A SEMANTIC MAP APPROACH TO ENGLISH ARTICLES (*A*, *THE*, AND Ø)

by

BRIAN C. BUTLER

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2012

DISSERTATION APPROVAL PAGE

Student: Brian C. Butler

Title: A Semantic Map Approach to English Articles (*a*, *the*, and Ø)

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Eric Pederson	Chairperson
Doris Payne	Member
Vsevolod Kapatsinski	Member
Lucien Brown	Outside Member

and

Kimberly Andrews Espy	Vice President for Research and Innovation Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2012.

DISSERTATION ABSTRACT

Brian C. Butler

Doctor of Philosophy

Department of Linguistics

December 2012

Title: A Semantic Map Approach to English Articles (*a*, *the*, and Ø)

The three structural possibilities marking a noun with an English article are *a*, *the*, and Ø (the absence of an article). Although these structural possibilities are simple, they encode a multitude of semantic and pragmatic functions, and it is these complex form-function interactions that this study explores and explains using a semantic map model. The semantic map that is proposed contains three dimensions, which I refer to as Grammatical Number, Referentiality, and Discourse Mode. Each of these dimensions contains a number of further semantic values or pragmatic functions – which I will label "attributes" – that are implicated in English article choice. Various semantic map versions are tested and compared with a methodological approach that uses data collected in a controlled protocol from an elicited conversational discourse. The version that performed best is used as a basis for proposing a comprehensive semantic map that includes the following dimensions and dimensional attributes: a Number dimension with 3 attributes (singular, plural, and uncountable); a Referentiality dimension with 11 attributes, including 7 referential attributes that describe kinds of identifiability (proper names, shared lexis, shared speech situation, frame, current discourse, identifiable to speaker only ["new reference"], and identifiable to neither speaker nor listener [non-specific]) as well as 4 non-referential attributes (categorization, general non-referential

expressions, finite verb [verb-object] "noun incorporation", and idioms); and a Discourse Mode dimension with 4 attributes (headline, immediacy, normal, and reintroducing).

This model of English articles contributes to the field of research on articles as well as to the field of English language instruction and learning. In addition, it is suggested that the methodological paradigm used to test the semantic map model may be useful as an experimental paradigm for testing semantic maps of other constructions and languages.

CURRICULUM VITAE

NAME OF AUTHOR: Brian C. Butler

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Emmanuel School of Religion, Johnson City, TN
Pacific Christian College, Fullerton, CA

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2012, University of Oregon
Master of Arts in Religion, 1998, Emmanuel School of Religion
Bachelor of Arts, 1992, Pacific Christian College

PROFESSIONAL EXPERIENCE:

Instructor (Graduate Teaching Fellow), American English Institute, University of Oregon, 2005-2012

Adjunct Faculty, World Languages Department, Brookhaven College, 2004-2005

Associate Faculty, Collin County Community College, 2005

English as a Foreign Language Instructor, English L-School, Ueda, Japan, 1998-2001

PUBLICATIONS:

Butler, B. Acoustic correlates of obstruent voicing distinctions in Yangon Burmese (Yangon Myanmar). *Journal of the International Phonetic Association*. (Provisionally accepted, pending revisions.)

ACKNOWLEDGMENTS

I would like to thank the faculty of the Linguistics Department at the University of Oregon for guiding me through the many years and projects of this doctoral program. I thank them also for providing me financial support in the form of a graduate teaching fellowship each term. That is something I will forever be grateful for. My assignment most terms was to teach at the University of Oregon's American English Institute, and I'd like to thank the AEI's faculty and staff at for their professionalism, good nature, and daily inspiration. Working there has always been fun.

I'd like to thank Doris Payne for her early guidance on the class projects that eventually turned into this dissertation as well as for serving on my dissertation committee; her numerous and extremely insightful comments and suggestions helped me immensely as I tried to improve the dissertation in its final stages. I'd also like to thank my other committee members, Vsevolod Kapatsinski and Lucien Brown, for their careful reading and helpful comments.

