

City University of New York (CUNY)

CUNY Academic Works

All Dissertations, Theses, and Capstone
Projects

Dissertations, Theses, and Capstone Projects

6-2014

Some complexities in English article use and acquisition

Victoria Somogyi

Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/gc_etds/286

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu

SOME COMPLEXITIES IN ENGLISH ARTICLE USE AND ACQUISITION

by

VICTORIA SOMOGYI

A master's thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2014

©2014
VICTORIA SOMOGYI
All Rights Reserved

This manuscript has been read and accepted for the
Graduate Faculty in Linguistics in satisfaction of the
dissertation requirement for the degree of Master of Arts.

Gita Martohardjono

Date

Thesis Advisor

Gita Martohardjono

Date

Executive Officer

THE CITY UNIVERSITY OF NEW YORK

Abstract

SOME COMPLEXITIES IN ENGLISH ARTICLE USE AND ACQUISITION

by

VICTORIA SOMOGYI

Adviser: Gita Martohardjono

Along with preposition use, native-like article use is one of the greatest difficulties for learners of English, particularly for those learners whose first languages do not have articles. And unlike many other areas involved in the mastery of the language, articles continue to present a challenge for advanced learners. A significant number of article usages are complex in that they are neither simple to explain (in that *the* does not encode definiteness or contextual uniqueness) nor strictly idiomatic. This paper a) provides an overview and critique of the scholarship on articles and their acquisition, b) takes a detailed look at several of the complex usages (noun phrases with genitive phrases, modified by the adjective “wrong,” or with a possible generic reading), and c) examines article acceptability judgments by English language learners with article-less L1s and by native speakers, both college students and adults.

TABLE OF CONTENTS

Introduction	1
Theories on the definite article.....	3
Theories on article learning	6
The study.....	15
The results.....	31
Conclusion	39
References	47

LIST OF TABLES

Table 1. Weak definites	31
Table 2. Noun phrases with genitive phrases	33
Table 3. Noun phrases with <i>wrong</i>	34
Table 4. Noun phrases with <i>people</i>	35
Table 5. Noun phrases with commercial institutions	37
Table 6. Noun phrases with a normally uncountable noun	38

LIST OF FIGURES

Fig. 1. Google Books n-gram showing the occurrences of <i>a wrong one</i> and <i>the wrong one</i> over time.	35
--	----

LIST OF APPENDICES

Appendix A – Questionnaire 42

Appendix B – Participants 46

1. Introduction

Along with preposition use, native-like article use is one of the greatest difficulties for learners of English, particularly for those learners whose first languages do not have articles. And unlike many other areas involved in the mastery of the language, articles continue to present a challenge for advanced learners. In fact, these small determiners probably play a large role in advanced learners' feeling that native-like English is an elusive goal.

But articles are more than tiny words with the ability to mar the fluency of a conversation or text. They play a significant role in the communicative function of language. They can differentiate phonetically identical words: *having an experience* is very different from *having experience*; *having a little money* is much nicer than *having little money*. They can establish presuppositions: a study on eyewitness testimony and memory found that witnesses are far more likely to say that they remember something if the questioner's *the* suggests its existence. *Did you see the broken headlight?* will receive more affirmative answers than *Did you see a broken headlight?* in both situations where the witnesses saw, in reality, such a headlight and ones in which they didn't (Loftus, 1996). And, more subtly, they can convey assumptions, emphasis and the speaker's conceptualization of a noun phrase.

A number of factors affect English language learners' ability to use articles in a native-like way. Studies have suggested many reasons for these problems: learners may make article errors due to misunderstanding the countability of the noun in the noun phrase, being unclear whether *the* encodes definiteness or specificity, viewing articles as adjectives and thus unnecessary if inessential to meaning. (White, 2009; Ionin, 2006; Yang & Ionin, 2009; Trenkic, 2008) Understandably, these studies have focused on article choices that are unambiguous for native speaker, and they have placed idioms into a separate category of article usage.

However, a significant number of article usages are neither simple to explain (in that *the* clearly encodes definiteness, or contextual uniqueness) nor strictly idiomatic. I wish to examine some of these uses, particularly of the definite article, *the*. Can these productive but complex uses be explained? Are native speakers uniform in their use of articles in these situations? Does explicit teaching play a significant role in learners' acquisition of articles? And what other factors influence their acquisition? These questions are part of a larger question: why are articles so difficult, and can a native-like level of use be attained by non-native speakers?

Many linguists and language theorists have discussed the meaning and function of articles in English. Although they have generated a variety of explanations about how articles function and about how native and other speakers learn the rules about article choice, none of the theories seems completely to account for the complex ways that articles are learned and used.

This paper will summarize the main theories about the meaning of English articles, particularly the definite article, and will chart some of the limitations of each argument. After establishing the historical and current state of the field, I'll examine the most prominent scholarly work on the difficulties faced by both humans and computers as they move from article-less languages to English.

Important to this overall discussion are the scholarly assessments of the complexities related to article usage. Many of these assessments attempt to answer some of the limitations of the most prominent general theories of usage, and in my review of the literature, I will concentrate on studies that examine several particular areas of complex, non-idiomatic uses of *the*: noun phrases including genitive phrases, noun phrases where both definite and null articles are grammatical, and noun phrases where the use of the definite article is connected to the meaning of the noun.

Yet even these studies do not fully explain many situations of article use in English and the ways in which native and non-native speakers interpret and select articles. To explore some of the remaining questions, I have designed a study that compares article judgments of learners and native speakers. The final section of the paper analyzes the results of the survey and offers some conclusions about the complexities of articles in English.

2. Theories on the definite article

In “On Denoting,” Bertrand Russell claims that *the* indicates that the noun refers to a unique entity. Christopherson (1939) takes a similar view. More recently, the notion of uniqueness has been expanded to include familiarity, which in a sense is pragmatic or contextual uniqueness. (Birner and Ward, 2012). Ionin (2006) describes *the* as showing both a specific referent and (+SR) and hearer knowledge of that referent [+HK] and thus comes up with [+definite]. Although learners of English from article-less L1s don’t quickly attain this understanding, it is probable that if the rules were this simple, they would do so much faster. A significant number of noun phrases use articles in a way that is complex, and native speakers are not always unanimous about article choices. Thus, the input is problematic.

We can, of course, see some of these uses as idioms, though these idioms may not be completely inexplicable—for example *on the other hand* and *by hand*. But there remain a number of situations in which article use is not fully explained by ideas of uniqueness (or familiarity.) Russell himself mentions a similar situation that does not fit his principle of uniqueness. "Now *the*, when it is strictly used, involves uniqueness; we do, it is true, speak of 'the son, of So-and-so' even when So-and-so has several sons, but it would be more correct to say 'a son of So-and-so'."(481) And Christopherson calls certain such uses “illogical.”(140) But it is not possible, from the standpoint of someone studying or aiding English language acquisition, to attempt to legislate or dismiss how native speakers use the definite article.

Attempts have been made to explain definite article use from a transformational/generative perspective, for example by claiming that *the* in a determiner phrase is indicative of a *that* clause at the deep structure level or that is anaphoric, referring back to a previous mention of the noun. Grannis (1972) finds these explanations inadequate to explain many uses of *the*.

According to Birner and Ward (2012), neither unique identifiability nor familiarity are the only conditions for the use of the definite article. They use the term “lack of relevant differentiability” to explain usages such as *open the window*, *take the elevator* and *go to the bank* (when there are multiple windows, elevators and banks).

Hawkins (1991) discusses pragmatics and Grice’s notions of conversational implicature in relation to the use of the indefinite rather than the definite article in sentences that introduce the existence of the noun. For example: *England has a prime minister* or *There is a first time for everything*. Even though the meaning of the definite article being used to show uniqueness would seem to apply here, the definite article contains the implication of existence, an implication that would make its use in a sentence about existence infelicitous.

In *Meaning and Grammar*, Chierchia (2000) writes of the idea that the definite article, *the*, is often considered to “presuppose that there is exactly one individual entity satisfying the nominal expression to which it is attached.”(144) This follows the ideas set out by Russell. Chierchia then goes on to mention usages that do not fit this definition: *the* used with mass nouns and plural nouns and a third situation which he explains with an example: “The cat hissed at the other cats she saw at the vet’s.”(144)

This example is fascinating. If we are allowed to imagine an omitted context, this sentence could follow a previous mention. (For example: “I owned a cat and a dog, both of them suspicious of their own kind. The dog ran away from other dogs.”) This contextualizing results in

the kind of the usage which appears to be the easiest for English learners from article-less L1s: the use of *the* with a noun phrase where the noun has been previously mentioned. (Sheen, 2007) If, however, the hissing cat sentence is at the beginning of a description or stands by itself, which is what I think Chierchia intends, the use of the definite article is harder to explain. It seems to indicate topicality.

