject also only an inconclusive
single instance. Graph B in itself is only useful for what it
subtracts from Graph A to throw the accuracy of *a* into higher
relief.

The equative *a* classification is almost identical to Graph
A in Category III (Fig. 6-7) across all five languages. The two
graphs are both concerned with *a*-accuracy, suggesting that some
aspects of accuracy might be better analyzed in terms of overall
article usage (with common nouns), an approach that will be
pursued in the next section of this chapter.

_____

[2]A noun phrase was only counted as irrealis mode when it
reflected an unreal conditional sense. Noun phrases in the real
conditional sense (e.g., *I read books if I have time*) were
counted as specific indefinites (Category III).

The combined negated Ø and *a* classification (Graph D) shows a fair amount of consistency across all five languages with accuracy not really increasing until the MM level and beyond. Four starred coordinates somewhat weaken the picture.

The negated Ø classification simply has too few instances (153 in the entire corpus, an average of 7.7 per subject) to be very conclusive. Accuracy is apparently higher for all groups at the LM level and beyond. The BA [+ART] subjects are typically less accurate for Ø-usage than the [-ART] group, even though the BA Chinese subject has no (O/1) correct usages.

The negated *a* classification (Graph F) is quite different from Graph C (equative *a*). In general, accuracy is higher for all subjects except the starred MM Japanese and the HM Chinese subjects, both of whom show an accuracy over 3O percentage points lower than their parallel performance with equative *a*.

Conclusions

From the data in the last section, the following conclusions can be drawn:

1. Category I: Generics are difficult to identify, especially at the BA and LM level. Generic *a* and *the* occur too infrequently to measure. There is no consistent pattern of increasing Ø-article accuracy across IL levels except for the Spanish speakers.

2. Category II: The [+ART] group has a high degree of accuracy in the second mention, shared knowledge, and postmodification categories. They both show marked weakness in the ranking adjective category and to a lesser extent in the shared knowledge of uniqueness category, the weakness most evident at the all-important (because the most dramatic changes occur here) LM and MM levels.

    The [-ART] group shows a characteristic rapid increase in the accuracy between LM and MM for all categories. Shared knowledge of uniqueness appears to be the most difficult aspect for this group, perhaps because such phrases are comparatively rare in English.

3. Category III: The [-ART] group shows a high degree of accuracy with Ø, with development always in the characteristic U-shape. *A*-accuracy improves at a slower rate than Ø-accuracy. Ø + noncount nouns appear to cause more difficulty than Ø + plural nouns for both [-ART] and [+ART] groups.

The [+ART] group is less accurate with Ø at the
early stages. *A*-accuracy is high for the German
subjects; for the Spanish subjects, it increases in
much the same manner as the [-ART] group's Cat. II-
accuracy.

4. Category IV: Equative *a*-accuracy for both [-ART] and
    [+ART] is identical to that in Cat. III, showing that
    a is used in the same way in both categories (further
    evidence for *a* as part of a different system).
    Negated *a*, however, generally shows delayed accuracy
    for all groups (except German, which has only starred
    coordinates), suggesting Klima's et al. (1966) delayed
    acquisition of negative and indefinite in L1
    acquisition.

USAGE
     Usage is investigated first within Bickerton's framework.
It is then examined in terms of overall article usage without
regard to Bickerton's categories. Finally, usage is analyzed in
terms of the native languages of the 20 subjects.

Usage within Bickerton's Four Categories
     The usage data are examined in terms of Bickerton's four
major categories or environments: generic, specific definite,
specific indefinite, and indefinite generic. Since they were
originally designed for interpreting usage rather than accuracy,
the results will no doubt be more useful than the same
categories were found to be for describing the accuracy data.
Figure 6-9 shows usage within the four categories by language
group. The three leftmost columns in each square depict the
usage of the subjects whose native language belongs to the
[-ART] group (Chinese, Japanese, and Russian). The two
rightmost columns in each square depict the usage of the
subjects whose native language belongs to the [+ART] group
(Spanish and German). The usage figures that determine the
height of the columns in Figure 6-9 are calculated by dividing
the number of occurrences of each article by the number of
common noun phrases within each category. In order to conform
to Bickerton's and Huebner's constraints (and thereby to allow
certain aspects of this data analysis to be compared with
Huebner's (1983b) longitudinal study), two of the classifi-
cations considered in the accuracy section are not counted here:
code 1d (preposition + Ø + singular count NP) and code 3d (*a* =
one).
     In Category I (generic *the*, Ø, and *a*), it is clear from
Figure 6-9 that Ø is the predominant article used by all
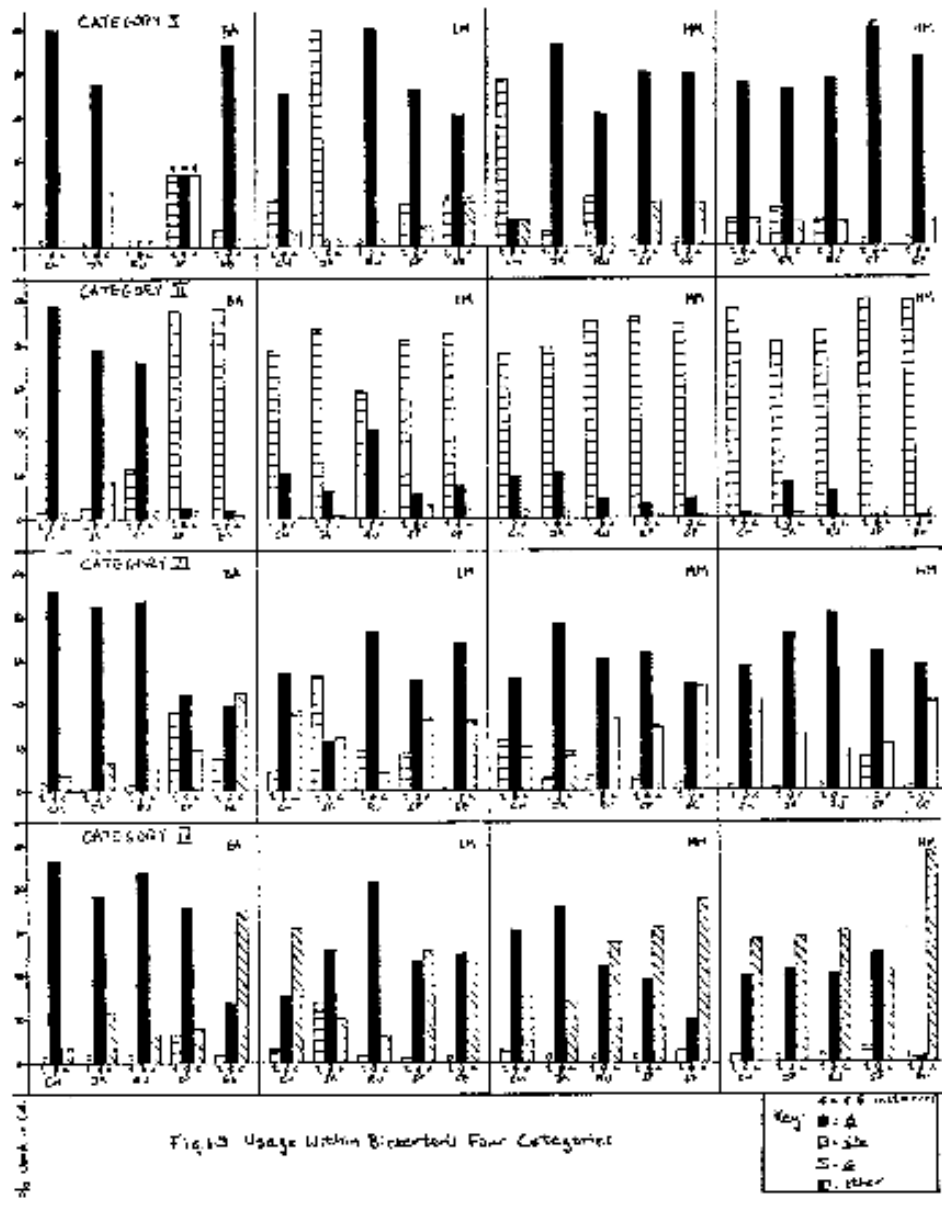language groups at all IL levels. *A* is used more at the MM and

45

Fig.13 Usage Within Bickerton's Four Categories

Key:
■ A
□ Able
▨ S-G
▤ Other

HM levels than at the BA and LM levels. Extreme generic *the* usage by the LM Japanese and the MM Chinese subjects probably reflects *the*-flooding at these stages. However, Category I differs from Categories II to IV in that, despite certain restrictions on their use, all three articles are allowed in this category. In Category II (specific definite *the*), the [-ART] group uses incorrect Ø increasingly less as IL level increases, with the most dramatic change occurring between the BA and LM levels. Simultaneously, *the*-usage by the [-ART] group increases dramatically at the same point. The (incorrect) usage of *a* occurs infrequently for all language groups, although the lower IL levels used it slightly more than the upper IL levels. [+ART] *the*-usage is high at the BA level, drops to a slight trough at the LM level, and then increases at the MM level to nearly 1OO%-usage at the HM level. This usage parallels *the*-accuracy (Fig. 6-6) very closely for this group.

In Category III (specific indefinite Ø and *a*), Ø-usage once again dominates across all five language groups. For the [-ART] group, *a*-usage increases with IL level whereas Ø-usage drops considerably at the LM level and then climbs again at the MM and HM levels. For the [+ART] group, *the*-usage (incorrect) is highest at the BA level but diminishes fairly consistently with increasing IL level.

In Category IV, (indefinite generic Ø and *a*), Ø-usage appears to decline with increasing IL level just as *a*-usage increases for all five language groups (with the exception of the BA German subject, who shows high *a*-usage even at the BA level). [-ART] *the*-usage (incorrect) flares a little at the LM level (once again reflecting *the*-flooding), but subsides by the MM level.

Highlights of Usage within Categories II and III
Usage in Categories II and III is broken down somewhat in this section because these categories embrace several functions and because the majority of noun phrases fall within these two categories.

Usage in Category II
It is apparent from Fig. 6-9 that *a*-usage plays a comparatively small role in this category. For this reason, Ø- and *the*-usage within Category II was plotted for all five language groups (Fig. 6-1O) without including *a*. Figure 6-1O shows that *the*-usage within Category II forms a virtual mirror image of Ø-usage. For the [+ART] group, *the*-usage begins near 1OO% at the BA level, drops to about 8O% at the LM level, and then returns to 1OO%-usage at the HM level. Ø-usage by the [+ART] group, on the other hand, starts near zero at the BA level, increases to about 1O% at the LM level, and then

47

Fig 6-10   Usage in Category II : φ (incorrect) vs. the (correct)
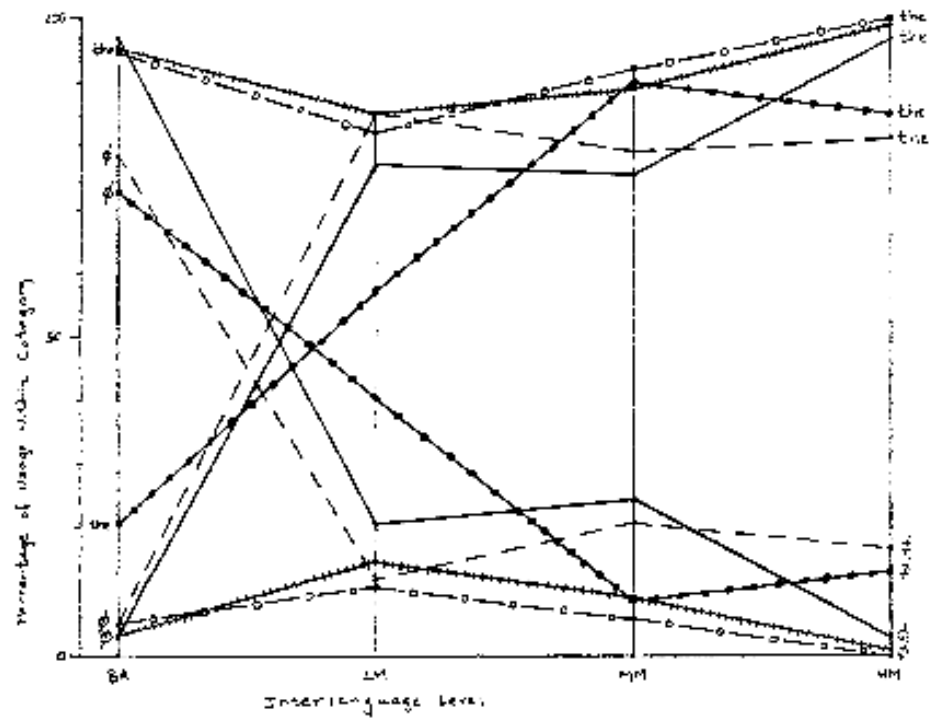
decreases gradually to zero usage at the HM level. For the
[-ART] group, *the*-usage starts low at the BA level, increases
dramatically for the Chinese and Japanese subjects to a peak at
the LM level, more gradually for the Russian subject to a peak
at the MM level, and then declines slightly to the HM level.
This dramatic rise in usage is what Huebner calls flooding.
Simultaneously, Ø-usage starts fairly high, drops dramatically
in a mirror image of the increase, and settles down to low usage
at the HM level. One implication of this apparent "replacement"
by the [-ART] group is that *a*-usage reflects the development of
a different system, namely the [±count] feature.

Usage in Category III
        Since all three articles play a role in Category III usage,
they are plotted for all five language groups in Figure 6-11.
The most obvious feature of Figure 6-11 is that Ø is used most
frequently by all five language groups whereas *a* and *the* compete
with each other at the bottom. *The*-usage is incorrect in this
category, and the [-ART] group appears to be testing the
hypothesis that *the* might work in this category as it did in
Category II. Peaks of *the*-usage occur at the LM level for the
Japanese and Russian subjects, at MM for the Chinese subject,
before settling back to near zero usage at the HM level. The
[+ART] group evinces no such "hypothesis testing" peak, but
rather assumes at the outset that *the* is correct in this
instance. They then appear to chip away at this hypothesis (in
a manner analogous to Huebner's "trickling") as they steadily
decrease to zero usage at the HM level. For both groups, *a*-
usage appears to generate no discernible pattern other than a
general tendency to an increase in usage with IL level. This is
most probably linked to mastery of the [±count] distinction
mentioned earlier and will be discussed in greater detail in
Chapter 7.

Overall Article Usage

        Article usage is examined in this section without regard to
the categories discussed in the preceding section. This will be
done in two ways: a) the percentage of article usage in the
entire corpus of common nouns for each language group and b) the
percentage of article overuse for each language group.

Flooding Beyond the Category (Usage/Total Noun Phrases)
        This section expands Huebner's notion of flooding to the
whole article system rather than just to the category in which
it is correct. The percentage of *the*-usage in the entire
corpus, i.e., the used/total NP's, is plotted in Figure 6-12.
Figure 6-12 shows that flooding as a function of the entire

Fig. 6-11 Usage in Category III : φ and a (correct) vs. the (incorrect)
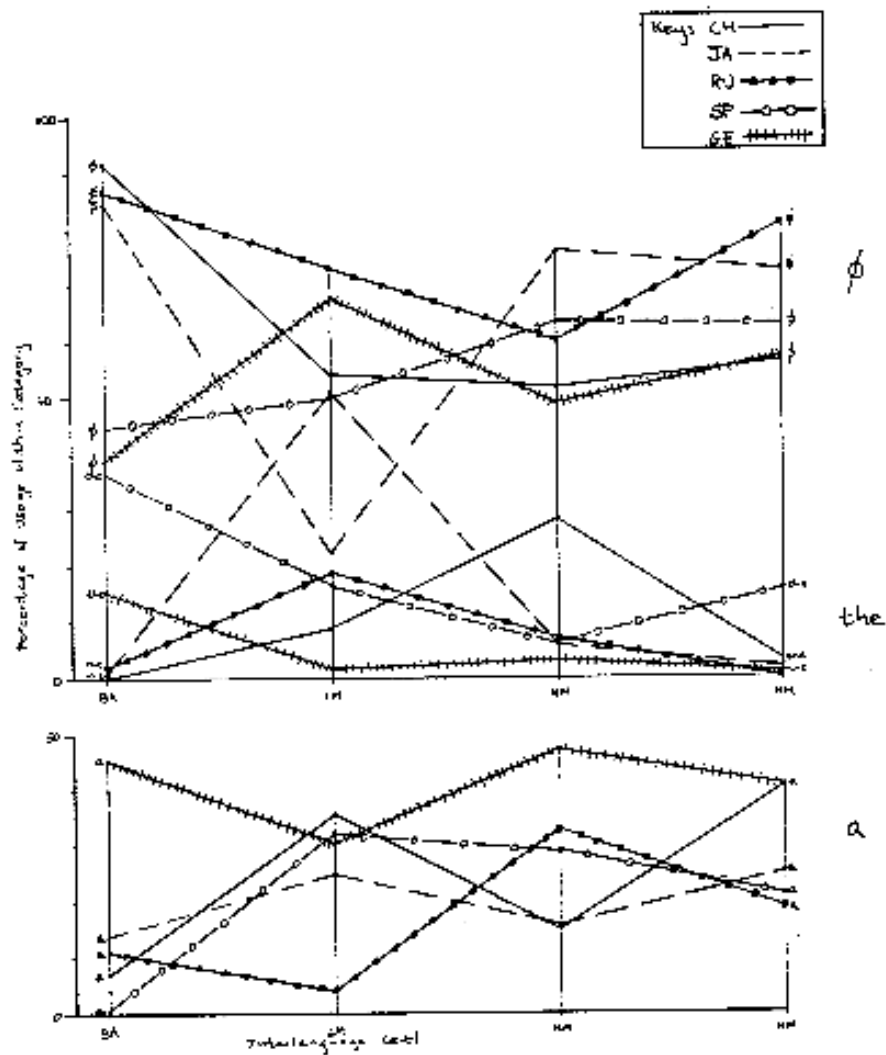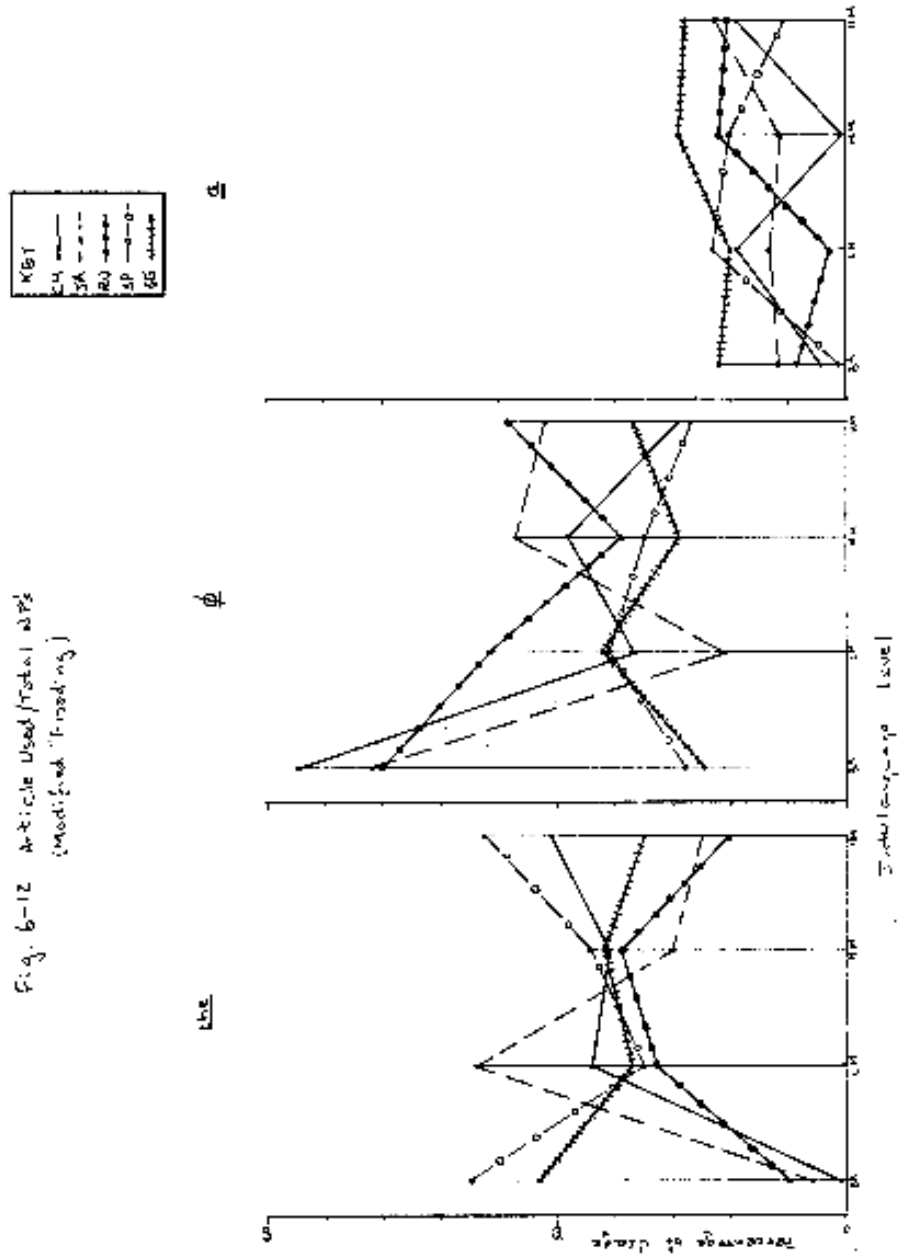
Fig. 6-12 Article Used (Total NP's)
(Modified Threading)

51

corpus of common nouns is really only exhibited by the Japanese subject at the LM level. The Chinese subject shows a peak of *the*-usage at the LM level, the Russian at the MM level. Whether this type of flooding is a universal or an individual strategy for [-ART] languages cannot be determined from this study, but if it does occur, it appears to do so between the LM and MM levels, especially as the Japanese subject was at the upper end of the LM level as indicated by her negation. The [+ART] group shows a trough at the LM level, recalling the mirror images of Fig. 6-10.