And I'd like to thank my advisor and committee chairman, Eric Pederson, for guiding me along the way. Our weekly meetings provided a framework for the fruitful conversations and a structure for the piecemeal progress that eventually produced this final product. I couldn't have hoped for a better, more positive, more perspicacious, or more supportive advisor.

I'd also like to thank the members of the Cognitive Linguistics Workgroup, which met weekly throughout my years in this program. In recent years, they heard many iterations of the various parts of this project and offered me much valuable feedback that shaped the eventual outcome.

All of these people have contributed to and improved the form of the dissertation as it appears here; however, I must take credit for any and all of its weaknesses.

Finally, I'd like to thank my family and friends for supporting me and keeping me sane in myriad ways, large and small. I feel fortunate to have you all in my life.

I dedicate this to the great eastern sun, to being hit with a shoe, to the clinking of a dish.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

The three structural possibilities marking a noun with an English article are *a*, *the*, and Ø (the absence of an article).[1] Although these structural possibilities are simple, they encode a multitude of semantic and pragmatic functions, and it is these complex form-function interactions that this study explores.

To date there have been many theoretical and applied treatments of the English article system. A number of researchers, including Christophersen (1939), Hawkins (1978), Reuland & ter Meulen (1987), Gundel, Hedberg, & Zacharski (1993), Haspelmath (1997), Lyons (1999), Abbott (2001, 2006, 2008), Givón (2001, 2005), Levinson (2005), and Riley (2007), have explored the aspects of definiteness and indefiniteness, contributing to our understanding of the article system, including especially its relationship to issues of referentiality. However, all the theoretical treatments, even those with the most extensive handling of the many details pertaining to article use, such as Christophersen (1939) and Du Bois (1980), have failed to produce a fully comprehensive model explicating the distribution of English articles in all (or even most) contexts. In addition to the theoretical investigations, much attention aimed at practical application has also been devoted to understanding article use, especially by language educators. Popular grammar books designed for English language learners, such as Azar (1999) and Murphy (2007), include detailed sections on article usage.

[1] Since this study focuses on the traditional English articles *a* and *the*, consideration of the distribution of other non-article determiners such as *some* and *this* will be left to a future treatment.

Unfortunately, their guidelines also do not adequately cover the full range of article usage found in English natural language. While certainly useful to a degree, they tend to posit rules and then offer lists of exceptions to the rules, an approach that seems necessary given the lack of an adequate comprehensive model, but one that can be frustrating to language learners and language teachers.

An adequate model of article usage must be cogent and comprehensive, and it must have its foundation in the functional realities of human language. This dissertation, therefore, offers a model explicating the range of functions of the English article system and suggests ways to test, refine, and expand it toward the goal of producing a truly comprehensive treatment. In order to investigate whether any proposed model accurately describes article use in different contexts, it is necessary to determine what testable hypotheses it makes. An approach which extracts testable explanations from the literature on English articles, constructs models from them, and then empirically tests the models, offers the possibility of comparing the predictive power of various claims;[2] this study follows such an approach.

1.1. Key contributions of this study

This work offers a number of new and useful contributions to the research on English articles. First, it offers a coherent model of English article use that integrates the many salient features of articles described in the literature. Many studies have focused on concepts related to articles such as definiteness, identifiability, specificity, referentiality, and cognitive framing. Other studies (particularly Du Bois, 1980) have

[2] Although this paper investigates English articles only, the approach used here could also be employed to map the functions of articles in other languages and compare them typologically.

added the idea of discourse mode to the discussion. Furthermore, all researchers apparently assume that grammatical number is an important (though sometimes, perhaps, obvious) detail. However, to date no one has incorporated all of these considerations by sorting through and interpreting the terminology, determining what is redundant and what is lacking, and integrating the remaining unique and salient features into a coherent, comprehensive model. The purpose of this work is to accomplish these goals. (If the reader wishes to preview the entire model, it can be seen in Figure 8 on pp. 145-146.)