It is this use of *the* as a topic marker that particularly concerns Epstein in "The definite article, accessibility, and the construction of discourse referents." (2002) Epstein argues that *the* can be used with singular nouns to mark topics. His examples are primarily literary, however, and could be seen as intended to produce a somewhat puzzled response rather than being neutrally communicative. "The stranger came in early February," begins H.G. Well's *Invisible Man*. "Did you hear about the fight?" asked his mother. If *the* conveys definiteness, definiteness being defined as a sign of hearer and speaker knowledge, these sentences will most likely be greeted by questions from the reader/hearer: "What stranger?" "What fight?" Presumably, this is the response desired by the writer/speaker.¹

Epstein's case becomes stronger with plural noun phrases. He reports a conversation in which the speaker explains that the problem with a particular location was the mosquitoes: "And we found a place that we thought, gee, this is such a nice campsite, and we couldn't figure out why nobody had gone down to it, until, about an hour later, when the mosquitoes hit." (350) This sentence brings up the issue noun phrases in which the inclusion or omission of an article is a matter of emphasis rather than grammaticality. *The pollution/Pollution in the city is decreasing.*

¹ *A stranger came in early February* is an acceptable sentence, certainly in writing, though inverting the order to delay the introduction of "a stranger" would be more natural in speech: *There was a stranger who came. . .* Or, using the colloquial [+specific][-definite] *this* (discussed by Ionin (2006)): *In early February, this stranger showed up.* In "Did you hear about a fight?" the indefinite article oddly seems to have only a reading of [-specific][- definite].

The problem with the campsite was mosquitoes/the mosquitoes. (It is possible that *the* means you will soon know what I mean—the theory of accessibility. This is perhaps sometimes tied to whether the noun is being seen by the speaker as generic. Perhaps the idea of the noun phrase as indicating a subset of all such things goes along with the idea that it is going to be a topic of further discussion.)

3. Theories on article learning

3.1 Studies on computer and articles

Native-like article use is also a challenge for computers and their programmers, particularly when dealing with translations into English from languages without articles, or the writing of learners with such languages as their L1s.

"Memory-based Learning for Article Generation" (Minnen et al., 2000) gives an indication of where the field stood in 2000. The authors mention previous work in the 1990s in which rule-based systems generated articles. Of these, they say, the highest accuracy, 88%, was achieved by Heine in 1998 and used a very small database of sentences.

In contrast to Heine, Minnen et al. approached machine learning based using the Penn Treebank. In addition to looking at the head noun of the NP, they looked at various features of nouns: their phrase's functional tag, the noun's semantic category, its countability, the presence of another determiner in the NP. When tested individually, most of these features, other than the head noun itself, did little or nothing to help with article prediction, but in combination they improved the performance slightly. Finally, with a combination of features, the authors achieved an accuracy of 83%, compared to a baseline of 70%, a figure that would have been achieved by guessing "no article" consistently.

They don't mention how they evaluate accuracy, but I assume they see how often their approach chooses the article seen in the original corpus. This approach seems to not explicitly

address the fact that articles, often, carry meaning. They do, however, point out the homogeneity of the corpus.

Their accuracy figures correlate with how much of the corpus they used in training; accuracy went up with quantity, suggesting that using a larger corpus would increase accuracy. Since their approach is based on parsing, they point out they would need to include a parser. This research seems to have been important in terms of the introduction of a corpus and ML into article choice, but have limited applicability to ESL article correction. The authors themselves describe their goals for automatic article generation as translation from languages without articles, correct article use in automatic summary generation, and the generation of articles in the synthesized speech of people requiring assistive technology in order to communicate. There is little chance for human input in these uses, particularly the first two. (I also find myself wondering if the *Wall Street Journal* has a different distribution of nouns than most ESL writing, perhaps having more non-count and proper names.)

Minnen's system was still apparently a leader in the field in 2007 when Turner and Charniak wrote "Language Modeling for Determiner Selection." In this article, the authors describe their system of article selection which, they say, exceeds the performance of "the previous best (Minnen et al., 2000)." They claim an accuracy rate of 86.63%. They also trained on the Penn Treebank but added data from an additional North American News Text Corpus on which they use an automatic parser. They look at the probability of *a/an*, *the* or no article occurring given the particular noun, its position in the noun phrase and its part of speech.

They point out that, while some of their errors truly are errors, which presumably could be easily corrected by a native speaking human, others are ambiguous, and the determiner cannot be determined even by the extended context. One example is particularly clear: "IBM, _____ world leader in computers, didn't offer its first PC until August 1981. . . ." Their program chose *a*

while the original was *the*, but we cannot know which is correct unless we know if the author thought IBM was #1 or in the top few. Here the article carries meaning that is not discernible even from the context, even by a human. Although I believe the authors give these examples to show that their program's errors weren't necessarily what we would call errors, such an example does raise the issue of to what extent correct article determination is possible by machine alone.

Although Turner and Charniak mention that a determiner selection program could "correct English text written by non-native speakers," their interest seems to lie primarily in machine translation. It is Han *et al.* who address ESL writing in particular by training their program not on correctly formed English text by native speakers, but on the corrected texts of ESL learners. They claim a significantly better performance, compared to an article determiner trained on native text, even one that is trained on much larger quantities of native text. While they limited themselves to preposition errors, it seems that this approach would be even more valuable in article selection. Certain article errors are far more common than others, and it seems as if training on such a corpus would be able to reflect that fact. For example, if article deletion errors are far more common than insertion errors (and in my experience with learners from article-less L1s, they are) the program could be trained to be more suspicious of a noun without an article than one with one. (Of course, this problem of omission varies a lot by writing level.)

Challenging the previous studies in which articles were removed from text and the computer attempted to replace them, in "Human Evaluation of Article and Noun Number Usage: Influences of Context and Construction Variability," Lee, Tetreault and Chodorow (2009) discuss the problems with evaluation of determiner selection programs: the fact that there is, not infrequently, more than one possible article. They address what I saw as a problem in the evaluation of both Minnen's and Turner and Charniak's programs: "using what the author wrote as the gold standard can underestimate the system's performance."

They classed nouns into different groups, depending on how variable the nouns were in terms of what article would be likely to accompany them. They removed nouns that could be determined with a very high degree of accuracy--pronouns and nouns that already have determiners such as "this," for example. (Interestingly, Minnen mentions pronouns being included, which does seem to give an oddly inflated accuracy rating, although of course they take this into account by putting their accuracy in context of a baseline that could be achieved randomly) Nouns were classed by entropy level, high entropy nouns being those that appear frequently with a variety of determiners and low entropy those that are far more predictable. They suggest that these classifications could inform a program's decisions; that it could approach suspected errors with low entropy nouns with more confidence than those with high entropy nouns.

They then gave two human raters sentences using the highest and lowest entropy nouns, asking them to select possible and impossible articles. The same raters saw the text as individual sentences and later within a larger context, and their performance was compared to each other and in light of the presence of context. There was actually a fairly high degree of disagreement between the raters, particularly without context, where one rater seemed to be more, in the authors' words, "imaginative": able to think of possible situations for the determiner noun combinations.

Context, Lee *et al.* determined, is helpful for humans in guessing what was the original, "correct" article, and could potentially be so for computers. Their main point, though, is that the evaluation of automatic article determination is underestimated when you are looking for a single correct answer. Unquestionably, if one is just looking for well-formed text, this is true. And clearly there are cases where two article choices are equally good and mean something very similar: for example, *The pollution in the city is getting worse* and *Pollution in the city is getting*

worse. These seem to me to have essentially the same meaning, whereas *She had experience with the computer*, *They presented us with chicken and lamb*, and *He has few friends* become quite different with indefinite articles added. So while it seems quite fair to say that many situations allow multiple correct articles, we also need to draw a distinction between choices that significantly change meaning and those that don't.

3.2 Studies on learner difficulties

Ekiert (2007) mentions a change in the focus of research on SLA of articles in the 1980s: "an interest has arisen in the semantics of interlanguage that motivated the general shift in focus in article research from the acquisition of morphology as form to a focus on morphology as the surface realization of an underlying semantic and conceptual system." Ekiert's study is of the acquisition of the notion of indefiniteness (and thus, the use of the indefinite *a*) by L1 Polish students of English. Her focus is the relationship of thinking and speaking. She clearly summarizes the categories of article use, citing Butler (2002), Huebner (1983), and Thomas (1989) as sources for the five -type system: generics [-SR, +HK], definites [+SR, +HK], referential indefinites [+SR, -HK], non-referential indefinites [-SR, -HK] and idioms.

Particularly interestingly, she writes in this piece that other languages, in this case Polish, have other non-article ways of encoding indefiniteness and so the concept is familiar. In her later piece (2010), she reaches a slightly different conclusion: that Polish shifts the burden of determining definiteness to the listener.

Studies on the acquisition of articles by learners with article-less L1s have focused, understandably, on article use that is relatively easy to explain. Ionin says that articles carry either definiteness or specificity and that in English, it is only definiteness that is expressed by the article choice. She looks at how English language learners from article-less L1s use articles and argues that they fluctuate between an understanding of articles as encoding definiteness and

encoding specificity. As their exposure to English increases, they gradually seem to realize that it is definiteness and not specificity that triggers the use of *the*.