The same formula applied to Ø-usage (i.e., Ø used/total NP's) shows troughs corresponding to the peaks of [-ART] usage. The [+ART] group shows corresponding peaks of Ø-usage where *the* had formed a trough.

The same formula applied to *a*-usage (i.e., *a* used/total NP's) produces no clear pattern except, as before, a trend to increased usage with increasing IL level. *A*-usage remains below the 30% level for all five language groups, leaving the "battle" to lie once again between Ø and *the*.

Overusage

The above formula indicates what percentage an article is used in the entire corpus of a single subject. But this does not take into account the [±count] factors for Ø and *a*, nor the factor of choice (e.g., *I ate* cake vs. *I ate* a cake). To capture the subjects choice of article, a formula was devised to get at the notion of "flooding" in another way. This formula counts the number of times an article was overused (i.e., used when it should not have been according to the English target) and divides that number by the total number of usages of that article, i.e., when the subject chose to use it. For example, the Japanese BA subject used *the* only once when she should not have in her total of 20 uses of *the* with common nouns. The coordinate for this subject is thus 1/20 = 5%. The results are plotted in Figure 6-13. Figure 6-13 shows that *the*-overuse peaks at the LM level for the Japanese subjects, at the MM level for the Chinese subjects. The Russian, Spanish, and German subjects show no such peaks of overuse, gradually (and correctly) reducing this tendency with increasing IL level. This is the first evidence of the Russian subjects behaving like their Indo-European cousins and not like the other [-ART] subjects.

Ø-overuse shows slight peaks for the Chinese and Japanese subjects at the LM level, but the Russian, Spanish and German subjects once again show only increasing reduction of such overuse with higher IL level.

A-overuse shows marked peaks for, curiously, the Japanese and Spanish subjects, but not for the Russian, Chinese, and

Fig. 6.13 Overusage of A/A cltes
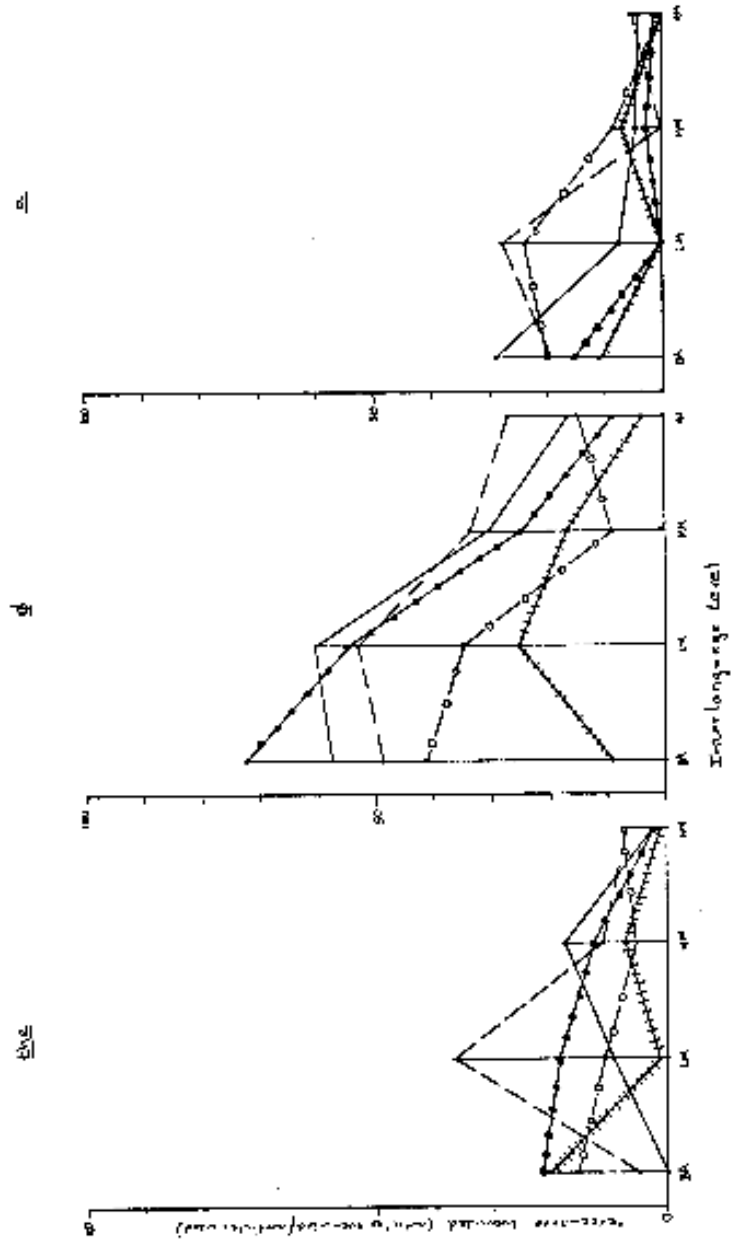
German subjects, suggesting that the former language groups have the most trouble with *a*, i.e., trouble with the [±count] feature.

Overall Article Usage for the Five Language Groups

In order to "paint a picture" of article acquisition, it is necessary to show which of the three articles is being used, how often, and with what degree of target-like use. Ideally, usage will approach 1OO%-accuracy as linguistic competence increases. In other words, let's say a native speaker of English uses *the* five times. This also reflects five instances when *the* is required (five obligatory contexts), and if the five usages are divided by the five obligatory contexts, the result (5/5) is one, or a 1OO% linkage between used and required articles. This used/required formula, which I call UOC-- used in obligatory context-- to parallel the well known SOC (supplied in obligatory context) is applied to the three articles for each language group by itself, and the results are plotted in Figure 6-14. Figure 6-14 shows that the development of article usage (UOC) creates a kind of funnel, wide at the BA level and increasingly narrow at the HM level as it approaches the 1OO%-accuracy line. The [-ART] group shows a wide funnel, the [+ART] group a narrow one as the latter clusters more consistently closer to the 1OO%-accuracy line across all IL levels. The top (widest part) of the funnel for the [-ART] group is defined by *a/the* and Ø, as Ø-usage gradually decreases and *a*-usage increases. In the [+ART] group, the Spanish subjects' *a*-usage forms the lower side of the funnel although it approaches the 1OO% line more rapidly than the [-ART] group. However, the German subjects' narrow funnel finds Ø- and th-usage making up the sides. This difference is explained by the fact that the Spanish subjects appear to treat *a* as a newly- acquired item, analogous to the [-ART] acquisition of *a* and *the*, whereas the German subjects treat *a* as an adjusted L1 item, analogous to the [-ART] acquisition of Ø.

Overall Article Usage: [+ART] versus [-ART]

Overall UOC (usage/required) data for the [+ART] group versus the [-ART] group is shown in Figure 6-15. Combining the two groups in this way is motivated by the fact that the different language groups show similar patterns of development in Fig. 6-14. Again, the broad-based funnel shape of the [-ART] group is readily apparent, with Ø-usage gradually decreasing and *a* gradually increasing. *The* rises rapidly from the BA level and then consistently drops to the HM level. This reduction of flooding Huebner (1983) calls "trickling." The Japanese LM *the*-flooder increased *the*-usage to above the 1OO% line; without her, the Chinese and Russian average 72%. Even so, the most dramatic aspect of the article acquisition process for [-ART] speakers appears to be the simultaneous increase in *the*-usage with a
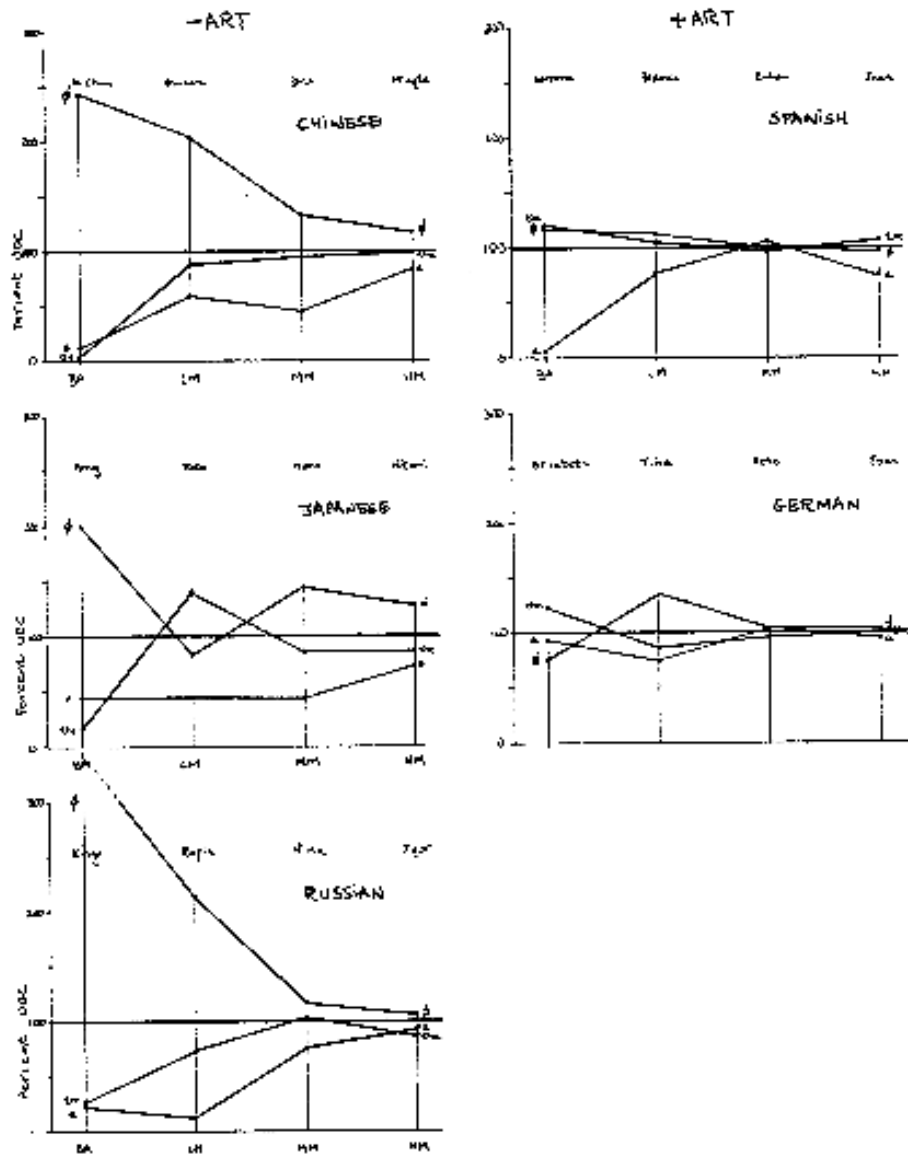
Fig. 6-14. Articles Used/Articles Required (UOC)

55

Fig. 6-15 Generalized Usage/Required (UOC)
[+ART] = Spanish and German
[-ART] = Chinese, Japanese, Russian

56
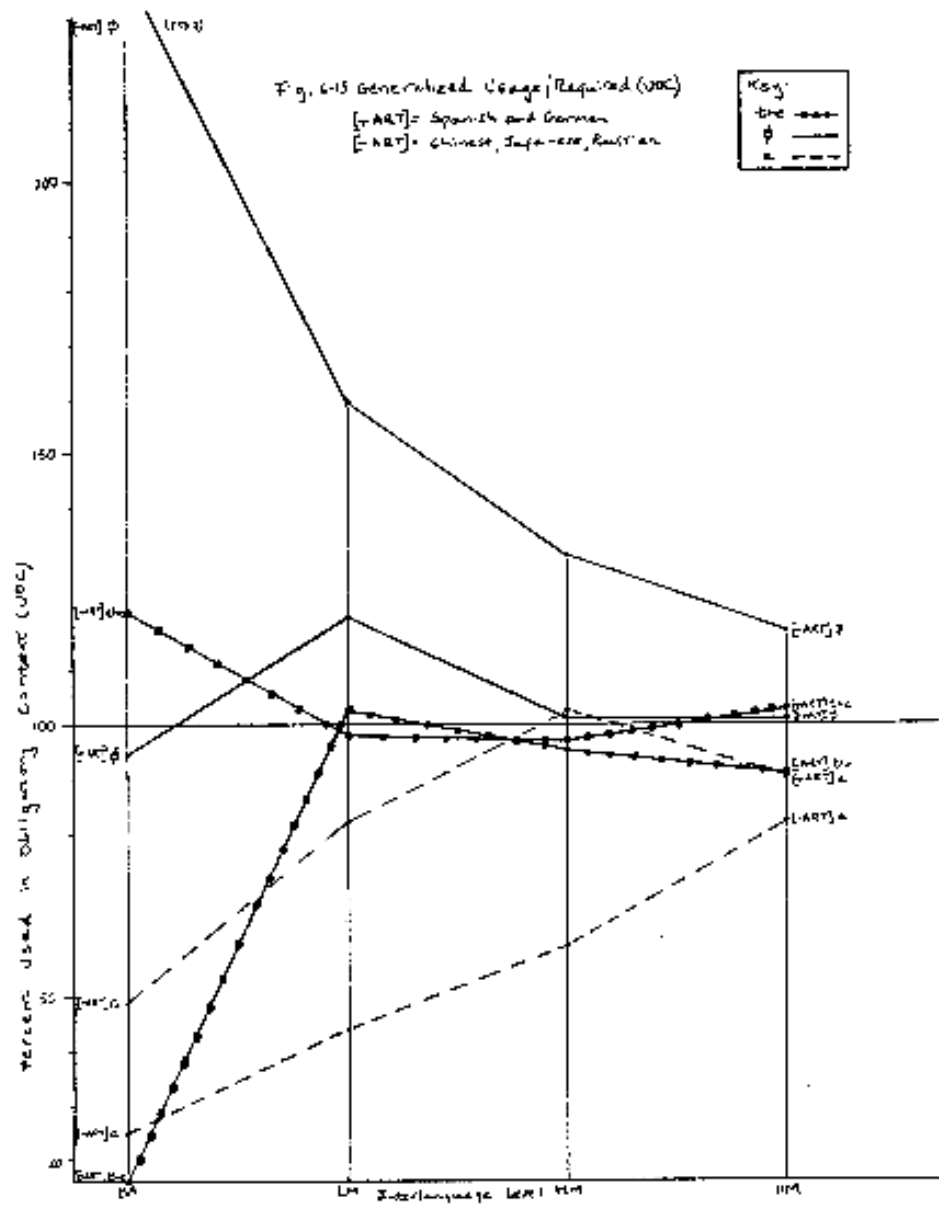
concomitant decrease in Ø-usage.

The [+ART] group, on the other hand, shows a high UOC for Ø and *the*, hovering more or less around the 100%-accuracy line across all IL levels. The most dramatic aspect of their article acquisition process is the rise (at almost a parallel rate to the [-ART] group) in *a*-usage, which has been linked throughout this study to the acquisition of the English [±count] feature and seems to constitute a process somewhat outside the acquisition of Ø and *the*.

Conclusions

The analyses in the foregoing sections lead to the following general conclusions:

1. Article usage at the BA level seems to reflect the article system, if any, that exists in the L1. The one exception to this is the Spanish usage of *a*, which seems to be acquired like a new lexical item.

2. The most dramatic acquisition activity for all subjects appears to occur at the LM level. For the [-ART] subjects, this includes a rapid increase in the use of *the* with a simultaneous decrease in the use of Ø. For the [+ART] subjects, this includes a rapid increase in the use of *a* for the Spanish subject and a considerable increase in the use of Ø for the German subject.

3. Acquisition of *a* and the [±count] feature takes place more slowly for the [-ART] than the [+ART] group, the former not achieving this until the HM level, the latter by the MM level.

4. The [+ART]/[-ART] distinction appears to be a useful and valid one in accounting for the acquisition of the English article system.

Chapter 7 Acquisition in Terms of Accuracy

     The discussion of the results from Chapter 6 has two goals.
Certain elements of the data will be used to justify the
pseudolongitudinal model used in this study by comparing them to
the findings of earlier studies. The data will then be
discussed in terms of the first hypothesis: that accuracy
(supplied in obligatory context or SOC) in all aspects of the
article system reflects increasing communicative competence.
Article usage data (used in obligatory context or UOC) will be
referred to in accounting for the shapes of the accuracy graphs.

Justification of Method
     This cross-linguistic interlanguage analysis of English
article acquisition is a logical consequence of the linkage
between negation structure and interlangauge level established
by Cazden, Cancino, Rosansky, and Schumann (1975). All the
subjects in the present study were assigned to one of the four
inter-language levels (basilang, low mesolang, mid-mesolang, and
high mesolang) based entirely on which negation structure (*no(t)*
+ verb, unanalyzed *don't*, auxiliary + negator, or analyzed
don't) was predominant in their English speech (see Table 5-1).
It is apparent from Figure 6-1 that the accuracy of article
usage with common nouns (i.e., excluding proper nouns) increases
for all five language groups, and that the negation criteria
that Cazden et al. (1975) described for Spanish speakers apply
to the interlanguages of the four other language groups as well.
     The claim that negation criteria effectively differentiate
the interlanguage levels can best be supported by comparing
certain aspects of interlanguage article usage in this
pseudolongitudinal study with those of true longitudinal
studies. The one factor that cannot be compared is the amount
of time between IL levels, as a pseudolongitudinal study
utilizes different subjects to represent different levels and
provides no indication of what the subject's IL looked like
before or after the moment of data gathering.
     Hakuta (1978) studied the English acquisition of a five-
year-old Japanese child. Using Brown's (1973) framework, he
charted the accuracy (SOC) of *the* at ten-week intervals. In
order to compare these results with those in my study, the SOC
for my four Japanese subjects is placed alongside Hakuta's
figure in Figure 7-1. Considering the fact that the subjects in
my study simply represent some point within the IL level, the
similarity in shape to Hakuta's figure, i.e., a peak followed by
a trough followed by a gradual rise, is noteworthy and suggests
that parallel acquisition characteristics may be taking place.
     Huebner (1983b) studied the English acquisition of a Hmong
man in his early twenties. He looked at the use of *the* in the

58

Fig 4-1 Comparison of Japanese Subjects with Uguisu (Maratu 1975)
in the SOC

A. Japanese Subjects: the SOC

B. Uguisu: the SOC
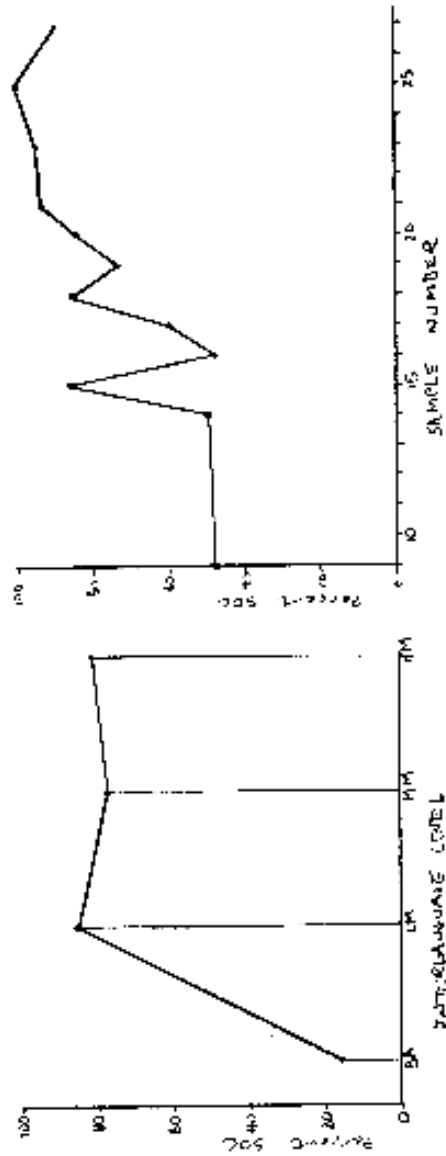
four environments (Categories I - IV) described by Bickerton. In order to compare his results with those in my study, I rendered Huebner's data table for the usage in Category II (specific definite *the*) and in Category III (specific indefinite Ø and *a*) in graphic form. Since I did not study Hmong speakers of English, I plotted Category II and III *the*-usage for the Chinese- and Japanese-speaking subjects against Huebner's figures as these southeast Asian languages are the closest to Hmong in my study. The results are shown in Figure 7-2. Considering once again that the subjects in my study represent some point within the IL level to which they belong, the similarity in shape to Huebner's figure (and this using a counting system completely different to that in the first comparison) is noteworthy and again suggests that parallel acquisition characteristics may be taking place.

Andersen (1977) looked at English article accuracy in a group of native Spanish-speaking students.   In arguing for a consideration of *a* and *the* as separate morphemes (and not as a single morpheme a là Brown (1973)), he ranked his subjects according to his Group Range method and plotted the accuracy (SOC) of *the*- and *a*-usage.   Andersen's is not a longitudinal study and therefore, like the present study, does not show the amount of time between IL levels. Nevertheless, to support the assumption that the data in the present study do reflect a developmental continuum, I plotted *a*- and *the*-accuracy for the Spanish subjects in my study against Andersen's figures. The results are shown in Figure 7-3. Although the lines of the two graphs are necessarily different in shape as Andersen's represents 76 subjects and mine 4, it is apparent that *a*-accuracy increases sharply and then gradually approaches 100%-accuracy in both graphs. Similarly, *the*-accuracy falls away from near 100%-accuracy at the outset and then returns, but never drops too far from the 100%-accuracy line.

Lamotte, et al. (1982) investigated English article acquisition in Spanish, Japanese, and Vietnamese speakers using the same model used in this study, i.e., a negation-based interlanguage analysis. Utilizing the TLU (target-like utterance) measure, among others, they reject the use of negation criteria for this purpose, stating, "A negation-based interlanguage continuum may not be the most accurate representation of second language acquisition for all language groups" (p. 8). They cite the lack of a "smooth progression" in the TLU of the Japanese and Vietnamese speakers as the reason for their rejection.