Second, this study uses the idea of semantic maps as a basis for integrating and organizing the functions of articles, something no previous work has done.[3] The approach here has been to incorporate the salient functions determining article usage into a semantic map framework. Parts of the resulting model (such as the identifiability portion) have likely accomplished this goal reasonably successfully and can be usefully employed to understand and compare articles and other constructions cross linguistically. However, other parts (such as the non-referential portion) might not find broad applicability in their current form since they are organized by an admittedly somewhat hodge-podge mixture of construction types and functions. Where this is the case (due to current inability to discover unifying principles to explain them), those parts of the model still need further development in terms of functions that are more typologically salient and consistent. However, the general use of semantic maps as a basis for comprehensive understanding of the function and use of articles is currently unique to this study, as far as I am aware.

[3] Croft (2003) does use a semantic map approach to look at issues of definiteness, and Gundel et al. (1993) offers an implicational hierarchy of givenness that could be fitted to a semantic map, but neither work seeks to address the entire system of English article usage.

Third, the model proposed in this study includes discourse mode as an important feature of the English article system. This idea originated with Du Bois (1980) but since then has received little attention in accounts of article use. However, including discourse modes resolves some issues of article use variability that otherwise might require an unnecessary (and strangely motivated) proliferation of "identifiability types," or situations that determine whether speakers can identify a referent when it is mentioned. For example, the addition of discourse modes to the overall model allows us to handle cases where a speaker clearly knows a referent is identifiable to the listener but acts (based on her use of articles) as though it is not, and, conversely, cases where a speaker clearly knows a referent is *not* identifiable to the listener but acts (again, based on her use of articles) as though it *is*. (In our model, the first case can be handled by the "Reintroducing" discourse mode, and the second case by the "Immediacy" discourse mode, both explained in detail in Chapters III and IV.) The demonstrated usefulness of Discourse Modes in our model suggests that concept of Discourse Modes should be included in the curricula of English language programs to enable learners to accomplish the communicative functions that the various Discourse Modes allow.

Fourth, in likewise manner, the model's explication of a somewhat complex Referentiality/Identifiability system suggests that English language educators should incorporate the elements of this more complex system into their program curricula. Doing so will enable learners to better utilize the English article system to accomplish the full functional range that it richly facilitates.

Fifth, by prominently integrating the grammatical number system as a dimension in the model, the importance of number in the lexical representation of nouns is

highlighted. This is important since it emphasizes the idea that without a rich representation of number (and particularly the *countability* aspect of the English number system) that matches the representation broadly shared across speakers, English articles cannot be consistently employed. While this might seem like an obvious theoretical point, it has not yet found broad application in the field of teaching and learning English as a second language, where rich grammatical countability information is not normally integrated in approaches to vocabulary acquisition of nouns. Second language learners of English should find it much easier to master the English article system if they explicitly acquired rich countability information as they learned new nouns, in the way that adult learners of Romance languages, for example, are explicitly taught grammatical gender information when learning nouns.

Sixth, this study tests a portion of the model it proposes by employing a methodological paradigm that has not been widely applied to the testing of semantic maps but which might prove broadly useful for such testing in the future. In order to evaluate a model that involves referentiality and discourse pragmatics such as the one proposed here, one has to know what the speaker thinks about the referents she mentions and what pragmatic intention she is trying to accomplish. The usefulness of the methodological paradigm employed here resides precisely in a design that seeks to manipulate the speaker's knowledge and pragmatic intent. (This was done in the study reported in Chapter III by doing two things: first, by showing the speaker a video in order to control her knowledge of the identity of referents in different scenes throughout the video, and second, by asking the speaker to respond to specific questions about the video in order to control the pragmatic intent of her responses.) When a speaker's

knowledge of referents and purpose in speaking are manipulated in the course of data elicitation, they can be used as independent variables in testing the model. The predictions of the model become dependent variables, and these can be compared against the actual linguistic structures spoken be the speaker. In this way, the output predictions of various competing models can be compared and evaluated to determine which best fits the actual spoken data. Therefore, although its implementation here involved some compromises in order to test certain aspects of the model at the expense of ignoring others (discussed further in Chapter III), this methodological paradigm offers the possibility of wide application to the testing of models of pragmatically-determined structures such as English articles.