Many researchers have looked at the problems English language learners have mastering native-like article use. Butler (2002) looked at article usage in Japanese EFL students. Her participants were asked to explain their choices, which seemed to show a significant number of misunderstood ideas affecting article use. Because they were interviewed in their L1, the participants gave, in occasion, interestingly thought out justifications for (un native-like) article choices, seeming to misinterpret the idea of hearer and speaker knowledge. (In describing why she chose *the* before *Japanese noblewoman* in *The tale of Genji was written by the Japanese noblewoman*, the participant explained that both speaker and hearer knew that the book was written by Murasaki).

Trenkic (2008) uses one of Butler's examples to argue against Ionin's specificity/definiteness setting. She says that Ionin's study is flawed because her examples of [+specific] [-definite] could also be interpreted as identifiable. Her claim is that learners with article-less L1s are not using a specific setting, but rather responding to explicitly stated knowledge, and using the definite article in sentences that express such knowledge. She also argues that those from article-less languages view them not as determiners, for which they have no category, but as adjectives, and that this view accounts for article omission--if an adjective is redundant in the discourse, it will generally be omitted, whereas an article does not carry semantic information in the same way--it carries information related to the discourse, and it is a noun phrase marker.

Ionin and her co-authors and Trenkic look, in L2 English users, at the relationship between article learning and what articles are actually used for--what information they carry.(Ionin, Baek, Kim, Ko & Wexler, 2011); Trenkic, 2008) Trenkic particularly mentions the

use of the term "definite" and "indefinite" and points out that almost all English learners have encountered at least these words if not explicit instruction involving them; she believes that many article errors stem from learners who misinterpret these two words.

Some researchers have been less concerned with what exactly needs to be learned and more concerned with the learning process. Sheen looked at the effect of error correction on the article usage of ESL students. This was a study, not specifically of article usage, but of the role of correction and particularly of metalinguistic correction (telling the student the rule along with supplying the correct answer). There was a correlation between the group with the explanations and long term improvement. The rule, interestingly, was the one most frequently taught with articles--that previous mention requires *the*.

Other researchers, while concerned with article learning, have focused on the source of problems outside of the understanding of the article. (White, 2009; Ekiert, 2007) White looks at article usage in students with L1s without articles (Chinese, Turkish, Korean, Japanese, and Thai) He asked students to complete an article elicitation task and to rate their certainty about their article choice. His results showed that students generally understood that *the* was [+definite] and didn't apply it based on countability or abstractness. However, uncountable nouns did incorrectly receive null articles even when the context required *the*. Countability, of course, did turn out to be a big issue with the choice between the null article and *a*. Perhaps most significant was the tendency to view abstract countable nouns as uncountable.

Explicit learning is not Ekiert's focus, but she does appear to possibly underestimate its possible influence. Although she states, "Indeed, in English, the nonreferential *there* indicates that the noun following the verb is nonspecific. According to Celce-Murcia and Larsen-Freeman (1999), 'some reference grammars state that only logical subjects with indefinite determiners occur with nonreferential *there*' (449)," she then goes on to conclude that her participant

correctly treated nouns after *there* as indefinite because he had the concept of indefiniteness, rather than having had a specific rule in mind. (Of course, this mastery may have been true for this participant.) In some cases, however, *there* doesn't introduce existence per se, but has other functions such as suggestions (*Who could we ask to take care of our rats? Well, there's the neighbor*) or announcements (*Look, There's the countess!*) (R. Fiengo, Linguistics 76500 Pragmatics and Discourse Analysis, April 2010) Other definite articles after *there* may indicate that the phrase introduces something other than the existence of the referent of the noun phrase but rather introduces a proposition (*There are the children to consider*) or may indicate what Epstein calls topicality ("There were the Useful Presents: engulfing mufflers of the old coach days, and mittens made for giant sloths. . ." from Dylan Thomas's *A Child's Christmas in Wales*.) It would be interesting to know whether learners placed the indefinite article after these non-existential *there* constructions.

3.3 Article uses that are neither idiomatic nor easy to explain

In "L2 English Articles and the Computation of Uniqueness," Yang and Ionin interestingly omitted one of their original items in their study when they found that native speakers also favored the response they had not anticipated (*I broke the leg of the chair*). This category is called "possessive weak definites" by Barker. Weak definites occur in noun phrases with a possessive prepositional phrase. According to Barker, the uniqueness requirement for the definite article is somehow not in effect when the head noun is semantically related to the genitive phrase that follows. Klein et al. extend the use of the term "weak definite" to cover quite different uses of definite article--with places of business (*the hospital, the bank*) and with media (reading *the newspaper*, heard it on *the radio*). While they see this as some sort of incorporation of activity and noun so that the noun no longer has reference (similar to the way the baby in

baby-sitting has no referent), I think it is also possible to view these uses as a kind of generic, where one item represents all the items in the class.

Raymond Murphy, author of a popular series of ESL textbooks, *Grammar in Use*, puts *the radio* into his generic category. (He explains that with animals, inventions and currencies, *the* means "a type of." However, he explains *the bank* with "The speaker is usually thinking of a specific bank or post office."(144)) This explanation, along with "we use *the* when we mean specific things or people" probably had no direct effect on any of the participants in Ionin's study, who often interpreting *the* as [+specific], but it is certainly possible that explicit teaching has some role in creating confusion in learners.

3.4 Less expected readings and native speaker differences

Just as article generation by computers ignores that in some situations there are multiple grammatical article choices and that in some situations that choice will carry meaning, so do some of the studies of acquisition. In Ogawa's study of English article acquisition by Japanese speakers, she includes a cloze, the target answer in parenthesis:

Son: Hey, Dad, can I have (the) car Friday night? I want to take Sally to a school dance.

Dad: Well, that depends. Don't you have (a) paper to write?

The target answer is certainly the most likely answer, if we had a corpus of parent-child communications of this type; however, if this family had multiple cars and had been discussing the paper recently, reversing the article choices would be fine.

Native speakers are not always in agreement about articles. A study using two human raters to judge noun phrases with low and high entropy nouns found that one rater was far more accepting, being presumably more imaginative, more willing to imagine contexts in which an article choice would be grammatical. (Lee, Tetreault & Chodorow, 2009) For instance, in her study of generics, Ionin (2011) designates *The dodo birds are extinct* as ungrammatical, yet the

sentence seems grammatically acceptable to me. And in her study with Kang, target native usage was based on the judgment of the preponderance of the native respondents, who were not entirely in agreement.

3.5 The variability of native speaker article judgments matters

The fact that native speakers are not in substantial agreement in interpreting articles is significant, for it shows that the acquisition of articles involves more complexity than acquiring a Universal Grammar setting. It also has a profound impact on the quality of the input that learners receive; not only do they encounter inconsistencies in exposure to native speech and writing, but explicit teaching will sometimes be contradictory.

4. The study

4.1 The questionnaire

An online questionnaire was created with two parts. (See Appendix A)

The first part had 23 sentences, ten pairs and one trio. Each pair or trio was essentially identical except that the article in one determiner phrase was different. The relevant determiner phrase was in bold. The order of the sentences was randomized. For each sentence, the participants were asked to choose whether it seemed good to them, bad, or somewhere in between. There were five possible degrees of acceptability.

The second part had five sentences with a blank in one determiner position. Participants were asked to “choose the article you think is best.” The options were *a*, *the* or *no article*.

The questionnaire asked the respondents to give their first language, and, if they were not native speakers, to say how long they had lived in the U.S. or another English-speaking country, and to rate themselves on their level of spoken English.

4.2 The participants

The questionnaire was given to two sets of students.

The L2 English students (NNSs) were students at the New School. Thirteen were in a high intermediate undergraduate ESL class at Parsons. Seven were taking an intermediate class to prepare them for matriculation. The high intermediate students self-rated their level of spoken English as ranging from low intermediate to advanced; ten considered themselves intermediate or high intermediate. Their first language was either Korean (9) or Chinese (4). They had lived in an English speaking country for between six months and ten years. The intermediate class consisted of speakers of Korean (4), Chinese (2) and Japanese (1). They had been in an English speaking country for between eight months and two years. (See appendix B.)

The native English speakers were 29 students taking freshman English classes at Emory & Henry College in southwestern Virginia. They were between 17 and 19 years old.

A second group of native English speakers were eight adults who taught at the same school.

4.3 The method

The participants asked to fill out an anonymous online survey. The survey had three parts. In the first part, the participants were shown one or two sentences and asked whether the phrases in bold sounded good to them. They were told that if the phrase sounded fine, they should check "Good." (0) If it sounded completely wrong, they should check "Bad." (4) If it was somewhere in between, they could check the intermediate possibilities, 1, 2, or 3.

The next page contained fill in questions, and the final page had demographic questions.

The task was not timed.

4.4 Categories of sentences

The first category looked at the influence of a genitive phrase on article choice. The first subtype were what Barker called “possessive weak definites.” The head noun is a relational noun and it is followed by a genitive phrase. The noun phrases were *side of a car* and *corner of a busy intersection*.