Since a "smooth progression" of article acquisition is not evident even in the longitudinal studies cited earlier (e.g., Hakuta 1976), I find this insufficient reason to reject the negation-based interlanguage continuum. Furthermore, there are

Fig. 7-Z. Comparison of Japanese and Chinese Subjects with Ge. (Huebner 1983b) in Category II and III the Usage.
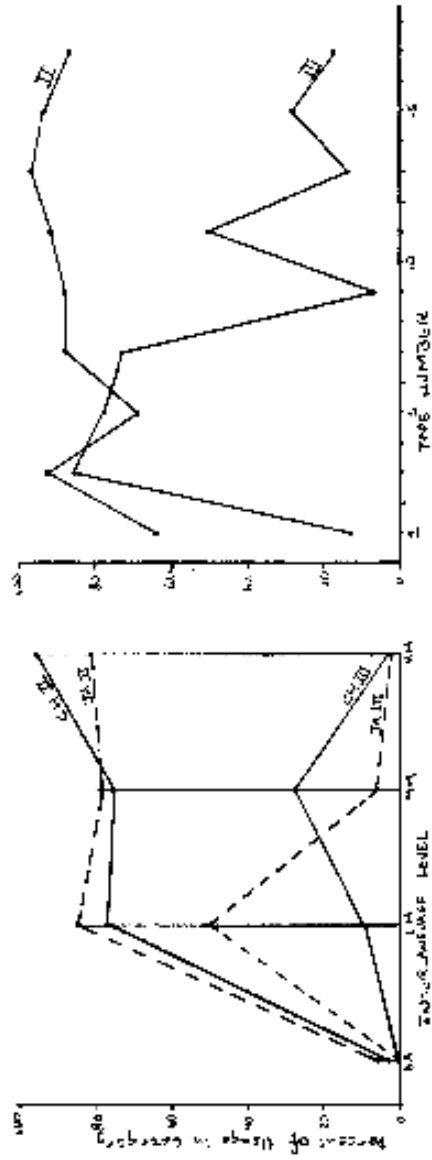
A. Chinese and Japanese the Usage in Categories II (cancel) and III (answer)

B. Gets the Usage in Categories II (lowest) and III (answer)

61

Fig. 4-3 Comparison of Spanish Subjects with Andersen's (1944) Spanish Subjects in the and a SOC

A. Spanish Subject the and a SOC

B. Andersen's Spanish Subject the and a SOC
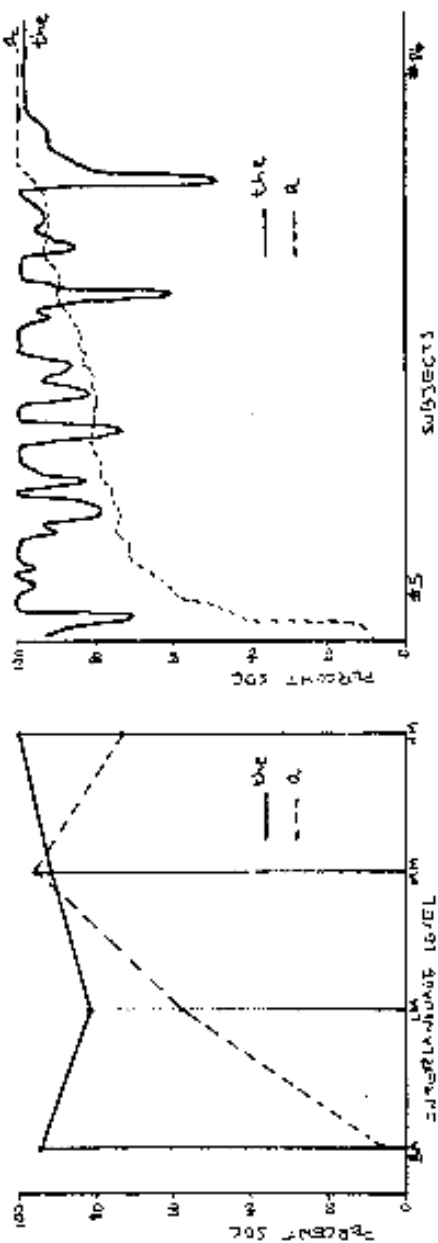
problems with the TLU measure, to be discussed below. Neverthe-
less, Lamotte et al.'s findings concerning article acquisition
by native Japanese and Spanish speakers are of interest for the
sake of comparison, and they are shown in Figure 7-4.

The similarities shown between the results of other
researchers and those of the present study, although
approximate, provide support for the assumption that the article
data in this study do reflect acquisition to some extent. In
other words, these pseudolongitudinal investigations roughly
parallel true longitudinal studies.   Further research using
this framework will have to be undertaken with larger numbers of
subjects to definitively establish the foregoing claim, but if
that is indeed done, the model described in this study could be
used as a basis for studying the acquisition of many aspects of
syntax without the need for longitudinal studies, whose time
requirement is in many cases prohibitive. At the least such
studies could be used to establish apparent trends, whose
reality could then be sought in true longitudinal studies.

The UOC Measure

This is a good point at which to discuss the advantages (at
least in dealing with article data) of the UOC (used in
obligatory context) calculation proposed in Chapter 6 as opposed
to the TLU (target-like utterance) measure devised by Stauble
1981. Figure 7-5 compares how the three measures depict *the*-
usage by the Spanish subjects as an example. The SOC line shows
simple accuracy. It cannot exceed 100%. The TLU is really a
kind of "docking" or penalty. It shows, for example, that the
95%-accuracy for Spanish BA *the*-usage is inflated, that it
should not be so high because the correct morpheme was used in
contexts where it should not have been. The UOC line shows the
number of times *the* was used divided by the number of obligatory
contexts. Since it shares the same denominator as the SOC
measure, it allows direct comparison. Furthermore, since it can
exceed 100%, UOC shows why the accuracy of *the* at the BA level
was so high: it is because *the* was being overused in a
considerable number of article contexts. Since ideally the
three measures (SOC, TLU, UOC) all equal one, they all tend to
converge at the HM interlanguage level. One suggestion that
would clarify matters would be to rename the term SOC (supplied
in obligatory context) because the word "supplied" does not
adequately distinguish "correct" from "used" morphemes (even
though SOC was originally meant to signify "correctly supplied
in obligatory context" (Schumann, personal communication)). A
better acronym would be AOC (accurate in obligatory context) to
differentiate it from UOC (used in obligatory context).

Fig 2-4 Comparison of Japanese and Spanish Subjects with Lemetti et al. (1982)
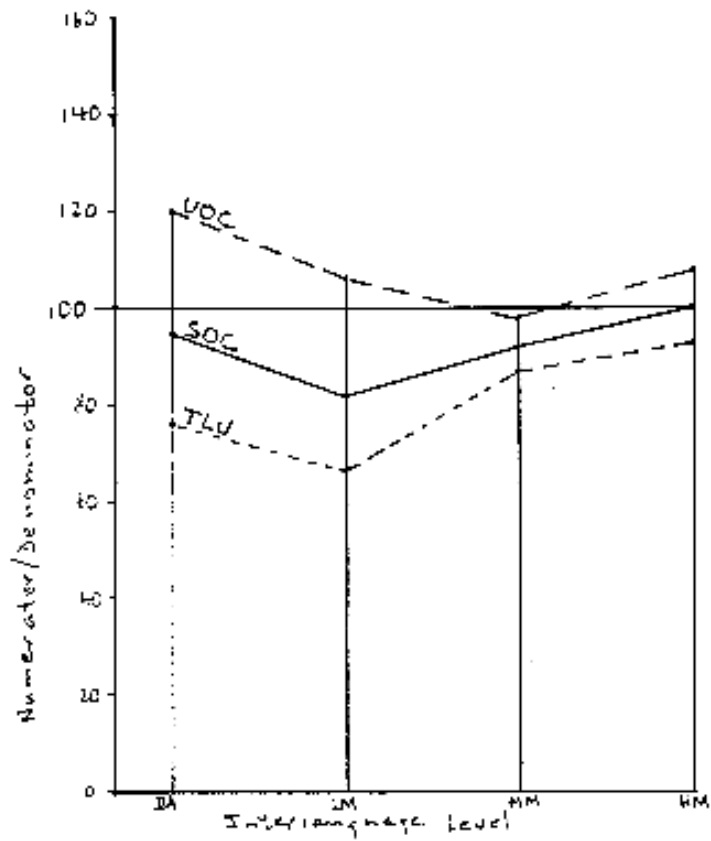Japanese and Spanish Subjects in this and a TLU
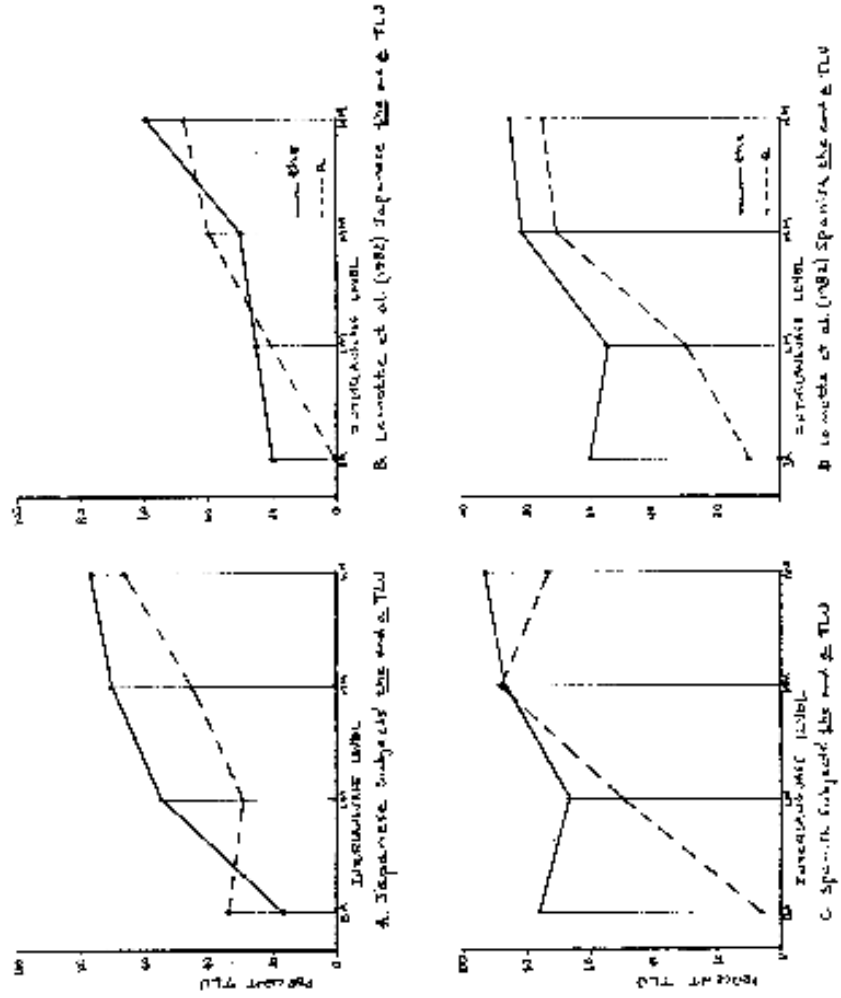
Fig. 7-5 Comparison of UOC, SOC, and TLU in the Spanish Subjects' Usage of the

65

Article Acquisition in the Five Language Groups
        The present study analyzes four "moments" of English IL
development in four different subjects who share the same first
language. For this reason, it can provide only a very
approximate picture of article acquisition. We do not know if
the "moment" that each subject represents is on an upward, a
downward, or a level trend or precisely where it might exhibit
radical departures from a trend. It is hoped that these
subjects reflect an IL development that is common to all
speakers of the same native language, but we cannot be sure. We
only know that these "moments" do not occur at the same IL level
and can thus be reasonably assured that they represent at least
successive stages of development.
        It is apparent from all the figures we have seen thus far
that article accuracy does not increase steadily and smoothly
with increasing IL level. Instead, it is characterized by
graphs whose peaks and troughs seem to be determined by the
presence or absence of an article system in the subjects' native
language, particularly at the lower IL levels. For this reason,
we will look at article acquisition first in the [+ART]
languages Spanish and German, and then in the [-ART] languages
Chinese, Japanese and Russian. Since acquisition implies
accuracy, the descriptions that follow will be based on the
accuracy (SOC) data from Chapter 6. However, since usage (UOC)
data can be very helpful in accounting for certain patterns of
accuracy, they will be considered too.

The [+ART] Languages

Spanish
        The definite article in Spanish is used in slightly more
environments than it is in English, but compared to the [-ART]
languages, it is roughly similar. For this reason, as shown in
Figure 7-6, *the*-accuracy in the Spanish BA subject is quite
high, probably reflecting the use of *the* very much as it is used
in Spanish. The high UOC for *the* allows for high *the*-accuracy
at this level. The fact that German *the*-accuracy is slightly
higher than Spanish *the*-accuracy might be accounted for by the
fact that the German definite article (nominative: *der*, *die*,
das) was usually expressed as *de* by the German subject (which
was always counted as correct) whereas the Spanish definite
article (nominative: *el*, *la*) is quite different from *the* and
requires the development of a new lexical item. This perhaps
accounts for the much lower BA TLU in Lamotte et al. 1982 (see
Fig. 7-4). The Spanish subjects never used *el* or *la* in their
speech. At the LM level, there is a marked drop in *the*-
accuracy, paralleling the decline in the number of uses of *the*.
This could reflect the "realization" on the part of the learner

66

Fig. 7-6. SPANISH Accuracy vs Usage (SOC)

A. φ SOC and UOC: Spanish

B. ½ SOC and UOC: Spanish

C. ing SOC and UOC: Spanish

67
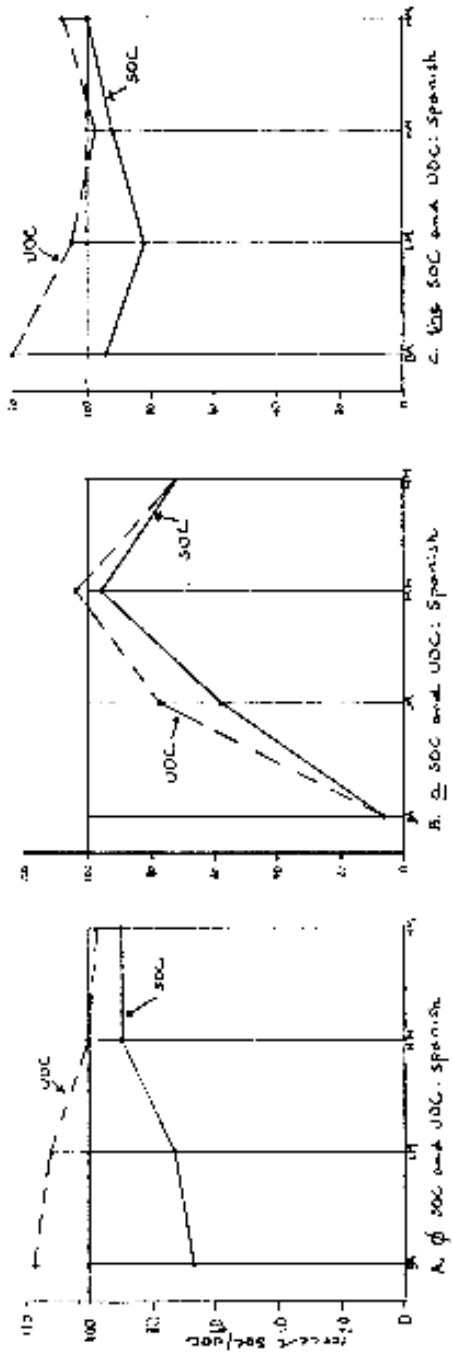
that although *the* is similar to Spanish *el/la*, it is not identical, thereby causing some hesitation in the blind application of the L1 article rules that the BA speakers used so effectively.

The three graphs in Figure 7-6 show that the BA subject used only Ø and *the* in her speech and *a* almost never. Yet at the LM level, *a* is used to a much greater extent, competing with *the* and Ø. This suggests that the LM subject has devised a new system that includes all three articles, although her hypotheses for Ø and *the* continue to overshoot the mark, perhaps showing residual L1 carry-over. The subject knows that only one article can precede a noun (there is little evidence of multiple article use in any of the transcripts in the corpus), and with *a* as a third choice, she appears to sacrifice *the* more than Ø, and thereby suffers a greater loss in *the*-accuracy.

Evidence that the LM subject is evincing a new system independent of her L1 comes from the following utterances:

26:1O I live on a̲ four floor
27:8   we are having laundry in a̲ same building
36:26 I think in a̲ day is better
37:26 [Int: Oh, they speak Spanish in the class?] Uh huh, yes, in a̲ class

These utterances would all be expressed using the definite article in Spanish. This suggests that the speaker has shifted from a simple reliance on L1 rules to an interlanguage hypothesis requiring the assignment of *a* to a singular countable noun since the head nouns in the four utterances are all of this type. From a slightly different angle, we could hypothesize that the speaker has become preoccupied with the complexity of the [±count] feature in English, causing her to override her L1 rules for the definite article. This scenario does not support the findings of Bertkau (1974) that article errors made by Spanish ESL students were more likely to be the product of simplification (e.g., Ø in place of *a* and *the*), although the LM subject is clearly trying to reduce the double load of determining the features of [±definite] and [±count] for every noun in the language.

Ø-accuracy at the BA level is similar to that of the German BA subject, although Ø-usage is much higher in the Spanish subject because *a* is so little used. At the LM level, Ø-usage drops slightly as accuracy increases, reflecting the fledgling interlanguage article system.

At the MM level, accuracy for all three articles hovers near the 9O%-accuracy level and the distance from usage decreases to approximately equal amounts for all three articles. The articles would be labeled "acquired" at this interlanguage
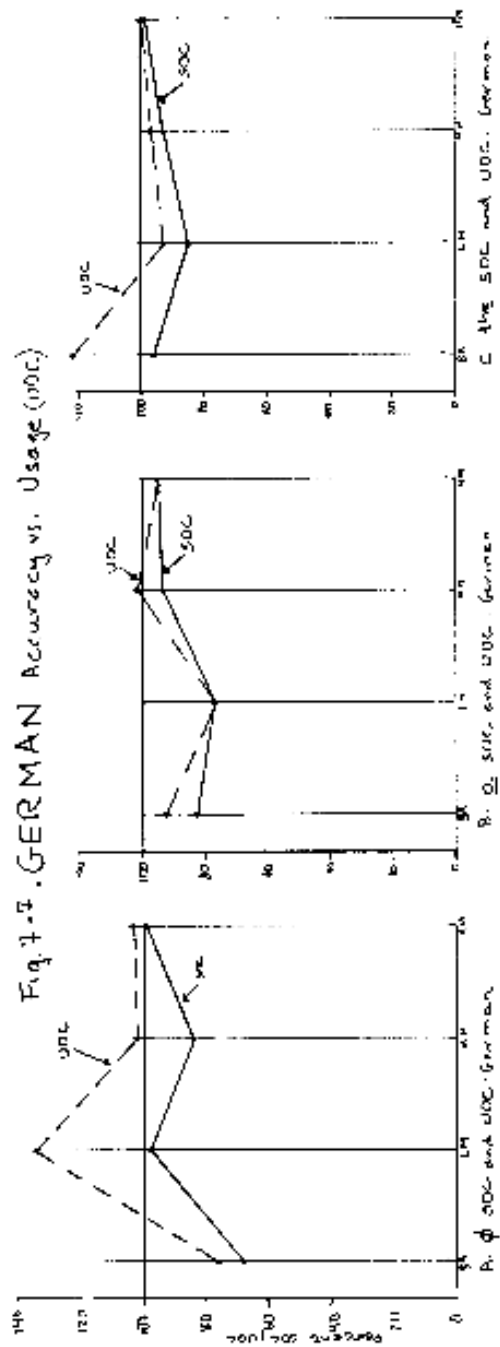
68

level using Brown's (1973) 90%-criterion. Article usage beyond this point reflects higher accuracy for *the*, equal accuracy for Ø, and lower accuracy for *a*, although the drop is partly dependent on the small number of noun phrases in the HM subject's transcript. This subject (Juan) had only 22 obligatory contexts for *a*, which makes his six errors more prevalent than they should be. The well over six errors made by three of the four remaining HM subjects is obscured by the high number of noun phrases in their transcripts. In other words, no great significance should be attached to the HM drop in Ø-accuracy and usage.

To summarize, article acquisition in the Spanish-speaking subjects in this study appears to take the following pattern:

BA   L1 rules are adhered to for Ø and *the*. *A* is used
        very little and *one*-use is far below that in L1.

LM   *The*- and Ø-usage drops as *a*-usage dramatically
        increases, leading to decreased *the*-accuracy,
        slightly increased Ø-accuracy, and considerably
        improved *a*-accuracy. This appears to reflect a
        developing system for article usage.

MM   *The*- and Ø-usage continue to fall and *a* to rise,
        while accuracy for all three articles hovers near
        90%. The article system appears to be nearly
        target-like at this level.

HM   Usage of all three articles hovers around the 100%-
        accuracy line, with the exception noted for *a*.
        Fluctuations at this stage probably reflect
        interference from other aspects of the target
        language, which at this IL level represents many
        acquired systems.

German
Figure 7-7 compares SOC and UOC for the German subjects. At the BA level, *the*-accuracy is very high, as it was for the Spanish BA subject. This no doubt reflects L1 article usage, as usage in both subjects is approximately the same, about 120%. Ø-usage at this level is much lower for the German subject, but accuracy is similar to the Spanish subject. A major difference from the Spanish subject is the German subject's high *a*-usage and accuracy. This would suggest that the German BA subject has retained a strong sense of the German [±count] system in her English IL. Why the German subject should do this and not the Spanish subject I cannot explain. However, the German subjects used German words in their speech to a much greater extent than

69

Fig. 7.7. GERMAN Accuracy vs. Usage (UDC.)