The second subtype were sentences in which the article could carry a [+ definite] meaning. The noun phrase was either *U.S. president* or *president of the U.S.*

The second category was looked at the role of semantics (and of collocation) in article choice. The first subtype were sentences in which the noun phrase contained the adjective *wrong*, the noun phrases being *wrong key* and *wrong turn*. The second subtype used the noun *people*, modifying it with either the adjective *English* or *American*.

The third category looked at two other atypical but fairly common types of article choice, examining access to rules in making acceptability judgments. Both of these usages are taught. The first is the use of articles with commercial entities and the second is with generally uncountable nouns where the indefinite article can indicate “a type of.” The noun phrases for the first type were *bank* and *deli*. And for the second subtype, the NP was *champagne*.

For each of the subtypes, an attempt was made to see if context could guide the participants to a particular, perhaps less popular article choice. In the first four and final subtype, this sentence was a fill in. In the fifth category with the commercial institutions, the sentences with *deli* had context that suggested a possible countable rather than institutional meaning.

4.5 Discussion of the categories

4.5.1 Nouns followed by genitive phrases

4.5.1.1 Phrases that Barker calls “weak possessive definites” in which the definite article does not express +definite.

Noun phrases: *side of a car*, *corner of a busy intersection*

Sample sentences:

Acceptability rating sentences:

Her bicycle hit **a side** of a car.

Her bicycle hit **the side** of a car.

Ms. Chou’s ice cream shop was located at **the corner** of a busy intersection.

Ms. Chou’s ice cream shop was located at **a corner of** a busy intersection.

Fill in sentence:

_____ corner of a busy intersection was chosen as the location for the yogurt store.

This construction came to my attention in the work of Yang and Ionin, who found themselves surprised that the native speakers did not follow the traditional [+definite] use of *the* in these constructions. Barker refers to these as “possessive weak definites” and argues that when the head noun is relational, it is possible for the article to combine with this noun rather than with the entire noun phrase. So, while an intersection has four corners, usually, and the phrase *corner of a busy intersection* should, logically, take *a* if the hearer doesn’t know which corner, Barker argues that *corner* is, in a sense, unique in that it is a unique relationship and that the genitive phrase is in fact an argument of the head noun.

Barker himself says that “What the speaker has in mind is a unique, specific relation, and that specificity is what the definite determiner is marking” and “the uniqueness requirement simply can’t apply to relations.” He also says “Unfortunately, I am not aware of any technically satisfying way of expressing this intuition formally. . .”) (110)

So perhaps we are looking at nouns that are by their nature a part of something else. Legs of tables and chairs seem to work this way. We seem to be reluctant to use an indefinite article, at least in some of these situations, but, when we want to call attention to a single indefinite entity of this type, we seem to prefer *one*. “A leg of the table is shorter” returns one Google search result while “one leg of the table is shorter” returns many.

Interestingly, *a* sounds wrong in these phrases, though we sometimes want to speak of the thing or person as one unspecified member of a set with multiple members.

Suppose a situation in which speaker and hearer know that Mara comes from a large family. If the speaker is inviting Mara and an undermined brother to an event:

? I’m going to invite a brother of Mara

** I’m going to invite Mara’s brother (* when any of the set of multiple brothers is being possible)*

I’m going to invite one of Mara’s brothers.

Perhaps one way to look at it would be to say that these words are not, in these expressions, being thought of in countable senses. The use of the definite article with a family relationship has continued, despite Russell’s objections: “Ms. Richardson was the daughter of Vanessa Redgrave and the film director Tony Richardson (Bruce Weber, *New York Times*, March 18, 2009). When Weber describes Richardson as the daughter of Vanessa Redgrave, he is not only not thinking of the set of Redgrave’s children, but perhaps not thinking of daughter in a countable sense. However, the article cannot be left out with the non-role relational nouns except when the noun is in an appositive phrase.

I have not seen any mention of articles before relational nouns in any grammar textbook.

4.5.1.2 Genitive phrases in which the article may carry +-definite meaning

Noun phrases: *president of the U.S.*, *U.S. president*

Sample sentences:

Acceptability rating sentences:

Q. Who was James Buchanan? A. He was **a president** of the U.S.

Q. Who was James. K. Polk? A. He was **the president** of the U.S.

Q. Who was Millard Fillmore? A. He was **a U.S. president**.

Q. Who was Warren Harding? A. He was **the U.S. president**.

Fill in sentence:

I know that Alexander Hamilton is on the \$10 bill, but I don't remember who he was or why he is on money. Was he _____ president of the U.S.?

We could certainly argue that when *president of the United States* does take an article, it is functioning in a way consistent with the traditional logic of definites/indefinites. If I am speaking of James Buchanan and thinking of the set of U.S. presidents in 1860, this is a set with a single member and thus “James Buchanan was the president of the United States” is both logical and grammatical. “James Buchanan was a president of the United States” suggests that I am thinking of the set of all presidents of the United States, a set with 43 members, of which Buchanan was just one. So here we see two acceptable choices, depending not on what the hearer knows, but on how the speaker is imagining the set from which the noun phrase comes.

However, by this logic, the minimal context given in the sample sentences does not suggest any particular reason for the speaker to assume a set of presidents at a particular time, since the interlocutor clearly doesn't even know who Buchanan was. Thus Russell and other

proponents of uniqueness would certainly prefer the indefinite article. Also if the article chosen were simply the result of how the speaker was imagining the set from which the president was chosen, it should not matter whether United States appears before the noun or in the prepositional phrase after it.

It does seem possible that, though there doesn't appear to be any semantic difference between *U.S. president* and *president of the U.S.*, the noun *president* is not exactly the same in the two cases. Perhaps, with the genitive phrase, it is a relational word, in Barker's terms. Certainly it refers to a role. Even without the genitive phrase, *president* seems to refer to a role, because it can be used without an article. With U.S. before the noun, however, we may be seeing the inability of the article to combine immediately with the noun.

For example:

He was president.

?*He was U.S. president.*

We also see a null article relational noun in "Verghese Kurien, *father* of the 'White Revolution' and *founder* of the cooperative dairy movement in the country, died in a hospital at Nadiad early on Sunday, aged 90." (*The Hindu*, September 9, 2012) (italics mine.)

None of the grammar textbooks I have seen mentioned that nouns used as roles appear without articles. "He was doctor to" and perhaps "father of the Revolution" are not uncommon, but are not discussed, even by books that go into a great deal of detail about geographical names.

Grammar and Beyond does specifically use the example *president* in its explanation of the definite article "Use *the* for: when there is only one: **the** president, **the** queen, **the** United States, **the** Alps"(104) (Their first precept, "Use *the* when both the write and the reader share common knowledge or information about the noun" would seem better explanation of these noun phrases, at least for the first two.)

In the CoCA Corpus, “The president of the United States” (or US) occurs has 3090 times, making it 19 times as common as the same phrases with an indefinite article (163 mentions).

When the country moves before the noun, definite articles are 5.7 times as common (86 to 15)

4.5.2 The influence of semantics and collocation on article choice

4.5.2.1 Noun phrases with the adjective *wrong*

Noun phrases: *wrong key*, *wrong turn*

Sample sentences:

Acceptability rating sentences:

We took **a wrong turn**.

We took **the wrong turn**.

I brought **a wrong key**.

I brought **the wrong key**.

Fill in sentence:

There are five keys here. Only one will open the car. Don't choose _____ wrong one.

Another exception to the traditional +definite uniqueness rule appears to be a collocation but, I would argue, is not a frozen idiom but also relates to the semantics of the head noun. These are noun phrases that contain the adjective *wrong*.

Wrong initially appeared as if it might work like *same*, which always takes *the* whether it is a noun or the adjective before a noun. For certainly we tend to say *the wrong answer* and *the wrong key* even when there are many wrong answers or wrong keys. However, *a wrong key* appears to be completely acceptable when the key is a musical or computer key.

So then the preference for the before wrong key is related to the meaning of key it occurs only when key is used in the sense of something that opens (or starts) something, either physically or metaphorically, rather than in any other sense. Key, when used in this sense, carries

the implication that it is right. A key to happiness provides happiness just as a key for the cellar door opens a lock, just as answers, by implication, are right answers, although this implication is not quite as strong. An answer to a question is implied to be right unless that implication is cancelled. We do move a bit further from the implication in sentences like. There are five answers to choose from.²

So in phrases where *wrong* cancels the implication of rightness that the head noun carries, a definite article seems to be the natural choice of native speakers.

Perhaps has something to do with countability. Perhaps wrong keys are uncountable? (Wrong answers are very countable, but perhaps not in every utterance.) Turns, of course, are not in themselves right, although if the destination is specified, they are right. The turn to the lake is the right turn to take to go to the lake.

4.5.2.2 noun phrases with a noun that has slightly different meanings when countable and uncountable/generic meaning

Noun phrases: *American people, English people*

Sample sentences:

Acceptability rating sentences:

American people are suspicious of the word “socialism.”

The American people are suspicious of the word “socialism.”

English people are suspicious of the word “communism.”

The English people are suspicious of the word “communism.”