A. φ SDC and UDC. German

B. α SDC and UDC. German

C. Avg SDC and UDC. German

70

all the other subjects in this study, even occasionally at the
MM and HM levels. Since English is a Germanic language, perhaps
German learners of English have the sense that there is a great
similarity between the two languages and this encourages them to
transfer lexical items. In the transcripts, although the
interviewers spoke only English, the BA and LM subjects often
used German words as if they would be readily understood by the
interviewer. For example, Elisabeth (BA level) says:

    4-22 for de dinner eingeladen [= invited]
    5-12 I have kennen gelernt viele [= met many] people
    9-22 three stufes [= levels] for the school

And Tina (LM level) says:

    1-1O dis is die...grenze [= border] von Austria
    6-1 I have a German freundin [= female friend]
    11-31 she go under estrich [= sidewalk]

None of the other subjects used first language lexicon so freely
in their interlanguages.

    At the LM level, *the*-usage declines with a corresponding
drop in accuracy in much the same manner as for the Spanish LM
subject. However, in contrast to the Spanish subject, *a*-usage
and accuracy drop for the German subject. This suggests that
rather than creating a new [±count] system as the Spanish
subjects appear to do, the German subjects simply modify their
German [±count] systems, with the distance between usage and
accuracy diminishing for both *a* and *the*. Like the Spanish LM
subject, however, the German LM subject also makes errors with
the that she would not have made if she had simply been
following the rules of her L1. Tina produces the following
utterances:

    1-31 I am go...to <u>opera</u>
    11-27 I see <u>first mal</u> [= time]
    15-4O Gretel is in <u>house</u>
    17-7 im [= in] Mexico to <u>congress</u>

These utterances would all be expressed with the definite
article in German. However, they seem not to reflect an
emerging article hypothesis but rather a simplification of
German prepositional phrase structure. In these four cases, the
article would have been combined with the preposition in German:

    to opera = zur (zu der) Opera
    [for] first mal = zum (zu dem) ersten mal

```
    in house = im (in dem) Hause
    to Congress = zum (zu dem) Kongress
```

This suggests that rather than grappling with a [±count] system
as Blanca appears to be doing, Tina has simply failed to "pull
apart" her contracted German preposition + article structures--
or dropped them altogether as in line 11-27. This accounts to
some extent for the marked increase in Ø-usage at the LM level
leading to higher Ø-accuracy than even the MM subject attains.
Whether this increase is characteristic of all German speakers
of English or simply of this subject I cannot say. Resolution
will depend on extending the same kind of analysis to several
German LM level speakers, which is beyond the scope of this
study. In any event, Tina's high Ø-usage and accuracy might be
a result of her tendency to drop potentially confusing aspects
of syntax (or better, her recognition that L1 rules no longer
apply), which would more closely agree with the simplification
strategies noted in Bertkau (1974).

       It is perhaps pertinent here to mention the peculiar nature
of the zero article. There are few syntactic environments in
English where Ø plays such a crucial role, and in terms of
acquisition it is certainly unique. Essentially, it requires
the recognition on the part of the learner that nothing, the Ø-
morpheme, equals something, the zero article. This is unlike
most other morphemes measured in morpheme acquisition studies.
In Brown, Cazden, and de Villiers, a morpheme is either P
(present), A (absent), OG (overgeneralized), or X (incorrectly
supplied). But for the zero article, present and absent (and
even overgeneralized and incorrectly supplied under some
circumstances) are indicated in the same manner, i.e., with
nothing, which creates a problem for the interlanguage
researcher: Is the morpheme being used or neglected?[3]   A similar
problem must occur for the ESL/EFL learner, which will become
apparent when we discuss article acquisition by the [-ART]
language groups.

       At the MM level, *a* and *the* usage hovers near the 90%-
accuracy level, as it did for the Spanish subjects, and these
articles can be said to be acquired at this point.   The drop in
Ø-accuracy, despite the similar level of usage to the Spanish LM
subject, is partially the result of [±count] problems, as is seen
in the following examples from Reto:

```
    1-30 they have a radar
    4-16 it's a small traffic
    10-53 they go with exposation (explosion)
```

_____
[3]This question led Lamotte et al. (1982) to the decision not to
investigate Ø-usage at all.

This might suggest that the establishment of an independent [±count] system appears later in the interlanguage of German speakers than it does for Spanish speakers. On the other hand, it may be unique to this subject.

Article accuracy at the HM level continues to approach the 100%-accuracy level as the distance between usage and accuracy decreases to zero.

To summarize, German acquisition of the English article system appears to take the following pattern:

BA   L1 rules are adhered to for all three articles.

LM   *The-* and *a*-usage drop with a corresponding decrease in accuracy. Ø-usage sharply increases, partly reflecting the simplification of an L1 structure (preposition-article contraction).

MM   *The* and *a* hover near the 90%-accuracy level, Ø somewhat lower (84%), possibly reflecting adjustment of the [±count] feature. The system is essentially acquired.

HM   Accuracy is near the 100% level for all three articles.

The [-ART] Languages

Chinese

The BA level Chinese subject depicted in Figure 7-8 shows a pattern of article usage that is quite different from the corresponding [+ART] subjects. Ø-"usage" is extremely high, reflecting not usage at all but rather the lack of any article system in the native language. Since *a* and *the* are used hardly at all, most noun phrases are "marked" with the Ø article, leading to nearly 100%-accuracy for this article.

At the LM level, Ø-usage drops to make way for a considerable rise in *a*- and *the*-usage, resulting in a slight drop in Ø-accuracy. This reflects a strong parallel with the acquisition in the [+ART] group: in both cases, the presence of a similar feature in the L1 leads to high usage and accuracy at the BA level which at the LM level must be decreased to make way for increasing use of the other articles, leading to decreased *the*-accuracy for the [+ART] group and decreased Ø-accuracy (or, at least, no increase for the Chinese subject) for the [-ART] group.

The rise to approximately 50% *a*-accuracy at the LM level suggests, as it did for a similar level of accuracy in the

73

Fig. 4-8 CHINESE Accuracy vs. Usage (UOC)

A. ∮ VOC and UOC: Chinese

B. ∮ SOC and VOC: Chinese

C. ∮ SOC and UOC: Chinese
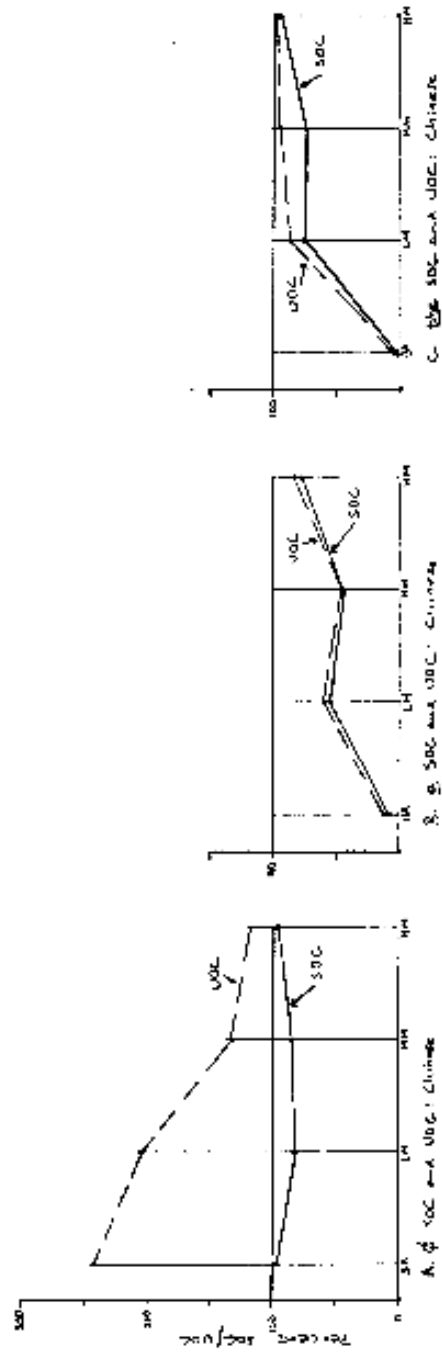
Spanish subject, that a new system is being developed that includes the features [±definite] and [±count]. The [±definite] feature seems to be of greater importance (or perhaps greater ease) for the developing interlanguage, as *the*-usage and corresponding accuracy are considerably higher than *a*-usage and accuracy. The same will be even more pronounced for the other [-ART] subjects.

At the MM level, Ø-usage continues to drop, leading to slightly increased Ø-accuracy. Can we say that the zero article is now being chosen as opposed to articles simply being dropped, as was clearly the case at the BA level? This is the problem when nothing equals something.

A-usage and accuracy drop at the MM level, but only slightly. Such a drop was described for the German MM subject as arising partially from adjustment of the [±count] feature. And indeed, Dr. X makes similar errors:

    5-25 know the basic of <u>foreign language</u>
    8-8 my interesting is to learn...<u>a art history</u>
    12-3 I stay in American (America) or <u>western country</u>

He also sometimes uses *one* instead of *a*, which is scored as being incorrect:

    1-45 I waste (visit) <u>one professor</u>
    2-22 I have /raiten/ (written) <u>one books</u>
    13-4 I have <u>one plan</u>

The-usage continues to climb at the MM level, but accuracy stays about the same. Unlike for the [+ART] group, the article system cannot be said to be acquired at this level, although a working system is well in place and the distance between usage and accuracy diminishes considerably for Ø.

At the HM level, Ø- and *the*-accuracy are very near the 100%-accuracy level. Ø is still being overused a little, with a corresponding underuse of *a*, again probably reflecting adjustments to the [±count] feature. The system can be said to be acquired at this point.

In summary, Chinese acquisition of the English article system appears to take the following pattern:

    BA L1 conditions (i.e., no articles used) dominate,
         resulting in very high Ø-accuracy.

    LM Ø-usage and accuracy decrease. *A*-usage and
         accuracy increase to around 50%. *The*-usage rises
         dramatically to almost 90%, with a corresponding
         accuracy of just over 75%.

75

MM   Ø-usage continues to fall as Ø-accuracy starts to
     increase. *The*-usage continues to rise while
     accuracy levels out. *A*-usage and accuracy fall
     slightly.

HM   Ø-usage declines to the least distance from the
     accuracy line, which is almost 100%. *The*-usage
     and accuracy converge at just below 100%. *A*-
     usage and accuracy come together at about 80%,
     possibly reflecting adjustment of the [±count]
   feature. The system is essentially acquired.

Japanese
     The BA level Japanese subject depicted in Figure 7-9 shows
a pattern of article usage that is quite similar to that of the
Chinese subject (Fig. 7-8). Ø-"usage" is quite high, with the
result that Ø-accuracy is almost 100%. *The*-usage is very low,
though not as low as it was for the Chinese BA subject.
However, *a*-usage is already at nearly 50%, with a corresponding
a-accuracy of about 35%. This suggests that the Japanese BA
subject is at a higher basilang level than her Chinese
counterpart, and indeed, her English is considerably more
fluent, with a much larger vocabulary. This indirectly supports
an earlier suggestion that control of *a* and the [±count] system
is a product of experience rather than strategy (e.g., *the*-
flooding). It is also interesting to note that for all the
subjects discussed so far, the distance between *a*-usage and *a*-
accuracy is invariably the smallest, often with no difference at
all. This suggests that *a*, unlike Ø and *the*, is used rather
deliberately and not just as a random guess. This is probably
why *a*-flooding (Fig. 6-12) does not occur.
     At the LM level, the picture is quite different from the
Chinese subjects. The LM Japanese subject is the *the*-flooder so
often mentioned in Chapter 6. The graph shows the effect of
this subject's strategy: *the* is used to such an extent that
accuracy climbs to almost 90%. Something has to give to allow
this much *the*-usage, and Ø-usage plunges with a corresponding
drop in accuracy. Huebner (1983b) found his Hmong subject to
the-flood to a similar degree and questions arise as to the
universality of this phenomenon: do all Hmong and Japanese
learners of English employ the strategy or only certain
individuals? Why didn't the Chinese LM subject flood to the
same extent? These questions can only be answered by conducting
the same study using many individuals with the same L1 and IL
level.
     Despite the LM Japanese subject's extreme flooding
behavior, *a*-usage and accuracy remain at about the same level as
for the Chinese LM subject, providing evidence once more that

76

Fig. 3-9 JAPANESE Accuracy vs. Usage (UOC)
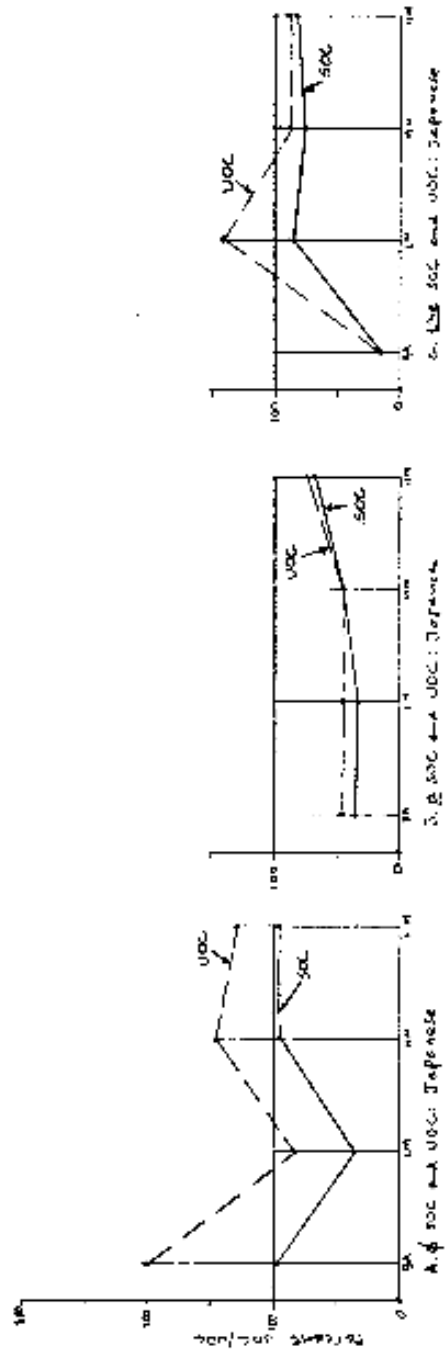
the control of *a* is part of another system (adjustment of the [±count] feature) that matures somewhat independently of the article system.

At the MM level, usage and accuracy for all three articles is very similar to the Chinese MM subject: Ø is still overused with an accuracy near 100%, *a*-usage and accuracy converge near 50%, and *the* is at around 80% accuracy, *the*-usage having dropped considerably. The usage and accuracy of *a* still remain well below that of *the* and Ø, which appear to be the competing hares in the race, while the turtlish *a* creeps along at its steady pace.

At the HM level, the Japanese subject continues to parallel the usage and accuracy of the Chinese subject. In contrast to the BA subjects, the language of the Chinese subject is somewhat more fluent than that of the Japanese speaker, suggesting that the former is a little further along in the HM level than the latter. This is reflected by the Japanese subject's slightly lower accuracy for both *a* and *the*. The system can be said to be acquired at this point.

In summary, Japanese acquisition of the English article system appears to take the following pattern:

BA L1 conditions (i.e., no articles used) dominate, resulting in very high Ø-accuracy and low *the*-usage [this subject's *a*-usage is quite high for a [-ART] basilang subject, suggesting that she is rather a late basilang, which her negation supports].

LM Ø-usage and accuracy fall in response to extreme *the*-flooding, which produces abnormally high *the*-accuracy [whether this pattern is true for all Japanese speakers remains to be studied]. *A*-usage and accuracy remain below 50%.

MM Ø-usage rises to a point close to that of the Chinese MM subject. *The*-usage falls from the flooded height while accuracy levels out. *A*-usage and accuracy stay at about the same level.

HM Ø-usage declines to the least distance from the accuracy line, which is almost 100%. *The*-usage and accuracy converge at just below 90%. *A*-usage and accuracy come together at about 80%, possibly reflecting adjustment of the [±count] feature. The system is essentially acquired.

With the exception of the degree of *the*-flooding at the LM

stage, the Chinese and Japanese subjects show very similar patterns of article acquisition.

Russian

The BA level Polish subject depicted in Figure 7-10 shows a pattern of article usage that is very much like the Chinese and Japanese BA subjects. Ø-"usage" is extremely high, leading to nearly 100%-accuracy. *A* and *the* are used seldom, more frequently than by the Chinese subject and less frequently than by the Japanese subject.

At the LM level, Ø-usage drops steeply, but accuracy remains about the same. As was the case for the Chinese and Japanese subjects, *the*-usage increases dramatically at the LM level, resulting in increased accuracy; however, this increase is not as steep as the increase shown by either the Japanese or Chinese subjects. *A*-usage and accuracy are actually slightly lower than that of the BA subject: presumably, it is the [±count] feature that is causing difficulties.

At the MM level, Ø-usage continues to drop, this time affecting accuracy slightly. *The*-usage has risen to just over 100% with a concomitant accuracy of around 90%. *A*-usage and accuracy (practically identical in these subjects) has risen considerably to around 80%. It should be pointed out that the usage (UOC) measure of over 100% cited above does not depict a one-to-one correspondence. It simply indicates that the number of times the article was used was slightly higher than the number of times it was required.

At the HM level, Ø-usage and accuracy converge near 100%. *A*-accuracy is over 90%. *The*-usage and accuracy converge at a slightly lower level than the MM subject.

In summary, Russian acquisition of the English article system appears to take the following pattern:

BA L1 conditions (i.e., no articles used) dominate, resulting in very high Ø-accuracy. *A* and *the* are used very little.

LM Ø-usage decreases and Ø-accuracy remains at the same level. *The*-usage and accuracy increase sharply. *A*-usage and accuracy decrease slightly.

MM Ø-usage continues to decline, this time resulting in slightly decreased accuracy. *The*-usage and accuracy rise to over 90%. *A*-usage and accuracy rise to 80%.

Fig 7-10 RUSSIAN Accuracy (SOC) v Usage (UOC)

80
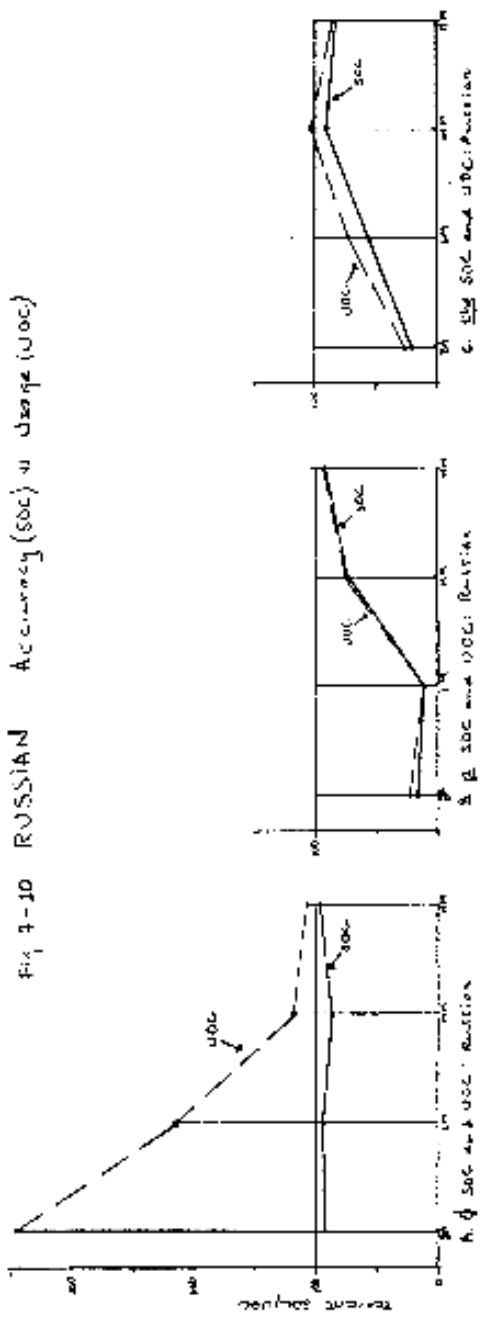
HM Ø-usage and accuracy converge near 100%. *The*-
      usage and accuracy declines slightly. *A*-usage
      and accuracy converge near 95%. The system is
      essentially acquired.

The Russian subjects behave in much the same way as the other
members of the [-ART] group, with two exceptions: *the*-usage and
accuracy climb less steeply and *a*-usage and accuracy climb more
steeply. In some ways, this places the Russian subjects midway
between the [+ART] and [-ART] groups. This might be because the
three Russian subjects (but not the Polish one) were all Jewish,
and as such they might well have been exposed to Yiddish, which,
like German, does have an article system. On the other hand,
Russian, German, and Spanish are all Indo-European languages and
perhaps it is this overall relatedness that makes the Russian
subjects more like the [+ART] group while clearly belonging to
the [-ART] group and, hence, "in between."

Conclusion
      What is clear from the foregoing description of article
acquisition by representatives of the five language groups is
that speakers whose first language contains an article system
behave in roughly the same manner and that speakers whose first
language does not contain an article system behave in roughly
the same manner. This conclusion supports the proposal in
Hypothesis I that the use of the English article system reflects
increasing communicative competence: if this were not the case,
there would have been no consistent pattern among any of the
language groups. The [+ART] group appears to acquire the
article system by the MM level; the [-ART] group appears to
acquire the article system not until the HM level. Finally, for
all subjects, *a* appears to function somewhat independently of
the article system, evincing for the most part a more gradual
acquisition process than the wildly fluctuating Ø and *the*, and
it is proposed that this is the result of the fact that *a*
represents acquisition and control of the [±count] feature. The
fluctuations in the use of Ø and *the* at various stages suggest
that it is these articles that are used in hypothesis testing
about the article system, and not *a*.

Chapter 8 Article Usage: Strategies of Acquisition

     The discussion of the results from Chapter 6 is continued
in this chapter with the focus now upon usage as opposed to
accuracy. The data are discussed in terms of the second
hypothesis: that the use of the English article system reflects
certain strategies of interlanguage development. But first, a
brief discussion of the place of the articles in current
syntactic theory is undertaken to see what light it can shed on
the acquisition process.