Fill in sentence:

_____ American people are often unwilling to eat dishes made with ingredients they don't recognize.

² In an earlier survey in which students were asked for acceptability ratings for *a wrong key* and *a wrong answer*, this difference was borne out. The native speakers strong disliked *a wrong key* and only moderately disliked *a wrong answer*. Non-natives liked both equally.

Another case of semantics is with the word *people*. Here, the word has both a countable sense and is also, in effect, the plural of person. Possibly one of the greatest difficulties for learners struggling with English articles is the countability of nouns. Furniture and information are, of course, confusing, because their uncountability is counter-intuitive, but for more advanced learners, the difficulty is more with nouns that can be countable or uncountable. “I was told abstract nouns are uncountable,” one advanced student told me, and while he may not have been told this quite so explicitly, abstract nouns are often given in grammar books as examples of uncountable nouns. For example, *Grammar for Writing 2* gives as one category of uncountable nouns, “Abstract nouns or concepts” and then proceeds with a list that includes many nouns that can be countable: *adolescence, anger, behavior, confidence, courage, education, importance, proof, time* etc. (Cain, 2012)³

ESL grammar books don't seem to take into account the fact that a great many English nouns can be countable or uncountable. (Two relatively simple ones are *work* and *art*. While *art* fits the rule of being countable when the speaker desires to talk about different types, *work* [in the sense of *artwork*] can be countable if the speaker really needs to count instances (*his three major works*. In the sense of *job*, of course, it is never countable.)

People is a concrete noun that has an always-plural and a countable sense. The Longman Online dictionary gives the first meaning as “used as the plural of 'person'” and the third as “the people who belong to a particular country, race, or area.” The example sentence is *He pledged*

³ Some of these such as *adolescence* and *courage* can be used as singular countable nouns but not as plural ones.

that he would never lie to the American people. Although this third sense is countable, Longman says it takes a plural verb.⁴

In this case, however, unlike words like *nature* and *experience*, where the countable and uncountable words are semantically very different, the two meanings of *people* are close enough that in many situations they could be interchangeable. An example of their not being interchangeable is a line from a radio interview with Michelle Bachmann on March 6, 2014: “I think the thing that is getting a little tiresome, the gay community, they have so bullied the American people, and they've so intimidated politicians.”

In the sentence acceptability sentences, *people* modified by the adjectives *American* and *English*, in an attempt to see if frequency of collocation is playing a role in the acceptability. For Americans, certainly, *American* is going to appear far more frequently. Also, the phrase “The American people” is, looking at the CoCA corpus, far more common than *American people*, appearing over 20 times as often. Of course, this corpus contains a lot of sources related to news and politics.

On the other hand, while the exact phrase *American people* may not be extremely common in grammar books, *the American people* would be very rare, in both texts and spoken English that are significant sources on input for most learners who are in the United States. A text from *Grammar and Beyond*, a cloze about articles, in fact, is an example: “There is _____ common tendency of people in _____ United States to say things like ‘Let’s get together sometime. . .’”(105-6) The student is given no possibility of putting an article before *people*. The countable sense of *people* isn’t really available here, though it would be if the preposition were *of*. But a definite article would seem possible more on the grounds of topicality or how the

⁴ Occasionally, when used with an indefinite article, it does take a singular verb. Benjamin Disraeli wrote “The phrase ‘the people’ is sheer nonsense. It is not a political term. It is a phrase of natural history. A people is a species. . .”(The Spirit of Whiggism, 1836)

speaker is conceptualizing the sets from which people are drawn—whether we are simply modifying people or whether we are thus choosing a subset.

Thus, the speaker's choice between using a plural noun as a generic (and then modifying it), or using it to refer to a subset of all possible items in that noun class, is a major source of inconsistent native speaker use of the null article versus *the*. In other words, often both choices are grammatical, but may not sound equally good to an individual speaker. An example from a student paper is *People/the people in Penn Station wait in front of the train board*. If the speaker first conceptualizes the noun phrase *people in Penn Station*, combining noun and prepositional phrase into the noun phrase, the article choice will, logically, be definite. But if the article choice is made before the combination with the prepositional phrase, the sense is of a modified generic, and the null article will be chosen. This ambiguity seems particularly common with the noun *people*, in that both approaches can result in a sentence that denotes exactly the same thing. The two sentences about Penn Station convey exactly the same meaning. Sometimes, again, it seems, particularly with the word *people* that the connotation is changed. My intuition is that using “*the* + modifier + *people*” has a different connotation than “null article + modifier + *people*” in phrases like *the Black people* and *the gay people* (all used generically) because the subset approach suggests strongly that the speaker isn't a member of that subset.

Adding to the complexity is the fact that in many contexts the choice of *the* or null article with a plural noun completely changes the referent. While *Then there were the mosquitos!* and *Then there were mosquitos!* in a narrative about the difficulties at a campsite convey the same actual information (though, indeed, the first may suggest that more is to be said about those mosquitos), *Let me tell you about the mosquitos!* and *Let me tell you about mosquitos!* are not equivalent because the second denotes an unmodified generic; the speaker is going to tell you about all the mosquitos in the world.

Although Ionin *et al.* claims that *the* with a singular noun is a common generic in English (Ionin, Montrul, Kim & Phillipox, 2011), I believe it is relatively rare, at least for learners exposed to native English use. She cites the reading of English dictionaries as a major source of input for the *the* + singular generic, but this is probably not a common source of input for learners living in an English speaking country. By far, the bare plural is the most common generic, but its meaning as generic is context dependent. For example, if a speaker complains that *People were talking on their cell phones and to each other* during a movie, the quantifier *some* is omitted but understood.

Another example from a student paper is *the social media*, a phrase that was corrected by a tutor to *social media* because she viewed it as a generic. Although this may have also been the student's idea, with the *the* the result of first language transference, it is perfectly possible to see social media as a subset of all media and thus accept a definite article.

Perhaps the common use of the null article with plural generics has an influence on its use with singular nouns meant in a generic sense. The use of the null article with *fashion industry* is extremely common in the writing of non-native fashion students at Parsons. The intuition that *fashion industry* is a generic type of phrase and the lack of awareness that *industry* is a countable noun are perhaps inextricable.⁵

4.5.3 The influence of collocation and access to article rules on article choice

4.5.3.1 The use of the definite article with commercial institutions

Noun phrases: *bank, deli*

⁵ It would be interesting to see if articles are omitted with singular nouns that have a kind of generic sense more often than with those that don't.

Sample sentences:

Acceptability rating sentences:

Is there an HSBC near here? I need to go to **a bank**.

Is there an HSBC near here? I need to go to **the bank**.

We need milk. Could you go to **a deli** for me? There are two very close to here.

We need milk. Could you go to **the deli** for me? There are two very close to here.

The use of the definite article with commercial institutions such as *bank* is both common and frequently taught, although the explanations given vary. In *Grammar in Use*, Murphy writes, “In the same way we say (go to) the bank, the post office: I have to go to the bank and then I’m going to the post office. (The speaker is usually thinking of a specific bank or post office)” (144)

This explanation is misleading on two levels. For one, it is a clear example of how explicit teaching contributes to the idea in learners that *the* is [+specific] rather than [+definite}. So while Samoan may, as Ionin points out, have articles that encode specificity, the fact that many of her study participants proceed on this assumption also, I would suggest, relates to the instruction and explanations they receive. Additionally, the speaker may indeed be thinking of a particular bank, but this is irrelevant in their use of the article. Certainly, the temptation of this explanation is obvious in towns that contain only one post office or bank, but in a city like New York, the explanation clearly doesn’t hold. A more convincing explanation refers to places that are connected to activities.

So *the* seems to indicate a commercial relationship, while a lack of article seems to indicate an institutional one. The major exception to this principle is *the beach* for which I don’t normally expect to pay. Several books speak of *going to the park* with no particular park in mind, and while I don’t believe I do this, I am willing to believe that some people do. However,

go to a park is more common than *go to the park* (17m to 1.4m on Google search) whereas *go to the bank* appears far more often than *go to a bank* (164 m to 29 m)

Perhaps one way of looking at this usage is generalization; in a sense, one bank is representing all banks. At any rate, we seem to abandon normal article use when we are referring to an activity more than to a place. With institutions we lose the article altogether; with commercial enterprises and beaches, we use *the*. Perhaps these are, in a sense, uncountable uses, and so resist *a*, which emphasizes countability.

4.5.3.2. The use of the indefinite articles to indicate a type of something that is normally uncountable.

Noun phrase: *champagne*

Sample sentences:

Acceptability rating sentences:

I will bring **champagne** to the party.
I will bring **a champagne** to the party.
I will bring **the champagne** to the party.

Fill in sentence:

I haven't had Perrier Jouet before. Is it _____ good champagne?

Now we get to an area in which countability would seem to be less complex.

Liquids are uncountable except when one is talking about types of a liquid. Another exception is when one is speaking of a portion. As portions become more common, the countable becomes more common. *A coffee* was relatively unknown in the U.S. until Starbucks and other vendors arrived, offering no refills and other less familiar drinks which were immediately counted (latte, for example.)

Here we have three sentences that are all, to me, acceptable. No article seems like a reasonable choice; *the* seems possible if there were either an assumption that this party would have champagne, that all parties have champagne, or if the champagne had already been discussed. In *a champagne*, the indefinite article means a type of champagne and it seems to carry the meaning “one, but you or I or both of us don’t know which one,” but this utterance is certainly less likely.

5. The results

5.1 The influence of genitive phrases on article choice

5.1.1 Weak definites

	Non-native speakers				Native speakers			
	High int.		Low int.		Students		Adults	
Her bicycle hit a side of a car.	2.62		2.14		3.41		3.38	
Her bicycle hit the side of a car.	0.77		0.57		0.00		0.13	
Ms. Chou’s ice cream shop was located at the corner of a busy intersection.	0.38		0.29		0.24		0.57	
Ms. Chou’s ice cream shop was located at a corner of a busy intersection.	2.33		2.86		0.67		2.13	
	a	the	a	the	a	the	a	the
_____ corner of a busy intersection was chosen as the location for the yogurt store.	15%	85%	0%	100%	14%	86%	29%	71%

Table 1. Weak definites

In some areas, the differences between the native speakers and the learners were not very large. Both liked the definite articles in both these phrases, though the native speakers did so more emphatically.

Interestingly, most of the learners completely accepted *the* in these phrases, even though this construction is not frequently taught or explained. Even a textbook devoted entirely to article use never addresses this usage (Claire, 1988).⁶ The lower level learners were, if anything,

⁶ Claire does give her readers precept 7.5 “Use the when the noun ‘belongs’ to a particular person group or place” which might be applicable, except an example is “*Mr. and Mrs. Jones were happy about the marriage of their son.*” So probably she means “use *the* before a genitive phrase,” though that clearly is imperfect advice. (35)

slightly more accepting than the higher level ones, so the difference doesn't seem to be connected to time in an English speaking country.

The one significant difference is that the non natives don't seem to distinguish between *a side of a car* and *a corner of a busy intersection*, judging both very similarly, while the native speakers, who love the definite article in both cases, have different feeling about the indefinite in the two, much preferring it with *corner* than *side*. This suggests a semantic relationship. Perhaps corners are more readily countable than sides of a car.

The native speaking students accepted *a corner*, but rated *a side of a car* essentially unacceptable (3.41) and the learners were less negative about (2.62 and 2.14) The lower the level of the learner, the more they found this article choice acceptable. There was a very high standard deviation (1.53) showing that, rather than individually finding this phrase between good and bad, there was a large spread of responses amongst the learners.

Here was one place where the adult native speakers here were different from the students, giving *a corner* 2.13 compared to the 0.67 of the students. Although the adults were, on average, less accepting than the students, with an average rating of 1.23 as compared to 1.10, this is the only item that they rejected to a much greater degree.

The fill-in sentence showed the strong preference that both NS and non-natives had for *the*. But the fact that two of seven adult respondents (and 15% of the student NSs) chose *a* for the fill in is interesting, given that the indefinite article at the beginning of a sentence is unusual.

5.1.2 Genitive phrases

	NNS			NS		
	High int.	Low int.		Students	Adults	
Q. Who was James Buchanan? A. He was a president of the U.S.	2.04	3.14		1.07	0.88	
Q. Who was James. K. Polk? A. He was the president of the U.S.	1.08	.71		1.38	2.13	
Q. Who was Millard Fillmore? A. He was a U.S. president .	1.92	2.14		0.50	0.63	
Q. Who was Warren Harding? A. He was the U.S. president .	0.85	1.14		2.76	3.25	
	a	the	--	a	the	--
I know that Alexander Hamilton is on the \$10 bill but I don't remember who he was or why he is on money. Was he _____ president of the U.S.?	46%	54%	0%	59%	28%	14%
	29%	71%		75%		25%

Table 2. Genitive phrases

The NNs preferred the definite article, accepting it in both phrases, slightly preferring it with *of the U.S.* but not significantly. This suggests that even if the genitive phrase of the previous section, the possessive weak definites, was eliciting the definite article in the learners, something else must be making it attractive in *He was the U.S. president*. Native speaking students rejected this with a rating of 2.76 and native speaking adults did so to an even greater degree with 3.33. Native speakers were more accepting of *the president of the U.S.* than *the U.S. president*.

The fill-in was designed to have a context that would move participants away from a traditional unique [+definite] definite article. The speaker doesn't know who Hamilton was and so is unlikely to have a particular time in mind; presumably s/he is asking if he belongs to the set of all U.S. presidents. English learners preferred *the* here still, with the preference being stronger for the slightly lower level learners. Native speakers preferred *a*, with the preference being stronger in the adults. No adult chose *the*.

The fact that a significant number of the student native speakers chose the in the fill-in suggests that collocation may be playing a role: *the president* is a much more common phrase than *a president*. For the non-natives, collocation probably also has a role. It is also possible that explicit teaching has an effect; there is, as Bunting *et al.* points out, only one president. (Bunting, Dintz & Reppen)

A few of the native speakers chose no article here, which is interestingly something that is not taught to learners in any of the books that I have seen. When a noun refers to a role, it allows no article, thus, perhaps, in some sense becoming uncountable.

5.2 The influence of semantics and collocation on article choice

5.2.1 Noun phrases with *wrong*

	Non-native speakers				Native speakers			
	High int.		Low int.		students		adults	
We took a wrong turn .	0.18		1.00		0.11		0.00	
We took the wrong turn .	1.08		2.00		1.61		1.75	
I brought a wrong key .	1.46		0.57		3.18		3.50	
I brought the wrong key .	0.61		2.29		0.07		0.00	
	a	the	a	the	a	the	a	the
There are five keys here. Only one will open the car. Don't choose _____ wrong one.	30%	62%	14%	71%	3%	97%	12%	88%
	--		--					
	8%		14%					

Table 3. Noun phrases with *wrong*

No huge differences is seen in the acceptability ratings, except the NSs reject *a wrong key* strongly while many of the NNSs accept it, and their preference for *a wrong turn* and *the wrong*

key are stronger than that of the NNSs. The low intermediate NSS didn't like *the wrong key* much. *We took the wrong turn* interestingly got a wide range of responses from the NSs with a mean of 1.61 and a standard deviation of 1.38.

For the fill-in, a few of the learners chose no article, presumably seeing *one* here as a quantifier rather than a pronoun.

Although the native speakers strongly prefer *the* for the fill in, the context/rule had an effect on a couple, who chose *a*. Although this choice seems less usual, it doesn't seem necessarily ungrammatical. But the phrase, *a wrong one*, seems to be becoming less popular with time.

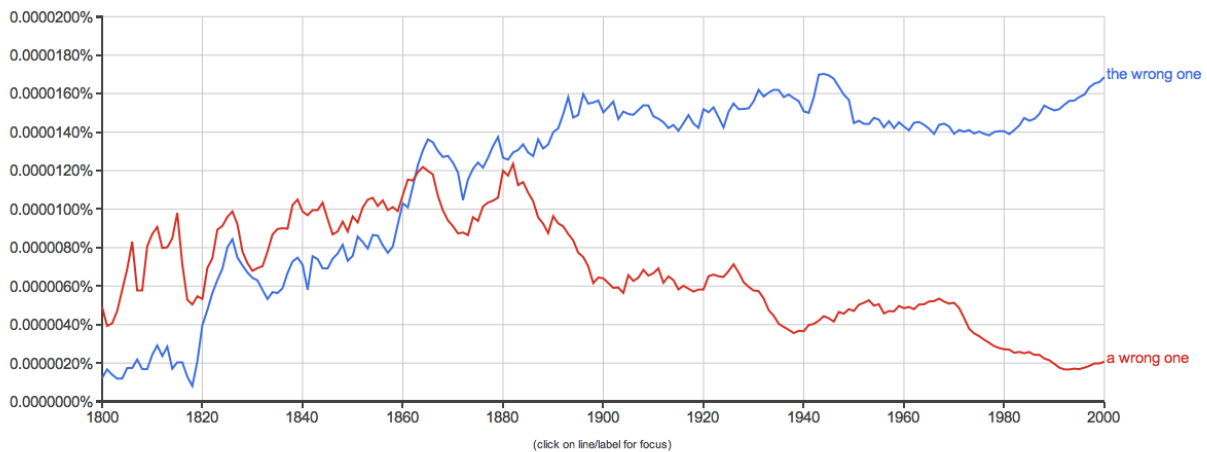


Fig. 1. Google Books n-gram showing the occurrences of *a wrong one* and *the wrong one* over time.