The Articles in Current Syntactic Theory
     Chomsky's current syntactic theory, i.e., the government
and binding (GB) framework, places the article system within a
category called "Specifier." The articles *a* and *the* constitute
noun-phrase specifiers, but little more is said about them in GB
theory. Certainly no mechanism is postulated for the initial/
subsequent mention rules of article use (i.e., *a/0 → the*), let
alone a framework for the determination of definiteness through
shared knowledge or implied postmodification.
     Jackendoff (1977) elaborates on the term NP *specifier* but
his primary concern is to account for the distribution of
phrases such as *the many apples* vs. incorrect *the all apples*.
In his X-bar syntax, the constituent "Art" includes words that
are semantically demonstratives (*a*, *the*, *this*, *that*, *these*,
those + partitive, *which* + partitive, *what*, *we*, and *you*-plural),
words that are semantically quantifiers (*any*, *all*, *no*, *some*,
each, *every*, *either*, *neither*, *both*), and words which are true
quantifiers (*many*, *much*, *several*, *few*-negative, *little*-negative,
one-nonpluralizable). The zero article (∅) is, however, nowhere
considered. Current GB theory thus provides little insight into
the basic article acquisition process except for the fact that
all noun phrases must have a specifier. In the accuracy data
discussed in Chapter 7, the recognition by the [-ART] group that
a specifier (in this case an article since no other specifiers
were investigated) is required seems to occur between the BA and
LM interlanguage levels. Otherwise, the traditional approach to
syntax reflected in the classification codes of this study
appears to be a more useful one in analyzing article acquisition
than the GB framework.
     The Columbia form-content approach, on the other hand, does
provide a principled framework for making the indefinite/
definite distinction based on the meaning the speaker intends to
convey. Although Garcia (1975) makes it clear that this is the
exclusive domain of native speakers of English, it was often
found that a consideration of form over function (i.e., the use
of the articles irrespective of category) provided useful
insights into the article acquisition process (e.g., the fact

that [+ART] *the* behaved across the four IL levels much like
[-ART] Ø). This approach will be used later in this chapter in
discussing usage strategies in terms of each article.

Article Usage in the Five Language Groups
     Article usage differs from article accuracy in that the
denominator of the usage formula is based on one of Bickerton's
four categories rather than obligatory context (these categories
were described in Figure 5-2 on page 21). As explained in
Chapter 6, the categories essentially differentiate generic from
specific noun phrases and then further subdivide generic into
true and quasi-generic (indefinite generics with *be* as the main
verb) and specific into indefinite and definite. The reason for
ignoring obligatory context is the notion that interlanguage
systems are created whether or not they are accurate, and useful
information about the acquisition process can be derived from an
analysis of such systems (Andersen 1977, Huebner 1983a).
     It must be remembered, however, that the subjects in this
study are at some point along the interlanguage continuum,
although we can be reasonably sure that the four subjects in
each language group represent successive stages of development
as indicated by their IL levels. Further research with a much
larger number of subjects with the same L1 and IL level must be
undertaken before any definitive statement can be made. The
present study merely reveals apparent tendencies.
     Usage is analyzed in the following sections first in terms
of the subjects' first languages and then in terms of the
individual articles. The focus will be upon the apparent
strategies that the 20 subjects use in acquiring the English
article system.

The [+ART] Group

Spanish
     Category I, the true generic category, allows the use of
all three articles. Since this category is not limited to one
article over another, it provides an initial suggestion as to
which articles are selected without the constraint of potential
inaccuracy (the remaining categories all have unallowable
articles). It can thus provide an indication of general
article usage, although problems with the precise identification
of generic noun phrases, especially with the sometimes sparse
contexts of the lower IL levels, as mentioned in Chapter 6,
render any statements made about apparent generic strategies
tenuous at best. Figure 8-1 shows that for the Spanish
subjects, the higher the IL level, the more Ø is used. Generic
a and *the*, however, are used increasingly less (for the most
part) with increasing IL level.

Fig. 8-1 SPANISH Usage Within Categories I-IV
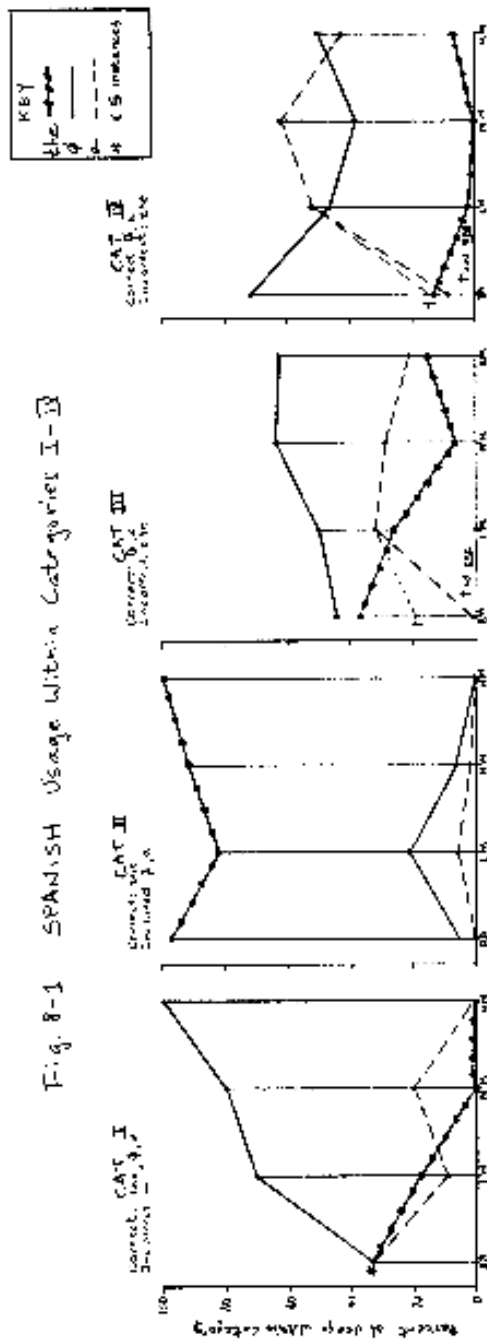
With the exception of the BA subject, a parallel pattern of usage is evinced in Category III, in which Ø and *a* are correct but not *the*. In fact, there is a drop in *the*-usage between the BA and LM levels in all four categories. As a usage strategy, it would appear that one of the first things a Spanish speaker does in acquiring the English article system is to reduce the use of *the*. This tendency is correctly reversed at the LM level in Category II, where *the* is a correct usage, but essentially persists in all four categories. The one exception is the somewhat anomalous HM subject, who exceeds the MM subject in overgeneralizing *the*-usage into inappropriate categories, namely III and IV, though at lower levels than the BA and LM subjects.

Since the vertical columns in Fig. 8-1 represent 100% usage within each IL level, one of the remaining two articles must be utilized when usage of the third is reduced. In Category II, there is a peak of Ø-usage at the LM level and a lesser peak of *a*-usage to make up for the diminished use of *the*. This raises the unanswerable question discussed in Chapter 7: Is it the Ø article or simply no article that is being used?[4] In other words, does this slight peak of Ø-usage represent the failure to use an article or is it a correct usage of the zero article? However, since *a*-usage also rises slightly at this level, it could reflect the dawning of the recognition that there is another system, namely the [±count] feature, that must be considered in choosing the correct article in English. The [±count] feature, as discussed in Chapter 7, requires the assignment of Ø or *a* depending on whether the noun phrase is a mass or a count noun. In Category III, where the indefinite articles are correct, *a* and Ø usage increases substantially (especially *a*) between the BA and LM levels. In Category IV, where again only the indefinite articles are correct, *a*-usage increases considerably with increasing IL level. One would expect this tendency to continue through to the HM level, but for the HM Spanish subject, as discussed in Chapter 7, *a*-usage is abnormally low. In Chapter 7, we also saw that *a*-accuracy tends to increase in a manner quite different from *the* and Ø, which led to the hypothesis that *a*, more than *the* and Ø, reflects the acquisition of the [±count] feature (especially as *a* seems to be acquired like a new morpheme for Spanish speakers). The relative independence of *a* from Ø is shown in the graph for Category I.

Finally, the U-shaped trajectories of *the*-usage in Category II (correct) and Ø-usage in Category IV (correct) suggest that

---

[4]The only way to determine the difference is to do a contrastive analysis between the L1 and English, as suggested by Andersen (1977) and discussed in Chapter 3. Such an analysis is beyond the scope of this study,

this pattern is associated with the correction of an L1 trait (both Ø and *the* exist in Spanish), although why this should occur in Category IV and not in Category III, I cannot explain. It remains to be seen whether this preliminary generalization is supported by the other language groups.

German

Figure 8-2 shows that article usage by the German subjects in Category I, where any article can be correctly used, is similar to the Spanish group in that *the*-usage generally decreases with increasing IL level. Ø-usage is quite different, however, showing a steep drop from the BA to the MM level. Nevertheless, Ø returns to 1OO%-usage at the HM level, as was the case for the Spanish HM subject. *A*-usage is also different in that at the BA level, the German subject does not use generic *a* at all. However, like the Spanish subjects, the German subjects evince a peak of *a*-usage at the MM level (although the German MM peak is over four times higher than the corresponding Spanish one) and a drop at the HM level which is lower than *a*-usage at the LM level. Category II usage is almost identical to that in the Spanish group: a U-shaped *the*-trajectory across all IL levels with the trough at the LM level, an equivalently mild peak of Ø-usage to compensate for the decline in *the*-usage, and very low *a*-usage.

Category III usage also parallels the Spanish group in that *the* is used increasingly less with increasing IL level. However, the German BA subject shows much higher usage of *a* than the corresponding Spanish subject, suggesting that the notion of *a* is transferred from German but not from Spanish, as was discussed in Chapter 7. However, the general levels of Category III *a-*, *the-*, and Ø-usage by the German subjects are similar to those of the Spanish group.

Category IV usage parallels the Spanish group in that *the* is used very seldom across all IL levels. Furthermore, between the LM and MM levels, Ø-usage decreases and *a*-usage increases in a similar fashion. In contrast, *a*-usage at the BA and HM levels is much higher for the German subjects whereas Ø-usage is much lower.

These data suggest that, like the Spanish subjects, German learners of English generally decrease their *the*-usage across all IL levels in all categories except Category II, where this trend is correctly reversed at the LM level. On the other hand, the German subjects tended to use *a* more freely when it is correct (Categories I, III, and IV) and less when it is incorrect (Category II) than their Spanish counterparts, revealing a more English-like control of the indefinite article. Is it the common heritage of English and German that allows for this ability?

86

Fig 8-2 GERMAN Usage Within Categories 1-IV
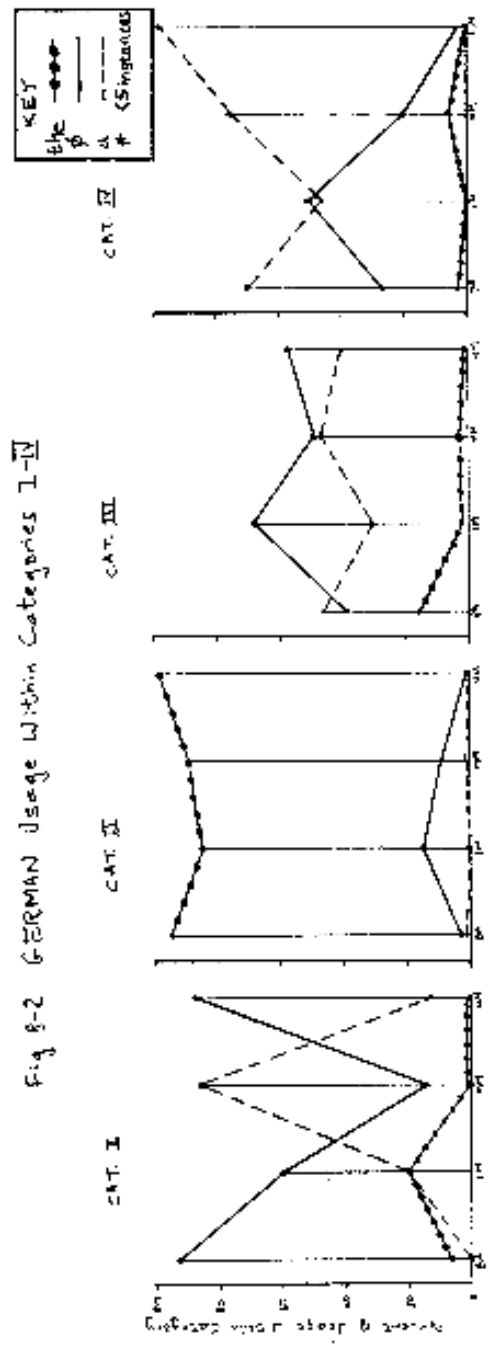
Finally, U-shaped trajectories occur with Ø in Category I, with *the* in Category II, and with *a* in Category IV, paralleling the Spanish performance only in Category II. This lends a little support to the hypothesis that U-shaped trajectories reflect the adjustment of a similar morpheme in the L1 (*a*, *the*, and Ø exist in German). However, it does not explain why such trajectories are not seen in all categories.

The [-ART] Languages

Chinese
     Category I usage, as described earlier, provides a picture of article usage without the constraint of perceived inaccuracy (the rules governing the choice of generic *the*, Ø and *a* are quite subtle). Thus, Figure 8-3 shows the Chinese subjects to have a peak of *the*-usage at the MM level and a U-shaped Ø-trajectory with a trough at the MM level corresponding to the *the*-peak. This peak is evidence of *the*-flooding, which was discussed at length in Chapters 6 and 7. A smaller but analogous peak occurs at the MM level in Category III, where *the* is not correct.
     Ø-usage declines initially in all four categories in much the same way that *the*-usage declined for the [+ART] group. It continues to decline in Category II, where Ø is incorrect, but turns around in Categories I, III, and IV, where Ø is allowed, at the LM or MM levels.
     A-usage, which, like *the*-usage, represents the acquisition of a new morpheme for the [-ART] group, shows a gradual rise from the BA to the HM levels in Category I. In Category II, *a* occurs, surprisingly, even less than it did in the Spanish group. This suggests that the Chinese subjects have little problem determining when *a* is incorrectly used, reducing the competition to Ø and *the*. In Categories III and IV, *a* follows a zigzag course with Ø across the IL levels.
     In terms of acquisition strategies, the Chinese subjects appear to first reduce the amount of Ø-usage, which means they recognize the fact that English noun phrases must have a specifier. By the LM and MM levels, they realize that Ø can be a specifier too, and they start to increase their Ø-usage. The second major strategy is to flood environments with *the* to some extent. *The*-flooding is much more constrained in the Chinese group than it is in the Japanese group as it does not occur at all in Category IV and evinces smaller MM peaks in Categories I and III. In Category II, where *the* is the only correct article, *the*-flooding occurs at the LM level, which leads to increased accuracy. Such constrained flooding is perhaps not flooding at all in Huebner's (1983a) sense as it is almost too selective. Perhaps there is some parallel notion of definiteness in Chinese

Fig 8-3 CHINESE Usage Within Categories I-IV
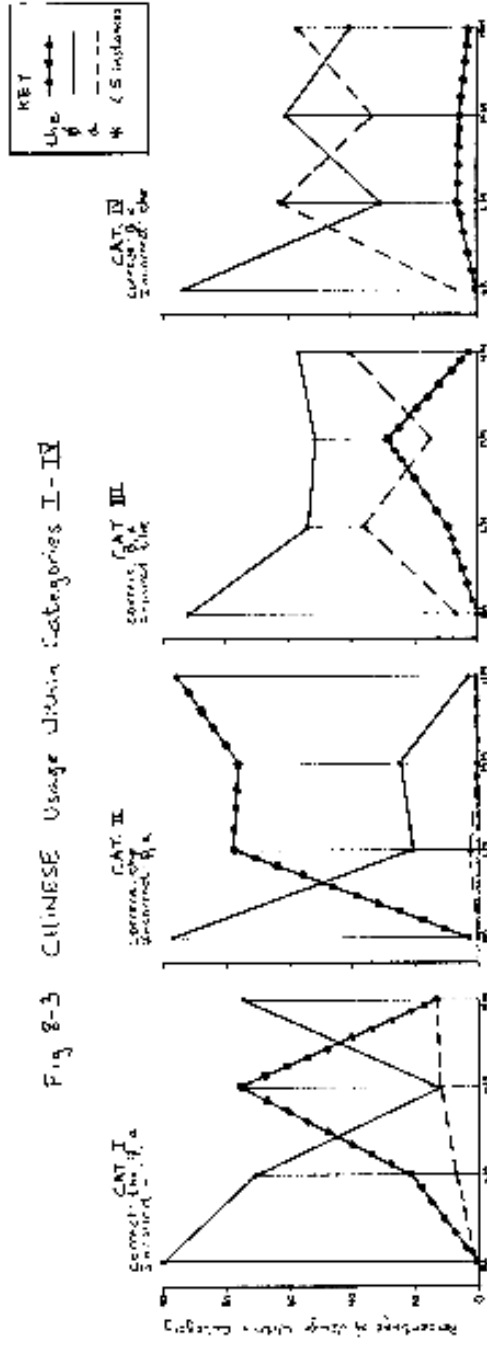
that fosters an understanding of English *the*-usage but which arises not at the outset as it does in the [+ART] group but rather later in the acquisition process (i.e., between the LM and MM IL level). Perhaps for this reason definiteness in Chinese is manifested outside the noun phrase. Further research is necessary to give any credence to these suggestions.

Japanese
        Japanese Category I usage, as shown in Figure 8-4, again provides a picture of unconstrained article usage. It shows extreme *the*-flooding at the LM stage and a higher use of *a* at the BA stage (the negation characteristics of the Japanese BA subject plus her overall fluency indicate her to be a rather high BA subject compared to the Chinese BA subject). Otherwise, Japanese Category I usage shows a similar pattern to that of the Chinese subjects: Ø plunges from a BA high to compensate for *the*-flooding and the percentage of use of all three articles at the HM level is almost identical.
        Category II usage is very similar to that of the Chinese subjects. The differences are the Japanese BA subject's usage of *a*, which decreases the amount of Ø-usage accordingly as *the*-usage is quite low, and the HM subject's persistent incorrect use of Ø in Category II (indicating, as does her negation, a lower HM than that of the Chinese HM subject), with a correspondingly lower usage of *the*, as *a*-usage is quite low.
        Looking across all four categories, we see consistent peaks of *the*-usage at the LM level. The Japanese LM subject is the *the*-flooder of this study, as mentioned many times before, and these data make an interesting statement about the subject's sense of the categories in which *the* is appropriate: *the* is used most freely in Category I, and increasingly less so across the four categories, with Category IV showing the least *the*-usage. As far as acquisition strategies are concerned, this suggests that there is something inherent in Category IV (indefinite generic noun phrases associated with the verb *to be* indicating equative, negated, interrogative and irrealis scope) that prohibits the usage of *the*, as is indeed the case for all the subjects, both [+ART] and [-ART], in this study. One explanation for this is that Category IV (described by the features [-HK] and [-SR] in Bickerton's semantic wheel) includes definitional contexts, which generally adhere to the Aristotelian definition formula, "An A is a B that C" (e.g., *A thermometer is an instrument that measures temperature*). It is quite possible that this definition formula is universal insofar as human beings recognize that in defining a species there can be no inclusion of a specific noun phrase, i.e., we define a species but we can only name or describe a specific object. This knowledge is probably not brought to bear at the BA level,

Fig. 8-4 JAPANESE Usage within Categories I - IV
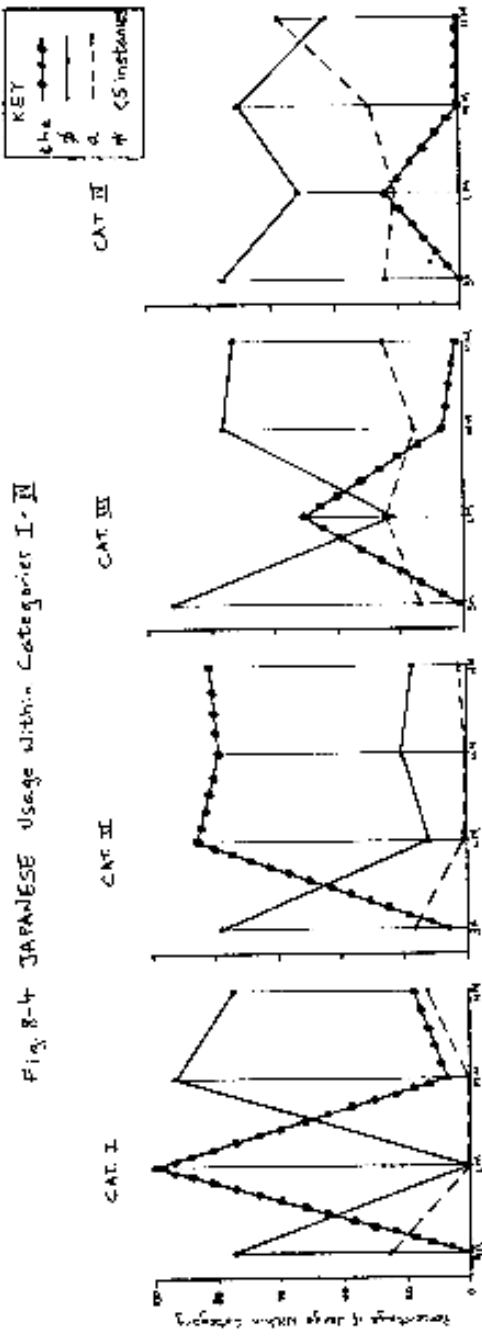
where *the* is simply not used by the [-ART] group. But the surge
of *the*-usage by all three [-ART] LM subjects suggests that by
this level, some concept of *the* as a definite marker must exist
in their interlanguages. Thus, even the *the*-flooder recognizes
(though to a lesser extent than any of the other subjects) that
definite *the* cannot be used when the noun phrase is not known to
the hearer ([-HK]) in Bickerton's schema), especially when the
noun phrase does not have a specific referent ([-SR] in
Bickerton's schema).

    Determining the appropriateness of *the* in Category III
([+SR][-HK] in Bickerton) is not so easy, and all three [-ART]
subjects evince peaks of *the*-usage at the LM (Japanese and
Russian) or MM (Chinese) levels that universally drop off by the
HM level. This would suggest that an association is first made
between *the* and [+HK] and that it is not until later that the
feature [+SR] is recognized as a second requirement for *the*
except in true generic environments (Category I). The fact that
the Chinese subjects do not decrease their *the*-usage until the
MM level suggests that they have more difficulty determining
when not to use *the* than the Japanese and Russian subjects.
This could be an individual variation (i.e., the Chinese MM
subject simply had more trouble with *the*); however, the fact
that the LM subject showed no peak of *the*-usage like the other
[-ART] subjects suggests that it is the L1 rather than
individual variation that accounts for the difference. Only
more research will resolve the uncertainty.

    It has already been noted that the Japanese BA subject uses
a more than the other [-ART] BA subjects. Having been eighteen
years in the U.S. (see Table 5-2), this subject is probably
fossilized, but perhaps a morpheme as frequent as *a* can be
acquired to a certain extent (her *a*-usage is still less than
half that of the German BA subject) before fossilization at the
BA stage. In any event, this subject floods *a* into all four
categories, indicating that she is aware of the morpheme but not
of the categories in which it is appropriate. *A* follows a
zigzag course of usage across Category III, as is the case for
all [-ART] subjects and even the German subjects (the Spanish
subjects would also evince a zigzag pattern if it were not for
the unexpected underusage of *a* by the HM subject). The zigzag
pattern I associate with the adjustment of the [±count] feature,
which must cope with a constant influx of new lexicon in
determining the distribution of *a* and Ø.

    Ø-usage shows a more or less U-shaped trajectory in all
categories except II, in which its use is incorrect and
consequently diminishes, much like Category II Ø-usage by the
other [-ART] subjects. This provides support for my suggestion
that the U-shaped usage pattern is characteristic of an element
that exists in the L1 but must be adjusted to fit the L2.

Russian

Category I usage by the Russian subjects, as shown in Figure 8-5, also functions as a microcosmic picture of unrestrained article usage as it did for the other subjects in this study. The one exception is the BA subject, who used no generic noun phrases whatsoever. Like the Chinese subjects, the Russian subjects (excluding the BA subject) show decreasing usage of Ø with a trough at the MM level corresponding to a peak of *the*-usage. *A* is not used at all until the HM level, but the percentages of article usage at the HM level are remarkably similar to the other [-ART] HM subjects, with Ø-usage at about 75% and *a*- and *the*-usage at around 15%.

Category II usage shows a steadily increasing usage of *the* with a peak at the MM level rather than at the LM level, which the Chinese and Japanese subjects had. This corresponds with a steadily declining usage of Ø, though there is a slight upturn at the HM level. Incorrect *a*-usage is almost entirely absent in this category, as was the case for most of the subjects in this study. This would suggest that *a* is not associated with the [+HK] feature until possibly the HM level, where it is finally realized that *a* can be associated with [+HK] in generic contexts only. This provides some support for the delayed pedagogical presentation of the concept of generic that will be taken up in Chapter 1O.

Curiously, although *the*-usage does not peak in Category II until the MM level, it peaks at the LM level in Category III, where *the* is incorrect. This LM peak is also slightly echoed in Category IV. As was the case for the Chinese subjects, it is difficult to say what it means when peaks of usage occur at different levels. One can only surmise that the Russian LM subject incorrectly associated *the*-usage with the [-HK] feature (the common characteristic of Categories III and IV), which may have caused her to flood with *the* to a lesser degree than the other [-ART] subjects at the LM level in Category II, since Category II requires [+HK]. However, if this were entirely the case, we should have seen much lower *the*-usage in Category II, not the over 5O%-usage that this subject showed. So perhaps this same subject did have some sense of the linkage between *the* and the [+SR] (specific referent) feature. This would account for her over 5O% *the*-usage in Category II and for her small peak of *the*-usage in Category III, which also shares the [+SR] feature. The connection between Bickerton's [+HK] and [+SR] features will be given a more global treatment in the next section of this chapter.

A-usage by the BA subject is also surprisingly high in Categories III and IV. Having been in the U.S. for 32 years, she too is probably fossilized. Like the Japanese BA subject, she must have acquired *a* to some extent despite her

93

Fig 8-5 RUSSIAN Usage Within Categories I-IV

94
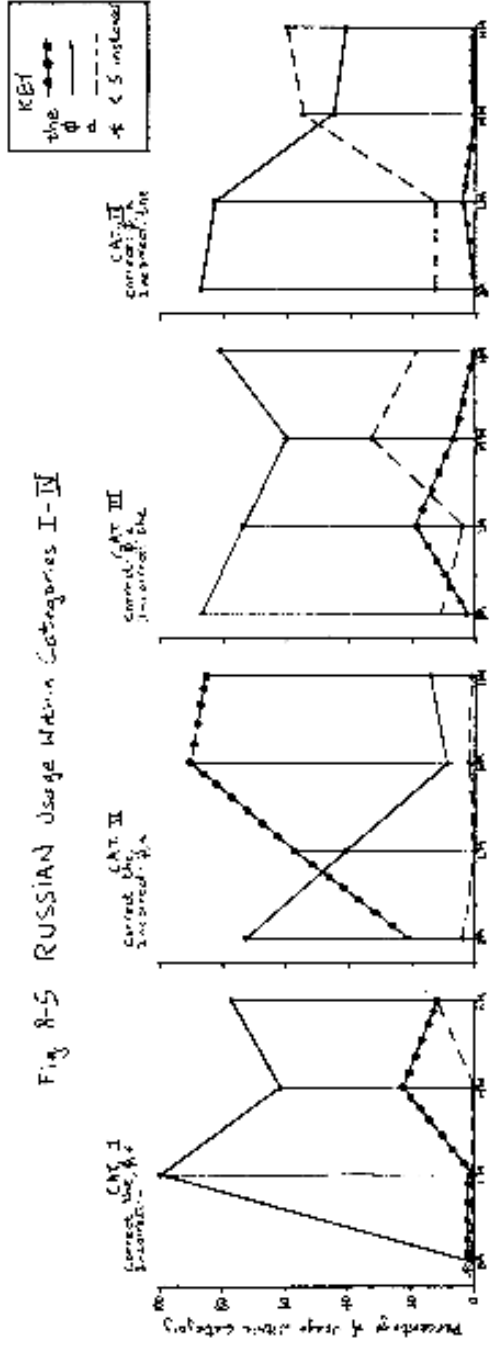
fossilization. But unlike the Japanese BA subject, she does not flood in every category but rather uses *a* only in Categories III and IV, where it is correct. These categories share the feature [-HK], and since she does not use *a* in Category I, where it is also correct, we can surmise that she associates *a* with the [-HK] feature. *A*-usage continues in a zigzag pattern across Category III, like all the other subjects, and in Category IV, a continuous increase in *a*-usage occurs, much like the Chinese subjects in the same category.

Ø-usage follows a U-shaped trajectory in Categories I, II, and III but not in IV. Ø is a feature of Russian and this pattern provides more evidence for the linkage of the U-shaped trajectory to the adjustment of an L1 characteristic.

Universal and Language-Specific Strategies

It is apparent from the preceding section that some of the strategies the subjects appear to use in acquiring the English article system are common to all five language groups, some common to members of the [+ART] group, some common to the [-ART] group, and some unique to a subject's language group or even to the individual subject. Bickerton's [+SR] (specific referent) and [+HK] (hearer known) features will be referred to in discussing these interlanguage systems.

Universal Strategies

Universal strategies are those that all five language groups appear to utilize in acquiring the article system. The broadest general strategy that can be identified is reflected in the U-shaped trajectories of Ø for the [-ART] group and *the* for the [+ART] group. The strategy is simply to reduce the usage of an element that occurs in the first language, somewhat analogously to Huebner's (1983a) concept of "trickling". This reduction in usage forms the left side of the U-shape; the right side, indicating increased usage of the same element, represents the apparent realization that the same element (keeping in mind the nothing-to-something Ø article for the [-ART] group) occurs in the target language.

The second universal strategy is reflected in the remarkably restrained use of *a* in Category II and *the* in Category IV by all five language groups. As shown in Table 8-1,

CATEGORY	FEATURES		DESCRIPTION
I	[-SR]	[+HK]	True generic (*the*, Ø, *a*)
II	[+SR]	[+HK]	Specific definite (*the*)
III	[+SR]	[-HK]	Specific indefinite (Ø, *a*)
IV	[-SR]	[-HK]	Indefinite generic (Ø, *a*)

Table 8-1 The Four Categories and Their Features

Category II is characterized by the features [+SR] and [+HK],
Category I by [-SR] and [+HK]. The fact that the [-ART] group
uses *the* to a greater degree in Categories I and II than in
Categories III and IV suggests that the first association made
is between *the* and [+HK]. This association leads to minimal
the-usage in the [-HK] categories III and IV.

   At the same time, the fact that Ø and *a* occur to a greater
extent in Categories III and IV suggests that an association is
made between the indefinite articles and [-HK]. This
association correctly leads to the lack of usage of *a* in the
[+HK] categories I and II and the lack of usage of *the* in the
[-HK] categories III and IV.

   The usage of *the* by the [+ART] group in Category II is very
similar to that of the [-ART] group from the LM level on. One
can thus surmise that the [-ART] group makes the association
between *the* and [+HK] between the BA and LM levels, leading to
the rapid increase in the usage characteristic (in this and
other studies) of the [-ART] group. At the same time, the
[+ART] group appears to realize at the same IL level that it
does not have to entirely do away with its L1 article system as
much of the usage of *the* in the L1 is appropriate in the target
language, so they reverse their tendency to decrease *the*-usage
and join the [-ART] group in increasing *the*-usage (correctly) in
Category II.

   The [+SR] feature differentiates generic from specific
usage, and the decreasing usage of *the* in Category I by the
[+ART] group suggests that the subjects are aware of a
difference between generic and specific but are not so
comfortable using generic *the*. It is not entirely fair to make
conclusions about generic usage based on Category I as there is
a considerable element of choice, which appears to reflect the
speaker's level of education. Thus, with the exception of the
Japanese *the*-flooder, who uses *the* indiscriminately in all
categories (though less in III and IV), the larger peaks of *the*-
usage in Category I are produced by the Chinese MM subject, a
professor of art history, and by the Russian MM subject, a
chemical engineer. Generally speaking, all the subjects tend to
use the generic Ø article more than any other, but this is
partly a result of the conversational nature of the speech
samples: Ø plus a pluralized countable or uncountable noun is
the less formal means of coding a generic noun phrase and
generic *the* is more a written than a spoken form unless a
formal, educated spoken register is adopted. That educational
experience can make a difference in generic article usage is
further exemplified by the fact that the three HM [-ART]
subjects use the three generic articles in remarkably similar
ratios (including 12-15% *the*-usage), and they have all attained
graduate levels of university education. The [+ART] HM subjects

are not so highly educated, and they do not use generic *the* at all.

The fact that generic usage occurs to a greater extent in the speech of the MM and HM subjects than in the BA and LM subjects suggests that the [+SR] feature does not arise as a serious consideration in article selection until these levels. The greater difficulty appears to lie in determining the boundaries of Category II and Category III, both of which share the [+SR] feature but differ in [+HK]. The [+ART] group uses *the* increasingly less as IL level increases, but the [-ART] group does not appear to realize the inappropriateness of *the* in Category III until some time after the LM level, and the Chinese subjects not until after the MM level. This whole process reflects Huebner's (1983a) notion of flooding followed by trickling.

Another universal characteristic is the zigzag shape of *a*-usage in Category III with generally higher *a*-usage at the higher IL levels. The only exception is the Spanish HM speaker, who has been described before as being somewhat anomalous in this study. This zigzag pattern I have associated in previous sections with the adjustment of the [±count] feature and the distribution of *a* with Ø.

[+ART] Strategies

The [+ART] group BA subjects show high *the*-usage in all categories. This usage generally drops with increasing IL level except in Category II, where *the* is correct. This is strong evidence of L1 transfer at the BA level, but it is apparently the definite article that is transferred and not so much the indefinite articles, even though the latter also exist in the [+ART] languages. This might be because, although the concept of *indefinite* exists in the L1, the [±count] feature must be adjusted to fit the target lexicon. A second minor strategy that the [+ART] subjects share is the replacement of *the* in Category II with Ø rather than *a*. This Ø must be interpreted as nothing rather than something, i.e., *the* is not being replaced with the zero article but is simply being used less often. This decreasing use of *the* is the only evidence in this study to support Bertkau's (1974) claim that simplification strategies predominate in article acquisition.

[-ART] Strategies

The most obvious [-ART] strategy is to flood all categories with *the*. This strategy was exercised to the greatest extent by the Japanese, to a lesser extent by the Chinese, and least by the Russian subjects. *The*-flooding occurred most commonly at the LM level but sometimes at the MM level, as shown in Table 8-2.

```
┌─────────────────────────────────────────────────┐
│ L1 [-ART] Low Mesolang (LM) Mid-Mesolang (MM)     │
│ Chinese       II,(IV)             I,III           │
│ Japanese      I,II,III,IV         ---             │
│ Russian       III,(IV)            I,II            │
│                                                   │
│ Note: Parentheses indicate minimal flooding.      │
└─────────────────────────────────────────────────┘
```

Table 8-2   Categories in which *the*-Flooding Occurred

The-flooding was not exercised by the [+ART] group except to a slight extent at the BA level. This is not flooding in Huebner's sense but rather transfer from the L1. A second strategy that the [-ART] subjects used was to gradually increase the use of *a* in Categories I, III, and, with the exception of the Chinese subjects, IV. The use of *a* is correct in these three categories. This strategy is also exercised by the [+ART] group if the German BA and the Spanish HM subjects are not included, making it a universal strategy. It was mentioned in Chapter 6 that *a* seems to be acquired almost independently of *the* and Ø as it cannot be applied without knowledge of the [±count] status of the noun phrase. Although Ø is also dependent on the [±count] assignment, it also indicates indefinite plural countable status and, in its "nothing" role, the failure to use any article at all, and is thus bound up with *the* as well as *a*.

Subject-Specific Strategies
     The most obvious subject-specific strategy is the Japanese LM subject's excessive use of *the* in all categories. She is excessive, however, only in her degree of flooding, as all the [-ART] subjects reveal flooding to a certain extent. It is in determining the universality of certain strategies that these pseudolongitudinal interlanguage analyses reveal their greatest weakness: since there is only one subject per IL level per L1, it is impossible to tell, in this case for example, whether all Japanese LM speakers will flood to the same extent. It should be remembered, however, that Huebner's (1983b) Hmong subject flooded with *the* to an even greater extent than the Japanese LM subject in this study. Does this mean that some [-ART] language groups flood with *the* to a greater degree than others? Or does it simply mean that some subjects flood with *the* more than others? True longitudinal analyses appear to be prey to the same weakness. The only way to solve this dilemma is to undertake multiple studies with subjects sharing the same L1, whether they be true or pseudolongitudinal analyses. This would allow a statistically valid comparison.
     Another apparent subject-specific strategy subject to the same dilemma as Japanese *the*-flooding is the Spanish and German BA use of *a*. In Categories III and IV, the Spanish BA subject uses *a* 0-10% (15-20% if incorrect *one* is included) whereas the

98

German BA subject uses *a* 45-70%. This difference can be interpreted in several ways: 1) the BA German subject's *a*-usage is abnormally high, 2) the Spanish subject's *a*-usage is abnormally low or 3) the Spanish subjects have more trouble using *a* than the German subjects and vice versa. Considering the fact that *a*-usage by the LM, MM, and HM subjects (with the exception of the anomalous HM Spanish subject in Category IV) is otherwise fairly similar, the question posed above is a serious one: Would other BA Spanish subjects show similarly low *a*-usage and other German BA subjects similarly high *a*-usage? An answer to the first question is suggested by Figures 7-3 and 7-4 in Chapter 7. In both of these figures, Spanish *a*-usage is shown to be quite low at the beginning levels. We can thus presume that the Spanish BA subject's *a*-usage is typical. Without such figures for German speakers of English, the question of the representativeness of the BA German subject's performance remains open. This is further complicated by the fact that the German BA subject speaks Swiss German, and in that dialect, a sound similar to "a" [namely /ə/] can be used in place of the indefinite articles (nominative: *ein*, *eine*) in a way that is not generally practiced in Standard German. Is it this dialectal usage that makes her *a*-usage so high? Once again, only multiple analyses can resolve the question.

A final language- rather than subject-specific strategy concerns the rate of removal (trickling) of *the* from Category III environments by the German subjects. The German BA subject's usage of *the* in this category is 16% whereas the Spanish BA subject's use of *the* is more than twice the amount at 36%. This usage is reduced by the German subjects to 2% by the LM level and stays at or below that level across all IL levels. The Spanish subjects, on the other hand, only reduce their *the*-usage to 6% by the MM level. This suggests that German speakers have less trouble identifying environments where *the* is not appropriate than the other [+ART] group.

Conclusion

There are strategies of English article acquisition that are used 1) by all five language groups, 2) by members of the [+ART] or [-ART] groups, 3) by the speakers of a particular language and 4) by a particular subject.

All five language groups show a U-shaped trajectory for an article that occurs in the first language. For the [+ART] group that article is *the*; for the [-ART] group that article is Ø. In addition, all five language groups show a zigzag pattern of *a*-usage (with the exception noted for the Spanish HM subject) in Category III, possibly reflecting the adjustment of the [±count] feature in distribution with Ø.

The [+ART] group shows the highest *the*-usage at the BA

level in all categories but II. In Category I, *the*-usage
generally decreases with increasing IL level. The [+ART] group
also shows a slight tendency to use Ø in Category II at the LM
level. The [-ART] group tends to flood with *the* in all
categories, especially in I and II. Furthermore, *a*-usage
generally increases with increasing IL level in Categories I and
IV (in which it is correct) for this group.

The only language-specific strategy is the German subjects'
tendency to remove *the* from Category III faster than their
Spanish counterparts. As far as the unique strategies of
particular individuals are concerned, the Spanish HM subject is
somewhat anomalous in using *the* too much in Categories III and
IV and *a* too little in Category III, and the Japanese LM subject
seems to flood with *the* to an inordinate degree. Finally, it
cannot be determined from the present study whether the German
BA subject has abnormally high *a*-usage in Categories III and IV
or whether this is a characteristic pattern for the English
interlanguage of native Standard German and/or Swiss German
speakers. Further research with multiple subjects is needed to
substantiate any of the conclusions of this chapter.

Nevertheless, article usage seems to function well as an
unconscious indicator of overall interlanguage level,
unconscious because the articles are used so frequently in their
English speech that the subjects have little time to dwell on
article choice. It is this very unconsciousness that allows a
measure of generalizability from a single subject. A more
precise proposal for using article usage as a potential
interlanguage measure is taken up in the next chapter.

An Interlanguage Grammar of Low Mesolang Article Usage
The present study presents and discusses the data in terms
of the article acquisition patterns of language groups as
manifested horizontally across IL levels. Since the major
activity in article acquisition appears to take place between
the basilang and low-mesolang levels for all language groups,
the data are here discussed vertically in terms of interlanguage
level rather than horizontally in terms of native language. It
has already been concluded that BA article usage appears to
reflect the native language of the subject (see Figure 6-9).
Thus, the [-ART] group shows a predominance of Ø usage, although
it has been explained that at the BA level, this Ø must be
interpreted as non-use of the article system rather than the
application of the zero article. The [+ART] group, on the other
hand, uses *the* to a great extent, especially in Categories II
(specific definite) and III (specific indefinite). Indefinite
a, on the other hand, was only used in an L1-like manner by the
BA German subject; the Spanish BA subject used *a* very little.
In describing an interlanguage grammar of article usage, it is

100

clear that the [+ART]/[-ART] distinction must be maintained.

[-ART] Low Mesolang Article Usage

At the LM level, all [-ART] subjects show a remarkable rise in *the*-usage with a concurrent decrease in Ø in the specific definite environment described by Category II (see Figure 6-1O). This represents an overall increase of 66% (from BA 9% to LM 75%) between these two interlanguge levels. Clearly, the LM subjects have shifted from a BA reliance on L1 Ø to a new hypothesis that marks noun phrases with *the*. This hypothesis appears to include the notion that in order to be marked with *the*, the noun phrase must describe something that is known to the hearer ([+HK]). For this reason, LM [-ART] subjects also use *the* with noun phrases that belong in the specific indefinite environment described by Category III. This also shows that the second feature of article usage, specific referent ([+SR]), appears to be considered less at the LM stage; otherwise, subjects would not use *the* in Category III.

With the dramatic rise in *the*-usage at LM, L1-like Ø-usage dramatically falls (BA 82% minus LM 23% equals 59%). It must be presumed therefore that the first hypothesis that a [-ART] subject makes about the article system is that *the* must be used before noun phrases. This accounts for the *the*-flooding shown by the Japanese LM subject in all categories. Within the same IL stage, this hypothesis is modified such that *the* is used with noun phrases describing things known to the hearer [+HK]. This accounts for the fact that the Chinese and Russian LM subjects do not flood with *the* in Category IV (indefinite generic usage) and that the Japanese subject floods less with *the* in this category.

At the same time, the notion of indefinite *a*-usage slowly begins to grow at the LM level. The [-ART] application of *a* is not nearly so haphazard as that of *the*. This is made quite clear in Figures 7-8 through 7-1O, which show that the distance between *a*-UOC (used in obligatory context) and *a*-SOC (supplied in obligatory context) is practically nil for all the [-ART] subjects. Such precision suggests that *a* is not used until its correct environment has also been determined. This explanation accounts for the remarkably diminished usage of *a* in Category II, the specific definite category in which *a* is not allowable in English.