5.2.2 Noun phrases with *people*

	Non-native speakers				Native speakers			
	High int.		Low int.		students		adults	
American people are suspicious of the word “socialism.”	0.85		0.17		0.37		0.75	
The American people are suspicious of the word “socialism.”	2.92		3.0		0.14		0.13	
English people are suspicious of the word “communism.”	1.17		0.17		0.57		0.63	
The English people are suspicious of the word “communism.”	2.54		3.33		0.96		0.13	
	the	--	the	--	the	--	the	--
_____ American people are often unwilling to eat dishes made with ingredients they don’t recognize.	8%	92%	0%	100%	31%	69%	13%	88%

Table 4. Noun phrases with *people*

Native speaker students accept either a definite article or no article with *American people* and *English people*, but they, overall have a slight preference for *the American people* and *English people*. They were least fond of *the English people*. Here I think we are seeing the result of frequency of collocation; Americans hear much more about the American people. The adult native speakers didn’t have the same preference; presumably the phrases are both reasonably common, or the adults see that they logically should be.

The fill-in was designed to increase the attractiveness of no article, since the uncountable meaning of *people* seems less available here where the Americans are exhibiting varying behavior and not acting as a political unit or in a political sphere. A sizeable percentage of the NS students still chose the frequent collocate, *the*. It is very likely that they hear this frequently but don’t fully recognize a semantic difference in the noun that follows.

Although one adult NS chose *the* in this sentence, they seemed to be less affected by collocation without meaning, or had had sufficient exposure to collocations such as *the English people*.

In contrast, the non-natives liked the article-less phrases and disliked the ones with the definite article, and this was more pronounced in the lower level learners. This suggests to me that this collocation is much less familiar to them; it was not a group of students who would have much exposure to news reports and political rhetoric in English.

5.3 The influence of rules on article choice

5.3.1. Commercial institutions

	Non-native speakers		Native speakers	
	High int.	Low int.	Students	Adults
Is there an HSBC near here? I need to go to a bank .	2.61	2.29	1.76	1.50
Is there an HSBC near here? I need to go to the bank .	0.77	0.85	0.07	0.38
We need milk. Could you go to a deli for me? There are two very close to here.	1.92	2.00	1.76	1.63
We need milk. Could you go to the deli for me? There are two very close to here.	1.46	1.29	0.31	0.75

Table 5. Noun phrases with commercial institutions

In both cases, the context makes clear that the speaker is not talking about a bank or a deli known to them (or, of course, to the listener). And in both sentences the noun phrase follows *go to* which suggests the (commercial) activity normally associated with the particular commercial entity. (I believe that using *find* instead would elicit very different responses, a preference for the indefinite article.)

The native speakers liked *the* with both *bank* and *deli*, and had very varying opinions on *a*. The learners dislike *a bank* more than the native speakers did, and liked *the bank*, though not as strongly. Like the native speakers, the learners preferred *the deli* to *a deli*, but not as strongly.

Go to the bank appears frequently in ESL textbooks and in real life, so it isn't surprising that it was accepted by the learners. The learners did see some difference between *the bank* and *the deli* sentences, though whether it was the noun or the structure of the context that affected the article acceptability is unclear.

5.3.2 Normally uncountable noun phrase

	Non-native speakers				Native speakers			
	High int.		Low int.		Students		Adults	
I will bring champagne to the party.	1.83		1.86		0.18		0.00	
I will bring a champagne to the party	1.15		1.57		3.82		3.88	
I will bring the champagne to the party.	1.69		3.14		0.31		0.25	
	a	the	a	the	a	the	a	the
I haven't had Perrier Jouet before. Is it _____ good champagne?	69%	23%	43%	0%	66%	3%	50%	0%
	--		--		--		--	
	8%		53%		31%		63%	

Table 6. Noun phrases with a normally uncountable noun

NS students, and to only a slightly lesser degree, NS adults, reject *a champagne*, not seeing the possibility of “a kind of” as the justification for the indefinite article. Many of the learners accept “a champagne”; whether they are more conscious of this rule or are unaware that

champagne is normally uncountable is unclear, though the fact that the more advanced learners are more accepting suggests an awareness of use of *a* for types.

The lower level learners rejected *the champagne* strongly and the higher level ones did so to a lesser degree. This might be evidence of a belief that there is one right answer; additionally, it might show an inability to understand *the* as possibly indicating associative anaphora, or it might reflect a lack of awareness that speaker and listener might reasonably expect champagne at a party. Of course, the sentence is also possible if the champagne has already been discussed.

6. Conclusion

A great many things contribute to the difficulties that learners of English have mastering native-like article use. Certainly, transference from first languages with no articles leads learners to leave countable nouns with neither any determiner nor the plural -s. Both *a* and *the* are usually unstressed and thus are somewhat difficult to hear. But it is also the case that article use is complicated. And many of these complexities are either ignored by instructional texts or are treated as idioms or individual cases.

Previous mention is the most frequently taught rule for the use of *the*. Yet it probably accounts for a relatively small percentage of definite articles, particularly in more advanced texts. Explanations given by teachers and texts do stress the specific reference [+SR] quality of the definite article and give a little less emphasis to the hearer knowledge [+HK] component.

Certainly it seems as if many learners would like a general definition of *the*, but both uniqueness and familiarity are, as Birner and Ward pointed out, not necessary or sufficient. The closest to a single explanatory theory is Epstein's notion of accessibility, though his idea that *the* indicates low accessibility is perhaps an indirect way of saying that we tend to use pronouns

rather than nouns for highly accessible referents. A more accessible way of wording his idea might be: *The* is telling the listener, “You know what I mean or soon will.”

We then have several categories of problematic definite determiner use:

- (a) the noun is relational and is followed by a genitive phrase explaining that relation
- (b) the noun is associated with an activity rather than a thing and thus comes to be treated as somewhat generic.

In both of these cases, the nouns seem to be classed as on a continuum in that many seem to accept both this partitive or incorporational reading or a more ordinary object/entity referent reading.

- (c) something complex is happening with the semantics of the noun phrase, as in *the wrong key* and *the wrong one*. These expressions appear frozen and idiomatic at first glance, but aren't. *A wrong key* is an acceptable usage to native speakers if the key is musical or on a keyboard. *The wrong one* appears to be moving toward idiomatic status but has not yet achieved it.

- (d) the choice of the null or definite determiner with mass, plural and non-count nouns creates often subtle differences in meaning or emphasis.

One of the reasons articles are complicated for learners is that they are intimately tied with the semantics of the noun phrases they precede. Few nouns are always countable or uncountable, and many have perhaps a quasi-countable sense, the situations where *the* is acceptable to native speakers but *a* is either unacceptable or less acceptable and the null article is not possible. For adults from article-less L1s who are learning academic English, the rapid acquisition of nouns, nouns that are often complex and abstract, native-like article use is particularly challenging.

Article use is neither simple nor insurmountable. Advanced adult learners should not despair of acquiring native-like article use. Learners have the ability to acquire complex and

unexplained article usage with reasonable exposure, but the process can be slow; even young native-speaking adults probably are continuing to acquire semantic knowledge that will change how they use articles. Learners might also benefit from the reassurance that in many cases, more than one article choice is possible.

Appendix A - Questionnaire

People often have different ideas about what sounds right to them. The following questions have no right or wrong answers. I am interested in how good a phrase sounds to you.

2. For each of the following items, say if you think the article sounds good, bad, or somewhere in between.

	Good				Bad
Is there an HSBC near here? I need to go to a bank .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Her bicycle hit a side of a car.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who was James Buchanan? He was a president of the U.S.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We took a wrong turn .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will bring the champagne to the party.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who was Millard Filmore? He was a U.S. president .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I brought the wrong key .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The American people are suspicious of the word "socialism."	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Her bicycle hit the side of a car.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who was James K. Polk? He was the president of the U.S.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We need milk. Could you go to a deli for me? There are two very close to here.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ms. Chou's ice cream shop was located at the corner of a busy intersection.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who was Warren Harding? He was the U.S. president .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I brought a wrong key .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The English people are suspicious of the word "communism."	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We took the wrong turn .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ms. Chou's ice cream shop was located at a corner of a busy intersection.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will bring a champagne to the party.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We need milk. Could you go to the deli for me?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

There are two very close to here.

American people are suspicious of the word socialism.

I will bring **champagne** to the party.

English people are suspicious of the word "communism."

Is there an HSBC near here? I need to go to **the bank**.

Fill-ins

For the following sentences, pick the article that sounds best to you.

3. I know that Alexander Hamilton is on the \$10 bill, but I don't remember who he was or why he is on money. Was he _____ president of the U.S.?

- a
- the
- no article (blank)

4. _____ corner of a busy intersection was chosen as the location for the yogurt store.

- a
- the
- no article (blank)

5. There are five keys here. Only one will open the car. Don't choose _____ wrong one.

- a
- the
- no article (blank)

6. _____ American people are often unwilling to eat dishes made with ingredients they don't recognize.

- a
- the
- no article (blank)

7. I haven't had Perrier Jouet before. Is it _____ good champagne?

- a
- the
- no article (blank)

Tell me a bit about you. . . .