It has been stated many times in this study that *a*-acquisition was generally slower than that of the other two articles and that this was probably because subjects required a certain amount of linguistic exposure before the [±count] feature could be correctly assigned. The precision described above suggests that *a* is possibly not used until some conception of the [±count] distinction has been acquired. This conception

begins to be built between the BA and LM stages and, as more
lexicon is acquired with its attendant [±count] features, it
continues to grow at a slow but constant rate, as shown in
Figure 6-15. It must also be assumed that with the gradual
acquisition of the [±count] feature, the English-like usage of Ø
must also be applied to [-count] noun phrases. The figures for
Ø-usage therefore reflect a combination of diminished use of L1
Ø and increasing use of target Ø.
    In summary, [-ART] low mesolang article usage can be
described as follows:

> *the*: 66%-usage in Cat. II shows that *the* has definitely
>   entered the lexicon (it was 9% at BA). This shows an
>   awareness that English NP's require a specifier
>   and suggests that an L1-independent system has been
>   created. *The* appears to be first associated with
>   [+HK].

>   Ø: 23%-usage in Cat. II shows that Ø has been largely
>   replaced (it was 82% at BA). This probably shows not
>   only decreasing L1-like Ø-usage but also an
>   awareness of target Ø-usage with [+count] plural and
>   [-count] noun phrases, although the ratio cannot be
>   precisely determined.

>   *a*: 24%-usage in Cat. III shows that *a* is used more than
>   it was used at BA (where it was 10%). This probably
>   reflects the beginning of a second L1-independent
>   system, i.e., the [±count] feature, and its marked
>   absence from Cat. II suggests that the [+SR] feature
>   of article usage may be at work to a certain extent
>   (Categories II and III are distinguished by [+SR].)

[+ART] Low Mesolang Article Usage
    At the LM level, the [+ART] subjects show a slight decrease
in *the*-usage with a concurrent increase in Ø in the specific
definite environment described by Category II (see Figure 6-10).
This represents on overall decrease of 12% (from BA 95% to LM
83%) between these two interlanguage levels. The LM subjects
appear to be rectifying their BA hypothesis that English noun
phrases must have a specifier (as is the case in their native
[+ART] languages) and that the article slot must be filled with
the.
    Like the [-ART] LM subjects, the [+ART] BA subjects appear
to link *the*-usage with [+HK], as *the* is used in both Category II
and Category III (where *the* is not allowable) but not in
Category IV (where *the* is also not allowable). By the LM level,
this notion has been largely corrected, although the Spanish LM

subject continues to use *the* in Category III to a greater extent than the German subject. Since the [+ART] languages by definition have an article system, the [+ART] subjects should not have to go through a "discovery phase" in which they learn that every English noun phrase requires a specifier. Indeed, already at the BA stage, the [+ART] group appears to make the hypothesis that *the* is the dominant article specifier, just as does the [-ART] group at the LM stage. Considering the claim in Chapter 7 that the [+ART] group appears to acquire the full article system by the MM level and the [-ART] group by the HM level, it could be said that the [-ART] group is, in some respects, one interlanguage level behind the [+ART] group as far as article usage is concerned.

The rectification of the [+ART] BA hypothesis described above essentially requires reducing the frequency with which *the* is used. Thus, [+ART] Ø-usage slightly (but incorrectly) increases in Category II. The [+ART] group does not have to build a new conception of a zero article the way the [-ART] group does because [+ART] languages already have such a concept. However, this group does have to correctly adjust the [±count] feature, so the amount of Ø-usage must be considered a reflection of the slightly reduced overuse of BA *the* plus the use of Ø with [-count] noun phrases.

The notion of indefinite *a* should also be familiar to the [+ART] subjects. As discussed earlier in this chapter, however, the Spanish BA subject treats *a* like a new lexical item. The BA German subject, on the other hand, uses *a* more than even the LM German subject does. Reasons for this have already been suggested. At the LM level, both [+ART] subjects use *a* to the same degree in Category III (32% combined), which is not so very different from that of the [-ART] group (24%). This suggests that both the [-ART] and the [+ART] groups have to grapple with the [±count] feature but that the latter has a head start, as it were, since the concept of an indefinite article is not new. In fact, [+ART] *a*-usage at the BA level is 24%, exactly that of the [-ART] group at the LM level. This is further support for the suggestion that the [-ART] group appears to be one interlanguage level behind the [+ART] group. The roughly parallel increase of *a*-usage between the two groups is evident in Figure 6-15.

In summary, [+ART] low mesolang article usage can be described as follows:

> *the*: 83%-usage in Cat. II shows that *the* is used less
>   than at the BA level (where it was 95%). The
>   hypothesis of the [+ART] group at the LM level
>   appears to be that *the* is used to a lesser degree
>   in English than it is in the L1. Thus, LM *the*-
>   usage appears to represent a language system that

is more target- and less L1-oriented, though
perhaps not as thoroughly as the [-ART] group.

Ø: 13%-usage in Cat. II shows that Ø is used more
than at the BA level (where it was 4%). Since
the-usage is decreased, Ø somewhat represents the
non-use of *the*. It probably also indicates a
growing understanding of the [±count] feature, and
to a greater extent than was true for the [-ART]
group.

a: 32%-usage in Cat. III shows that *a* is used more
than it was used at the BA level (where it was
24%). This probably reflects the further growth
of the L1-independent [±count] feature, and its
marked absence from Cat. II suggests that the
[+SR] feature of article usage may be at work to
a certain extent.

Chapter 9 An Interlanguage Measure Based on the Article System

   The interlanguage measure on which the present study is
based is verbal negation as described for Spanish speakers in
Cazden, Cancino, Rosansky and Schumann (1975). This measure was
satisfactory for determining the interlanguage level of the
subjects in this study because it was desired to gain a picture
of article acquisition across the entire interlanguage
continuum.

Problems with the Negation Criteria
   There is reason to believe, however, that the negation
criteria do not discriminate so well at the upper end of the
interlanguage continuum. One reason for this belief is based on
an interlanguage analysis of a Romanian speaker I undertook in
1985 (Master 1985). The predominance of analyzed *don't* negation
structures in this subject's speech placed him squarely at the
high mesolang (HM) level. Yet he exhibited no correct passive
or present perfect structures in his spontaneous speech, and he
still made numerous errors with the article system, leading me
to wonder whether this subject had been correctly classified.
The effectiveness of the negation criteria as an interlanguage
level discriminator seems to diminish somewhat at this level
(this may also account for the anomalous Spanish HM subject's
behavior and even for the HM Russian subject's unexpected
reduction in *the*-usage in Category II).
   A second reason to question the effectiveness of the
negation criteria at the HM level is that three HM subjects had
to be discarded in the course of conducting this study. After
checking their transcripts against the recorded interviews,
pulling and classifying the noun phrases for each subject,
tallying the data, and rendering the results in graphic form, it
was immediately apparent that these subjects were actually at a
lower IL level, compared to other speakers of the same native
language, in their article usage. The three subjects who were
rejected were native speakers of Russian (Ana), Polish (Sofia),
and Japanese (Mrs. Irie). They were all older women who had
been in the U.S. for a considerable period of time. A possible
explanation for their HM classification (which is based on the
predominance of analyzed *don't* in spontaneous speech) is that
over such a long period of English exposure, the negated
auxiliary verbs *doesn't* and *didn't* must have been picked up,
even though many other features of the language had become
fossilized at a lower IL level.
   These questions about the efficacy of the negation criteria
as an interlanguage level indicator led me to consider some
aspect of the article system as a possible replacement or at
least as a second indicator of IL level. The rest of this

105

chapter is devoted to the search for an element of article usage that may serve in this capacity.

An Article-Based Interlanguage Measure

The one basic requirement of an interlanguage measure (ILM) is that it increase with increasing interlanguage level. If there is a decrease (or, to use the words of Lamotte et al. 1982, if there is not a "smooth progression") in any pattern of accuracy or usage, that pattern must be rejected for the simple reason that a decrease would allow at least two possible placements from an ILM score, i.e., on either side of the peak, and there would be no way to determine which side was correct. In the present study, this means that patterns such as the widely-occurring U-shaped trajectory and the zigzag patterns associated with *a*-usage cannot be candidates for an ILM.

A careful perusal of all the figures generated in this study shows that there is only one that consistently increases with IL level irrespective of first language: the SOC (correct/required ratio) for all three articles in common (i.e., excluding proper) noun phrases considered together. This pattern is indicated by a double line in Figure 6-1.

Even though the common SOC increases for all five language groups, the score ranges must be considered in terms of the [+ART]/[-ART] distinction as the [+ART] group has been shown in this study to acquire the article system before the [-ART] group. The score ranges shown in Table 9-1 are determined by the lowest and highest common SOC scores attained within the IL level.

IL Level	[-ART] Range	[+ART] Range
BA	36 – 55	69 – 84
LM	55 – 71	72 – 87
MM	73 – 85	89 – 92
HM	83 – 92	95 — 97

Table 9-1 Actual Score Ranges for Common SOC

With the exception of the high BA [+ART] score, the score ranges are remarkably continuous with very little overlap. The high end of the BA [+ART] range is that of the German BA subject. It was suggested in Chapter 8 that the *a*-usage figures for this subject were perhaps abnormally high. The common SOC score may thus be a little high, too. For this reason, and because it suits the purpose of this chapter, I have lowered this score range to fit the other scores. With this and a few minor arbitrary changes to remove overlap and fill gaps, Table 9-1 is modified to Table 9-2.

IL Level	[-ART] Range	[+ART] Range
BA	O – 54	O – 71
LM	55 – 72	72 – 88
MM	73 – 84	89 – 93
HM	85 – 92	94 — 97

Table 9-2 Modified Score Ranges for Common SOC

Testing the Measure

Table 9-2 represents an interlanguage measure based on the SOC of articles used with common nouns. The obvious way to test this measure is to determine the common SOC score for a large number of interlanguage subjects and to see how well those scores fit this system in comparison with a subject's negation criteria, the 2O%-5O%-7O% morpheme acquisition criteria as described in Chapter 5, and, of course, subjective judgments of the subject's IL level.

While it is hoped that data such as that described above will one day be forthcoming, another way to test the article-based IL measure is fortunately at hand, namely the subjects who were rejected in the course of this investigation. Three of these subjects were described in Chapter 8: Ana (Russian), Sofia (Polish) and Mrs. Irie (Japanese). A fourth, Tomo (Japanese) was rejected because, although his negation criteria placed him at the LM level, he never used *the* and only rarely *a* and thus appeared to have an atypical interlanguage structure, which is somewhat confirmed by his 2O%-5O%-7O% morpheme acquisition data (see also Niimura 1986 for a treatment that includes Tomo). These four subjects all have [-ART] languages as their native tongues, and their common SOC scores are shown in Table 9-3.

Subject	NL	IL Level (neg-based)	Common SOC Score	IL Level (art-based)	N (# of NPs)
Tomo	JA	LM	43.6	BA	4OO
Ana	RU	HM	49.1	BA	38O
Irie	JA	HM	65.3	LM	4O6
Sofia	PO	HM	73.4	MM	374

Table 9-3 Common SOC Scores of Four Rejected Subjects

Table 9-3 shows that, according to the article-based IL criteria proposed in this chapter, the four subjects were placed at too high an IL level by their negation criteria. Ana may be anomalous as it is somewhat suspicious that the article criteria would place her at one end of the English IL continuum and the negation criteria at the other. In any event, to check the result of these subjects' revised IL classifications, their common SOC scores for each article are plotted against the same SOC scores of the subjects in this study who share the same native language (or, in the case of Sofia, a similar one). The

results are shown in Figure 9-1.

Figure 9-1 shows the degree to which the rejected subjects fit the common SOC pattern of subjects with the same L1. The placement of the four rejected subjects is determined by the total common SOC score (T). The closeness of fit can be estimated for each article by the shortness of the dotted line between the vertical placement line and the nearest intersection with the appropriate article SOC line. Each rejected subject is discussed below.

In the Japanese group, Tomo's total common SOC (T) is lower than the lowest total common SOC on the graph. For this reason, his vertical placement line is outside the confines of the figure. His *a*- and *the*-SOC are both lower than that of the BA Japanese subject and his Ø-SOC is higher. This places him squarely at the low BA level, as predicted by the article-based IL measure.

Mrs. Irie's total common SOC (T) puts her vertical placement line to the right of the LM column. However, her Ø- and *the*-SOC scores would tend to pull her closer to the LM column. Her *a*-SOC score, on the other hand, would pull her towards the MM column. In Chapter 8, it was noted that the BA Japanese subject, an older woman who had lived in the U.S. for 18 years, tended to use *a* more than expected. Mrs. Irie is also an older woman, and she has lived in the U.S. for 56 years. Since she could be expected to use *a* even more than the BA Japanese subject, the pull towards the MM column can be disregarded. This places Mrs. Irie closer to the LM level, as predicted by the article-based IL measure.

In the Russian group, Ana's total common SOC (T) puts her vertical placement line to the left of the LM column. Her high Ø-SOC would tend to pull her, like Tomo, to the BA column. Her *the*-SOC would also tend to pull her towards the BA column. Her *a*-SOC, however, would tend to pull her towards the MM column. Like the Japanese BA subject who used *a* more than expected, the Russian (actually Polish) BA subject, a woman who has lived in the U.S. for 32 years, also used *a* more than expected, so this tendency can be ascribed to Russian as well as Japanese native speakers. Although Ana has not been in the U.S. as long as the other women, there is reason to discount the strength of the *a*-SOC pull towards the MM column. This places Ana closer to the BA level, as predicted by the article-based IL measure.

Finally, Sofia's total common SOC puts her vertical placement line to the left of the MM column. With the shortest dotted lines of the four rejected subjects, Sofia fits all the common SOC patterns remarkably well. Her Ø-, *the*-, and *a*-SOC tend to pull her slightly away from the MM column though not out of its range. Furthermore, her total common SOC score of 73.4 places her at the bottom of the MM level according to the ranges
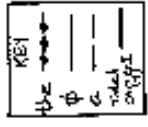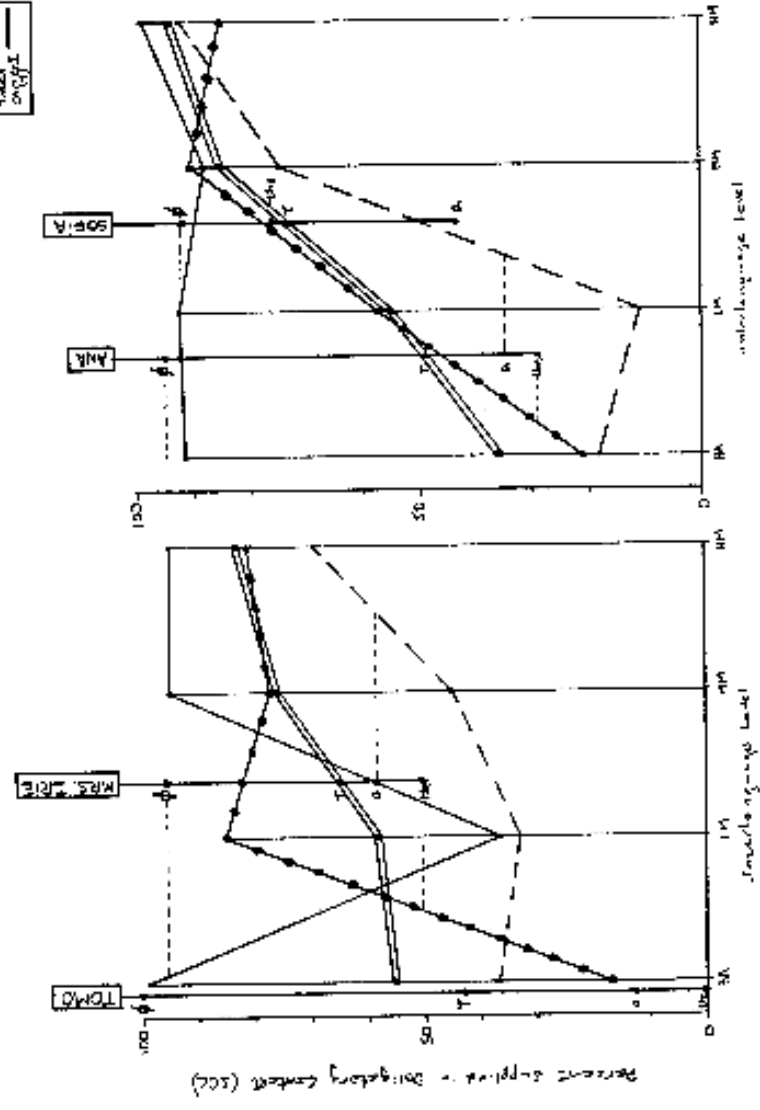
Fig. 5-1 How Rejected Subjects Fit into the Common SOC Patterns

109

(7O-83) presented in Table 9-2. Thus, she would be fairly likely to be assigned to the MM level, as predicted by the article-based measure.

Other Potential Article-Based Interlanguage Measures
     It was stated earlier that the only aspect of article usage that consistently increased with increasing IL level irrespective of first language was common SOC, as described above. However, there are two other aspects that, while not achieving the absolute consistency of common SOC, come reasonably close to achieving the same, i.e., they are consistent with some exceptions. These include 1) common *a*-SOC and 2) second mention *the*-SOC.
     Common *a*-SOC, or the degree of accuracy with which *a* was supplied in obligatory contexts, is first described in Figure 6-1. In this figure, it is apparent that only the Japanese subjects clearly show consistently increasing *a*-accuracy with increasing IL level. However, when the three [-ART] subjects are collapsed into a single group and the two [+ART] subjects are collapsed into a single group, their respective *a*-accuracies do increase consistently (and in a remarkably parallel manner) with increasing IL level, as shown in Figure 6-3. Thus, *a*-accuracy is a good second candidate for an interlanguage measure based on article usage. A replication of the present study with 5-1O subjects in each cell (instead of the single-subject cells on which this study is based) may well show that *a*-accuracy is a consistently reliable measure. This would be logical in light of the fact that *a* seems to be acquired more gradually than and somewhat independently of Ø and *the*.
     Second mention *the*-SOC, or the accuracy with which *the* is supplied in second mention contexts, would be a desirable IL measure because it is the only one based purely on discourse considerations, which some linguists consider to be the only true domain of article usage. Second mention *the*-SOC is described in Figure 6-6, Graph A. It is apparent from this figure that, with the exception of the HM Russian subject, the [-ART] group's second mention *the*-SOC increases with increasing IL level. If the individual [-ART] SOC scores are collapsed, they increase as follows: BA 12.O%, LM 8O.5%, MM 92.4%, and HM 92.9%. Second mention *the*-SOC is therefore a possible interlanguage measure, but only for subjects whose first language is Chinese, Japanese, or Russian, or ideally, any language that does not have an article system. The collapsed scores for the [+ART] group (BA 91.6%%, LM 84.6%, MM 91.7%, and HM 1OO%) do not increase with increasing IL level; second mention *the*-SOC is therefore not a possible interlanguage measure for this group.

Conclusion
     The only element of the article system that rises
consistently across all four IL levels irrespective of L1 and
therefore serves as the best interlanguage measure from the
present study is the total accuracy of articles with common noun
phrases, or total common SOC. The total common SOC scores of
all twenty subjects in this study were used to create separate
score ranges for the [+ART] and [-ART] groups.
     The efficacy of the ranges is then tested with four
subjects who were analyzed and then rejected from the present
study because their negation criteria placed them at too high an
interlanguage level. The new IL level predicted by the article-
based measure appears to place these four subjects at a position
that is more or less consistent with the other subjects in the
study. Once again, testing this measure with a statistically
appropriate number of subjects is required before any final
claims can be made. Nevertheless, the results of the
application of the measure are promising as 24 subjects (2O plus
the four rejected subjects) have been placed with a reasonable
degree of certainty at a point on the interlanguage continuum.
     Two other possible IL measures based on the article system
are *a*-SOC and [-ART] second mention *the*-SOC. The former needs
further research to substantiate its candidacy as an IL measure
and the latter can only apply as an IL measure to subjects whose
first language does not contain an article system. This
somewhat limits its usefulness.
     It is probably necessary to have a battery of tests that,
when applied to an interlanguage sample, can accurately assess
the IL level of the subject. The 2O%-5O%-7O% morpheme accuracy
levels described in Chapter 5 and the negation criteria on which
the present study is based would be included in such a battery.
The common SOC measure determined by the present study is a good
candidate for inclusion in the battery, and an *a*-SOC measure and
possibly a second mention *the*-SOC measure for [-ART] languages
may, with a larger replication of the present study, prove
themselves good candidates as well. Such a battery of placement
devices would help to avoid any individual variation or L1
interference that might skew a subject on one measure but not on
several (Celce-Murcia, personal communication). An index that
is a function of different measures could also be devised
(Kirsner, personal communication). Nevertheless, if the common
SOC measure does, with further research, prove itself to be
capable of accurately determining IL level for a variety of
subjects, it would certainly be a less time-consuming placement
device than either a battery of tests or an index would be.

Chapter 10 Pedagogical Implications and Suggestions for Further
                Research

     One of the reasons for undertaking the present study was to
provide a basis for a sound pedagogical approach to teaching the
English article system to nonnative speakers of English. Most
pedagogical approaches to teaching grammar depend on a perceived
hierarchy of difficulty, a natural and intuitive response to
human learning capability: teach the easy aspects first and
build upon them towards the more complex. This final chapter
discusses methods of teaching the article system in terms of
what this interlanguage analysis tells us about how the system
is acquired for the most part outside the pedagogical sphere.
It will conclude with suggestions for further research.

Pedagogical Implications
     Although there are numerous syntactic descriptions of
various aspects of the English article system (see "The Syntax
of the English Article System" in the References) and many texts
that provide a (sometimes overly complex) list of rules with
examples of article usage, there are comparatively few attempts
in the literature to provide a coherent grammar for teaching the
articles as a system. These include Whitman (1974), McEldowney
(1977), Pica (1983), and my own text (Master 1986a).
     Whitman (1974) bases his pedagogical sequence on the
assumption that English article structure is "a sequence of
quantification and determination rather than a choice between
specified and unspecified" (p. 253). In this manner, he
anticipates Jackendoff's (1977) focus on complex determiner
configurations (e.g., *both of the two white powders*) rather than
on the simpler elements of article usage. Whitman delineates
six steps for teaching the system:

    1) QUANTITY. Count nouns and *one* vs. *a.*
        Example: This is a book *vs.* These are (Ø)four books.

    2) GENERIC PLURAL
        Example: All apples are red *vs*. (Ø)Apples are red.

    3) NONCOUNT NOUNS. Noncount *vs.* count and *a lot of* vs.
                        much and *many*
        Example: He drank a lot of water *vs.* He bought a lot
                 of books.

    4) DETERMINERS. *Which*-NP questions and first/subsequent
                     mention.
        Example: Which books are red? The books on the table.
                 I read a book. The book was called Dracula.

5) QUANTITY AND DETERMINER.
    <u>Example</u>: One of the books on that table is blue.

6) GENERIC ARTICLES
    <u>Example</u>: A stone is a chunk of rock *vs.* (Ø)Stone is a
                        good building material.

    Whitman's sequence deals with the quantificational aspects
of article usage before the specified/unspecified distinction.
In spite of his claim that article structure is a sequence of
quantity and determination rather than a choice between
specified and unspecified, this choice must be made in Steps 4
and 5 (e.g., *Many a book has been lost*). He presents the
generic articles last. The finding in the present study that
the [±count] feature appeared to cause the most persistent
difficulty in article acquisition and that *the* appears to be the
first of the articles with which the [-ART] group makes
interlanguage hypotheses warrants placing [±count] and
[±definite] distinctions before [±generic] in a pedagogical
sequence. The finding that accuracy and usage of the generic
articles appears to depend on education and experience plus my
hypothesis (Master 1984) that generic articles have a particular
function in the structure of written discourse suggest that the
generic is relatively less salient than the specific realm and
justifies its placement at the end of the instructional
sequence. In sum, Whitman's sequence, [±count] > [±definite]
>[±generic], is supported by the implications of the present
study.
    McEldowney (1977) takes more of a form/content approach to
the teaching of the articles. She says that four types of
meaning are communicated by the presence or absence of *a*, *the*,
or *-s* in various combinations in noun phrases: a) general or
particular, b) any or special, c) countable or uncountable, and
d) singular or plural. She then cites three universal types of
error which she claims occur irrespective of L1: a) omission of
a/*the*/*-s*, b) wrong insertion of *a*/*the*/*-s*, and c) confusion of
a/*the*/*-s*. With these two taxonomies in view, McEldowney
proposes the following "stages of learning":

1) CLASSIFICATION.
    a + N (any one) vs. *the* + N (the special one).
      <u>Example</u>: Choose a bag. vs. Take the red bag.
    N + *-s* (plural classification)
      <u>Example</u>: These are bags.

2) PLURALITY
    some + N + *-s* (any ones)
      Example: Choose some bags from the collection.
    the + N + *-s* (the special ones)
      Example: Take the red bags.

3) MASS OR SUBSTANCE
    N (the substance in general)
      Example: Mud is found at the bottom of rivers.
    some + N (any substance)
      Example: There was some mud on his shoes.
    the + N (the special substance)
      Example: Point to the black mud.

4) NUMBERED SPECIFIC; GENERIC
    numeral N + *-s* (any numbered ones)
      Example: Choose six pens from the collection.
    a + N
    the + N   (ones in general)
    the + N + *-s*
      Examples: An elephant never forgets.
               The elephant never forgets.
               The elephants never forget.

    McEldowney's sequence links the English articles to the three concepts of *any* (*a*) to mark choice, *special* (*the*) to mark specification, and *general* (*-s* and later *a* and *the*) to mark generalization. Her sequence can be characterized as [±definite] > [±count] > [±generic] and as such differs from Whitman's sequence by inverting the first two elements. Her neglect of the zero article, in my opinion, places an added burden on the learner. It is, in my experience, easier for students to grasp Ø as an article that stands in opposition to *a* than as an opposition to both *the* and *a*. The former defines Ø as an article with specific domains of usage whereas the latter fosters the unproductive generalization that sometimes you use the article, sometimes you don't, thereby undermining the very concepts that McEldowney hopes to inculcate. Teaching the concept of the zero article may help the learner to correctly identify the environments (e.g., Category II) in which only *the* (and not Ø) is correct, perhaps leading to reduced flooding, especially by the [-ART] group, in inappropriate categories. The fact that flooding occurs at all, however, suggests that learners try to increase their usage of the articles (especially *the*) because they hear them so frequently, particularly at the LM level and beyond. Their first hypotheses probably try to link the article to a specific meaning (HK, "known to the hearer" was suggested in Chapter 8), and it is in this regard

that McEldowney's notions of *a* = any and *the* = special might be useful. To this writer, her sequence of presentation seems overly complex, especially the first stage distinction between *a* and *the*, as *the* requires, in addition to the "special" status that an object can take on in a specific environment, an awareness of the whole domain of shared knowledge and implied postmodification, which McEldowney's simple tasks and sentence models exclude. She claims that with her framework as a basis, "blends" can be made from "the situation [of the moment] into the world...As long as the three concepts of 'any', 'special' and 'general' are kept clear throughout stages I to IV, the blends occur without much special effort on the part of either teacher or learner. Any actual practice exercise that is felt necessary for 'in the world' only succeeds if the three usage areas are again controlled" (p. 111).

From the findings of the present study, there is no clear reason to delay the teaching of the [±definite] distinction until after the [±count] distinction, especially as the latter appears to take longer to acquire. The greatest motivation for teaching the [±count] distinction first is a pedagogical one: since *the* can occur with both singular and plural nouns, the [±count] distinction must be made first in order to correctly mark the noun with *-s*. Furthermore, natural texts can be found (or acceptable ones created) wherein only Ø and *a* occur whereas texts with *the* alone would be rather unnatural.

Pica (1983) argues not for a new pedagogical sequence but for the inclusion of discourse-related rules in the presentation of the English article system. She based her research on a perusal of the kinds of article rules typically presented in ESL grammar texts and compares them with the article use of native speakers in requesting and giving directions and ordering food at restaurants. She concludes that "article use may have more to do with communication and communicative competence than with grammar and linguistic competence" (p. 231) and makes the following recommendations for instruction:

1) Since articles are often not necessary in immediate environments, activities like ordering food should be practiced first "as a non-frustrating lesson for beginning students."

2) First mention *a* and subsequent mention *the* are easy to teach from a pedagogical point of view but are not used as frequently as proforms in natural speech.

3) Since assessing the knowledge of the hearer is often no simple matter, students should be encouraged to always use *the* with a qualifying description rather than just

115

a bare noun (e.g., *the nearest post office* vs. *the post office*; *the university bookstore* vs. *the bookstore*).

4) dialogs should be used to provide students with relevant examples of article use and the effect of using an incorrect article should be discussed with the class to increase awareness of native usage.

5) Students should be engaged in experiences outside the classroom to foster natural acquisition.

Pica's points are generally well taken, especially if spoken communicative competence is the goal and especially for students at lower levels of proficiency. I question whether taking students to a restaurant in order to give them "a nonfrustrating lesson" in the fact that correct article usage is not usually necessary in an immediate environment has any value other than a gastronomic one.   But I generally agree that the articles should be taught like many other aspects of syntax, i.e., with sufficient room for inductive generalization from natural speech and other elements of the communicative approach, and not just with a list of vague and often confusing rules.
     With more advanced students, especially if the goal is written competence (where article errors really stand out), Pica's suggestions would be helpful but insufficient, basically because the articles are still presented as a list of rules, albeit discourse-based ones, and not as a complete system.
     The pedagogical sequence for teaching the article system in my own text (Master 1986a) was presented in Chapter 6 as the basis for the classifications that were used in tallying the data. That sequence is based on the answers to a set of six questions that must be asked before assigning the article to any noun that occurs in discourse:

1) Is the noun countable or uncountable?
2) Is the noun indefinite or definite?
3) Is the noun postmodified or not?
4) Is the noun generic or specific?
5) Is the noun common or proper?
6) Is the noun idiomatic or not?

The purpose of these questions is to make intermediate to advanced ESL students aware of the system and not just of individual article use. It basically adheres to the same general sequence posed by Whitman (1974), [±count] > [±definite] > [±generic], but adds the dimensions of [±proper] and [±idiomatic] to fill out the whole domain of article usage.

116

In a quasi-experimental pretest/posttest control group design (Master 1986b), I tested the efficacy of using this pedagogical sequence in a classroom setting. After approximately six hours of instruction spread over an eight-week period, the experimental group showed a significant ($p < .05$) increase between pretest and posttest scores whereas the control group showed no significant increase. This suggests that systematic instruction can foster accelerated learning of the article system.

The three classes that comprised the control group made no significant increase despite the fact that teachers corrected article errors on student compositions and, in one case, even assigned an article exercise for homework. This suggests that the article system must be taught systematically over a considerable period of time in order to induce accelerated learning. Otherwise, improvement in article usage appears to occur at its own rate, as is evident from the present study. For the control group in the pretest/posttest study, this rate was 3.6 times slower than that of the experimental group.

In conclusion, the sequence [±count] > [±definite] > [±generic] appears to produce the best results in improving learners' control of the English article system. Based on the present study, it seems wise to provide students with [-ART] native languages more instruction in the use of *the* than the [+ART] students, if only because the latter would find such instruction obvious and boring (except in the subclassifications of ranking adjectives and shared knowledge of uniqueness, as shown in Figure 6-6). On the other hand, both groups can benefit from continuous exposure to the [±count] features of the noun phrases they encounter as well as practice with the subtleties of generic usage. However, as a general principle, the articles must be taught as a system over a considerable period of time in order to bring about a more rapid improvement than normal exposure to the language will produce.

Suggestions for Further Research

The greatest need for further research, as has been mentioned throughout this dissertation, is the replication of the present study with a much larger number of subjects. In other words, there should be at least five if not ten or more representatives of each language group at each IL level (i.e., 100 to 200 subjects). This would allow the use of statistical procedures and should provide significant results as opposed to the rather speculative ones that characterize this interlanguage analysis of article usage.

To test the validity of the proposed article-based interlanguage measure, it would be useful to determine the percentage of correct articles with common noun phrases (common

SOC) in a number of interlanguage analyses of spoken discourse that have been classified by the traditional negation criteria, and then to see if the article-based measure predicts the same interlanguage level. At the upper mesolang levels, however, the results from the four rejected subjects in this study should be borne in mind: the negation characteristics of fossilized speakers (particularly older ones) may place them at too high an interlanguage level. The actual ranges of common SOC scores provided by this study should first be confirmed and, if necessary, adjusted by the results of the larger replication proposed above. In order to have sufficient data to carry out these suggestions, it is essential that applied linguistics courses in interlanguage analysis include article analysis in their syllabi in order to build up the data base in this potentially useful area of research.

In terms of the efficacy of pedagogical treatments of the article system, a thorough contextual analysis of article usage in both spoken and written discourse is needed in order to identify the frequency with which native speakers of English use the various aspects of the article system. Keeping in mind the necessity to teach the whole system, such an analysis would help to assess the value of time spent on any one aspect. Pica (1983) has begun this research with her discourse-based analysis of article usage in speech acts such as ordering food in restaurants and asking for and giving directions. Similar research is required in other domains of spoken as well as written discourse. A thorough contextual analysis of article usage coupled with the results of a wider replication of the present study would provide a very solid basis on which to build a sound pedagogical presentation of the English article system.

REFERENCES USED IN THE DISSERTATION

Andersen, R. 1977. The impoverished state of cross-sectional
     morpheme acquisition/accuracy methodology. Paper presented
     at the Los Angeles Second Language Research Forum (SLRF).

Bailey, C. J. N. 1973. *Variation and linguistic theory.*
     Arlington, VA.: Center for Applied Linguistics.

Bailey, N., Madden, C., & Krashen, S. 1974. Is there a
     natural sequence in adult second language learning?
     Language learning 24.2:235-243.

Berko, J. 1958. The child's learning of English morphology.
     Word 14:15O-177.

Bertkau, J. 1974. An analysis of English learner speech.
     Language learning 24:279-286.

Bickerton, D. 1975. Creolization, linguistic universals, natural
     semantics and the brain. Paper presented at the
     International Conference on Pidgins and Creoles,
     University of Hawaii, January 6-11, 1975.

Bickerton, D. 1984. The language bioprogram hypothesis.
     Behavioral and brain sciences 7:173-88.

Bowerman, M. 1982. Starting to talk worse: clues to language
     acquisition from children's late speech errors. In S.
     Strauss (ed.) *U-shaped behavioral growth*. New York:
     Academic Press.

Brown, R. 1973. *A first language*. Cambridge, Mass.: Harvard
     University Press).

Brown, R., C. Cazden & Bellugi-Klima, U. 1968. The child's
     grammar from I to III. *Minnesota symposia on child
     development*, v. 2, pp. 28-73. Minneapolis: University of
     Minnesota Press,

Cancino, H., Rosansky, E.J., & Schumann, J.H. 1978. The
     acquisition of English negatives and interrogatives by
     native Spanish speakers. In E. Hatch (ed.) *Second language
     acquisition*, pp. 2O7-23O. Rowley, Mass.: Newbury House
     Publishers.

Cziko, G.A. 1987. Testing the language bioprogram hypothesis: a review of children's acquisition of articles. *Language 62*,4:878-898.

de Villiers, J. & de Villiers, P. 1973. A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of psycholinguistic research 2*:267-278.

Dulay, H.C. & Burt, M.K. 1974a. Natural sequences in child second language acquisition. *Language learning 24*.1:37-53.

Dulay, H.C. & Burt, M.K. 1974b. A new perspective on the creative construction process in child second language acquisition. *Language learning 24*:253-278.

Dulay, H.C., Burt, M.K. & Hernandez, E. 1973. *Bilingual syntax measure* (Restricted Edition). New York: Harcourt Brace.

Garcia, E.C. 1975. Chapter II: The theory. *The role of theory in the linguistic analysis of the Spanish pronoun system*, pp. 38-58. Amsterdam: North-Holland Publishing Co.

Hakuta, K. 1976. A case study of a Japanese child learning English as a second language. *Language learning 26*:321-351.

Hatch, E.M. 1978. Acquisition of syntax in a second language. In J. Richards (Ed.) *Understanding second and foreign language learning: issues and approaches*, pp. 34-7O. Rowley, Mass.: Newbury House Publishers.

Huebner, T. 1979. Order-of-acquisition vs. dynamic paradigm: a comparison of method in interlanguage research. *TESOL quarterly 13*:21-48.

Huebner, T. 1983a. Linguistic systems and linguistic change in interlanguage. *Studies in second language acquisition 6*:39-52.

Huebner, T. 1983b. *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma Publishers.

Jackendoff, R. 1977. *X-bar syntax: a study of phrase structure.* Cambridge, Mass.: MIT Press.

Kellerman, E. 1984. If at first you do succeed... In S. Gass and N. Madden (eds.) *Input in second language acquisition*, pp. 345- 353. Rowley, MA: Newbury House.

Klima, E. S. & Bellugi-Klima, U. 1966. Syntactic regularities in the speech of children. In A.Bar-Adon and W.F. Leopold (eds.) *Child language: a book of readings*, pp. 412-424. Englewood Cliffs, NJ: Prentice-Hall.

Lamotte, J., Pearson-Joseph, D. P, & Zupko, K. 1982. A cross-linguistic study of the relationship between negation stages and the acquisition of noun phrase morphology. Unpublished paper, University of Pennsylvania.

Larsen-Freeman, D. 1978. An explanation for the morpheme accuracy order of learners of English as a second language. In E. Hatch (ed.) *Second language acquisition*, pp. 371-379. Rowley, Mass.: Newbury House Publishers.

Maratsos, M.P. 1976. *The use of definite and indefinite reference in young children*. Cambridge, Mass.: Cambridge University Press.

Master, P. 1985. An interlanguage analysis of a high-mesolang Romanian speaker. Unpublished paper, UCLA.

Master, P. 1986a. *Science, medicine, and technology: English grammar and technical writing*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Master, P. 1986b. Measuring the effect of systematic instruction in the English article system. Unpublished paper, UCLA.

Master, P. (1987). Generic *the* in *Scientific American*. *The ESP journal 6*.3:165-186.

McEldowney, P. L. 1977. A teaching grammar of the English article system. *IRAL 15*.2:95-112.

Neuman, R.A. 1977. *An attempt to define through error analysis the intermediate ESL level at UCLA*. MA-TESL thesis, UCLA.

Niimura, T. 1986. *The English interlanguage of a native Japanese speaker: temporality and transfer*. MA-TESL thesis, UCLA.

Parish, B., Tarone, E., & Taghavi, R. 1986. Article usage in interlanguage: A study in task-related variability. Paper presented at the 1986 TESOL convention.

Piaget, J. 1926. *The language and thought of the child*. New York: Humanities Press.

Pica, T. 1983. The article in American English: What the textbooks don't tell us. In N. Wolfson and E. Judd (eds.) *Sociolinguistics and language acquisition*, pp. 222-233. Rowley, Mass.: Newbury House Publishers.

Scollon, R. 1976. *Conversations with a one-year-old: a case study of the developmental foundation of syntax*. Honolulu: The University Press of Hawaii.

Selinker, L.1972. Interlanguage. *IRAL 10*:209-231.

Stauble, A.M. 1977. *An exploratory analogy between decreolization and second language learning: negation*. MA-TESL thesis, UCLA.

Stauble, A.M. 1981. A comparison of the Spanish-English and Japanese-English interlanguage continuum. Paper presented at the European-North American Workshop on Cross-Linguistic SLA Research, Lake Arrowhead.

Tarone, E. 1985. On chameleons and monitors. Paper presented at the Los Angeles Second Language Research Forum (SLRF).

Weisenburger, J. 1976. A choice of words: Two-year-old speech from a situational point of view. *Journal of child language 3*:275-281.

Whitman, R.L. 1974. Teaching the article in English. *TESOL quarterly 8*.3:253-262.

Yamada, J. & Matsuura, N. 1982. The use of the English article among Japanese students. *RELC journal 13*.1:50-63.

Zehler, A.M. & Brewer, W.F. 1982. Sequence and principles in article system use: an examination of *a*, *the*, and null acquisition. *Child development 53*:1268-1274.

LINGUISTIC DESCRIPTIONS OF THE ENGLISH ARTICLE SYSTEM

Aldridge, M. 1982. *English quantifiers.* U.K.: Avebury.

Burton-Roberts, N. 1976. On the generic indefinite article. *Language 52*.2:427–448.

Celce-Murcia, M. and Larsen-Freeman, D. 1983. *The grammar book*. Rowley, Mass.: Newbury House Publishers).

Chafe, W. 1969. English noun inflection and related matters from a generative semantic point of view. *POLA Report 2-6*. Berkeley: U.C. Berkeley).

Chafe, W. 1972. Definiteness. In J.B. Carroll, R.O. Freedle, & Committee on Basic Research in Education (eds.), *Language comprehension and the acquisition of knowledge*. Washington: V.H. Winston.

Chafe, W.. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (ed.), *Subject and topic*. New York: Academic Press.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press..

Christophersen, P. 1939. *The articles: a study of their theory and use in English*. Copenhagen: Munksgaard.

DuBois, J. W. 1980. Beyond definiteness: the trace of identity in discourse. In W. Chafe (ed.), *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Ellis, T. 1984. A review of "definiteness". Unpublished paper, UCLA.

Frank, M. 1972. *Modern English: a practical reference guide.* Englewood Cliffs, NJ: Prentice-Hall.

Garcia, E.C. 1975. Chapter II: The theory, pp. 38–58. *The role of theory in the linguistic analysis of the Spanish pronoun system.* Amsterdam: North-Holland Publishing Co.

Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London: Longman Group.

Heringer, J.T. 1969. Indefinite noun phrases and referential opacity. *CLS 5*:89-97.

Hewson, J. 1972. *Article and noun in English*. The Hague: Mouton.

Jackendoff, R. 1977. *X-bar syntax: a study of phrase structure.* Cambridge, Mass.: MIT Press.

Jespersen, O. 1924. *A modern English grammar* (completed and published by Niels Haislund in 1949). Copenhagen: Munksgaard.

Karttunen, L. 1968. What makes noun phrases definite? *Paper p-3871.* Santa Monica, CA: The Rand Corporation

Kirsner, R. 1979. The article system. *The problem of presentative sentences in modern Dutch*. Amsterdam: North-Holland Publishing Co.

Kramsky, J. 1972. *The article and the concept of definiteness in language*. Paris: Mouton.

Kruisinga, E. 1932. *A Handbook of present day English*, v. 2. Groningen: Noordhof.

Lackstrom, J. E., Selinker, L., and Trimble, L. 1973. Technical rhetorical principles and grammatical choice. *TESOL quarterly* 7.2:127-136.

Perlmutter, D. 1970. On the article in English. M. Bierwisch & K. Heidolph (eds.). *Progress in linguistics*, pp. 233-248. Paris: Mouton.

Postal, P. 1966. "On So-Called Pronouns in English", in F. Dineen S.J. (ed.), *Report of the 17th annual round table meeting on languages and linguistics*, Georgetown University Press, Washington, D.C. Reprinted in D. A. Reibel and S. A. Schane (eds.) (1969), *Modern studies in English: readings in transformational grammar*. Englewood Cliffs, New Jersey: Prentice- Hall.

Poutsma, H. 1914. *A grammar of late modern English*. Groningen: Noordhof.

Quirk, R., S. Greenbaum, Leech, G. & Svartvik, J. 1972. *A grammar of contemporary English*. London: Longman Group Inc.

Seuren, P. 1969. *Operators and nucleus*. Cambridge: Cambridge
      University Press).

Sloat, C. 1969. Proper nouns in English. *Language 45*.1:26–30

Smith, C. S. 1964. Determiners and relative clauses in
      generative grammar of English. *Language 40*:37–52.

Sommerstein, A. H. 1972. On the so-called definite article in
      English. *Linguistic inquiry 3*.2:197–210.

Stern, S. 1977. Generic noun forms-- usage differences.
      Unpublished paper, UCLA.

Van der Auwera, J. (ed.). 1980. *The semantics of determiners.*
      Baltimore: University Park Press.

Zandvoort, R.W. 1968. *A handbook of English grammar*. London:
      Longman Group Inc.