8. What is your first language?

9. If your first language is not English, how long have you lived in the U.S. or another English speaking country?

10. How would you rate your level of English?

- Low intermediate
- Intermediate
- High intermediate
- Advanced
- Near native
- Native speaker

11. If you wish, you may write comments here:

Appendix B – participants

Breakdown of participants by length of time in English speaking countries and self reported level

Pre-matriculated intermediate class (shown in blue in the in-text charts)

Length of time in English speaking country	Self-reported level				
	low intermediate	intermediate	high intermediate	advanced	near native
2 years			2		
≤ 1 year	1	3	1		

Matriculated high intermediate class

Length of time in English speaking country	Self-reported level				
	low intermediate	intermediate	high intermediate	advanced	near native
6-10 years			1	1	1
3-5 years		1	4	1	
≤ 1 year	1	1	2		

Although there is a correlation between time in the U.S. and self assessed language level, the survey answers of the students with the longer time in the U.S. (6-10 years) didn't show any significant statistical difference from those their more newly arrived classmates; this says nothing about the general effect of more time in the U.S. since the entire sample were students who had been placed in the same class based on a placement test that evaluated grammar, reading comprehension, speaking and writing.

References

- Avery, P., & Radišić, M. (2007). Accounting for variability in the acquisition of English articles. *Proceedings of the 2nd conference on generative approaches to language acquisition. North America* (pp. 1-11). Somerville, MA: Cascadilla Proceedings Project.
- Barker, C. (2004). Possessive weak definites. In J. Kim, Y. Lander, & B. Partee (Eds.), *University of Massachusetts occasional papers in linguistics, 29: Possessives and beyond: Semantics and syntax*. Amherst, MA: GLSA Publications. 89-133.
- Birner, B., & Ward, G. (2012). Uniqueness, familiarity, and the definite article in English. *Proceedings of the annual meeting of the Berkeley linguistics society, 20(1)*. Berkeley, CA: University of California.
- Bunting, J., & Diniz, L. (2013). *Grammar and beyond 4*. Cambridge, UK: Cambridge University Press.
- Butler, Y. G. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition, 24*, 451-480.
- Cain, J.S. (2012) *Grammar for writing 2*. New York: Pearson Longman.
- Celce-Murcia, M., Larsen-Freeman, D., & Williams, H. A. (1999). *The grammar book: An ESL/EFL teachers' book* (2nd ed.). Boston, MA: Heinle & Heinle.
- Chen, H.-J. H. (2009). Evaluating two web-based grammar checkers: Microsoft ESL Assistant and NTNU Statistical. *International Journal of Computational Linguistics and Chinese Language Processing, 14(2)*, 43-48. Retrieved from <http://www.aclclp.org.tw/clclp/v14n2/v14n2a2.pdf>.
- Chierchia, G., and McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics* (2nd ed.). Cambridge, MA: MIT Press.
- Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing, 27(3)*, 419-436.
- Christophersen, P. (1939). *The Articles*. London: Munksgaard, Copenhagen and Oxford University Press.
- Claire, E. (1988). *Three little words a an and the: A foreign student's guide to English articles*. McHenry, Illinois: Delta Publishing.
- Daglish, G. M. (1987). Some uses of computers in teaching English as a second language: The issue of control. In J. S. Blanchard & G. E. Mason, *The Computer in Reading and Language Arts* (pp. 81-94). New York, NY: Haworth Press.
- Davies, M. (2008-). The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at <http://corpus.byu.edu/coca/>.
- Ekiert, M. (2007). The acquisition of grammatical marking of indefiniteness with the indefinite article a in L2 English. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 7(1)*. Retrieved from <http://www.tc.columbia.edu/tesolalwebjournal>
- Ekiert, M. (2010). Linguistic effects on thinking for writing: the case of articles in L2 English. In Z. H. Han & T. Cadierno (Eds.), *Linguistic relativity in SLA: thinking for speaking*. Bristol, UK: Multilingual Matters. 125-153.
- Epstein, R. (2002). The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics, 12 (4)*, 333-378.
- García-Mayo, M. & Hawkins, R. (Eds.). (2009). *Second language acquisition of articles: Empirical findings and theoretical implications*. Amsterdam: John Benjamins.

- Grannis, O. C. (1972). The definite article conspiracy in English. *Language Learning* 22, 275-289. Doi: 10.1111/j.1467-1880.1972.tb00088.x.
- Han, N.-R., Tetreault, J., Lee, S.-H., & Ha, J.-Y. (2009). Using an Error-annotated learner corpus to develop an ESL/EFL error. *International Journal of Computational Linguistics and Chinese Language Processing*, 14(2), 161-180.
- Hawkins, J. A. (1991). On in(definite) articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics*, 27(2), 405-442.
- Ionin, T. (2006). This is definitely specific: specificity and definiteness in article systems. *Natural Language Semantics*, 14, 175-234.
- Ionin, T. (2009). Specificity. In L. Cummings (Ed.), *The pragmatics encyclopedia*. New York, NY: Routledge. 449-454.
- Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2011). That's the meaning: interpretation of definite and demonstrative descriptions on L2-English. In M. Pirvulecu (Ed.), *Proceedings of the 4th conference of generative approaches to language acquisition – North America, University of Toronto*. Somerville, MA: Cascadilla Proceedings Project.
- Ionin, T., Montrul, S., Kim, J., & Philippov, V. (2011). Genericity distinctions and the interpretation of determiners in LS acquisition. *Language Acquisition*, 18, 242-280.
- Ionin, T., Zubizarreta, M. L., & Bautista-Maldonado, S. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua*, 118, 554-576.
- Klein, N. M., Gegg-Harrison, W. M., Sussman, R. S., Carlson, G. N., Tanenhaus, M. K. (2009). Weak definite noun phrases: rich, but not strong, special, but not unique. In U. Sauerland & K. Yatsushiro (Eds.) *Semantics and pragmatics: From experiment to theory*. New York, NY: Palgrave Macmillan.
- Knutsson, O., Pargman, T. C., & Eklundh, K. S. (2003). Transforming grammar checking technology into a learning environment for second language writing. *Proceedings of the HLT-NAACL 03 workshop on building educational applications using natural language processing*. Retrieved from <http://acl.ldc.upenn.edu/W/W03/W03-0206.pdf>.
- Ko, H., Ionin, T., & Wexler, K. (2010). The role of presuppositionality in the second language acquisition of English articles. *Linguistic Inquiry*, 41, 213-254.
- Leacock, C., Gamon, M., & Brockett, C. (2009). User input and interactions on microsoft research ESL assistant. *Proceedings of the NAACL HLT Workshop on innovative use of NLP for building educational applications*, Rochester, NY: University of Rochester, 73-81.
- Lee, J., Tetreault, J., & Chodorow, M. (2009). Human evaluation of article and noun number usage: Influences of context and construction variability. *Proceedings of the third linguistic annotation workshop, ACL-IJCNLP* (pp. 60–63). Singapore: Association for Computational Linguistics.
- Liu, D., & Gleason, J. I. (2002). Acquisition of the article *the* by nonnative speakers of English: An analysis of four nongeneric uses. *Studies in Second Language Acquisition*, 24, 1-26.
- Loftus, E. (1996). *Eye witness testimony*. Cambridge, MA: Harvard University Press.
- Ludlow, P. & Neale, S. (2006). In M. Devitt and R. Hanley (Eds.). *The Blackwell guide to the philosophy of language* (pp. 274-287). Oxford, UK: Blackwell.
- Lyons, C. (1999). *Definiteness*. London: Cambridge University Press.
- Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Brockman, W., The Google Books Team . . . and Aiden, E.L. (2011). Quantitative analysis of culture using millions of digitized books. *Science* 331 [Published online ahead of print 12/16/2010].

- Minnen, G., Bond, F., & Copestake, A. (2000). Memory-based learning for article generation. *Proceedings of CoNLL-2000 and LLL-2000*. Retrieved from <http://lingo.stanford.edu/pubs/det-ml.pdf>.
- Murphy, R. (2004). *English grammar in use* (3rd ed.). Cambridge, UK: Cambridge University Press.
- Ogawa, M. (2008). The acquisition of English articles by advanced EFL Japanese learners: Analysis based on noun types. *Journal of Language and Culture. Language and Information*, 3, 133-151.
- Parrott, M. (2000). *Grammar for English language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Russell, B. (1905). On denoting. *Mind*, 14, 479-493.
- Rutherford, W. (1998). *A workbook in the structure of English: Linguistic principles and language acquisition*. Oxford, UK: Blackwell.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255-283.
- Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or Adjectives? *Bilingualism: Language and Cognition*, 11, 1-18.
- Turner, J., & Charniak, E. (2007). Language modeling for determiner selection. In *Human language technologies 2007: The conference of the north american chapter of the association for computational linguistics; companion volume, short papers* (pp. 177-180). Rochester, NY: Association for Computational Linguistics.
- White, B. (2009). Accounting for L2 English learners' article choices. *MSU Working Papers in SLS*, 1(1), 14-37.
- Yang, M., & Ionin, T. (2009). L2 English articles and the computation of uniqueness. In J. Crawford, K. Otaki, & M. Takahashi (Eds.), *Proceedings of the 3rd conference on generative approaches to language acquisition North America* (pp. 325-335). Somerville, MA: Cascadilla Proceedings Project. Retrieved from <http://www.lingref.com/document#2332>.