The full model (presented in Chapters III and IV) attempts to integrate all factors that the literature has suggested impinges on article usage, but undoubtedly the extant research has not uncovered the full range of factors that affect their use, and therefore it is likely that some things have been left out. In the future, the model must be tested on a wide variety of genres from a wide variety of speakers and authors to identify its weaknesses, expand its comprehensiveness, and refine its predictive power.

1.2. Structure of the dissertation

The organization of the dissertation proceeds as follows. Chapter II provides a review of the literature that deals with English articles and issues that are relevant to their many functions. Chapter III proposes, tests, and compares four competing semantic-map-based models designed to predict correct article usage based on the referentiality and discourse mode status of noun phrases in a discourse. Since testing constraints prevented

all salient features from being included in the tested models, Chapter IV discusses additional functions that are found to affect English articles and proposes a comprehensive model of English article usage that incorporates all implicated functions. Chapter V concludes the dissertation with a discussion of the benefits of this semantic-map-based modeling approach, consideration of its limitations, and directions for its future application.

CHAPTER II

THEORETICAL BACKGROUND

2.1. Definiteness

2.1.1. Various uses of the term "definiteness"

Definiteness is a quality normally associated in the literature with identifiability and/or specificity in a variety of linguistic forms. In English, these forms include common nouns marked with the definite article (*the car*, *the egg*, *the idea*, etc.), proper nouns (*Steve*, *Frodo*, *Crater Lake*, etc.), personal pronouns (*I*, *you*, etc.), and nouns modified by possessive pronouns or phrases (*my car*, the word *egg* in the phrase *a spectators egg*), as well as deictic terms such as demonstratives (*this*, *that*, *these*, etc.), location terms (*here*, *there*, etc.), and some senses of time terms (*now*, *then*, *long ago*, etc.) (e.g., Chafe & Li, 1976; Lyons, 1977; Hawkins, 1978; Du Bois, 1980; Fox & Thompson, 1990; Berezowski, 2001; Givón, 2001; Riley, 2007). Indefiniteness, in opposition to definiteness, is seen as a corollary and contrastive quality that is normally ascribed to singular common nouns that are marked with the indefinite article (*a car*, *an egg*, *an idea*, etc.) and to plural nouns that have no article, possessive, or demonstrative form. However, even though many forms have been identified as being involved in the phenomena of definiteness and indefiniteness in English, the English articles are central and prototypical among them (as noted even in the terms used to describe them: *the definite and indefinite articles*). As such, in this study, the full range of expressions that

have been considered part of the definite-indefinite spectrum will not be addressed, and instead we will consider only nouns (and thereby non-pronoun noun phrases) marked with the indefinite and definite articles – *a/an* and *the* – along with nouns not marked with any article, determiner, or possessive (i.e., the so-called "zero-marked" nouns, since an account of articles must also account for when they are not used).

What exactly "definiteness" (and its corollary, "indefiniteness") is remains a matter of debate in the literature. Most accounts focus on "familiarity" or "identifiability" or some similar concept on the one hand or on "uniqueness" or "specificity" on the other hand, and some try to include both. Those who speak of definiteness mostly in terms of familiarity or identifiability are referring to the cognitive access of a referent's identity in the memory of the listener, and this access can stem from things such as previous mention of the reference or a related frame or context in the discourse (e.g., Christophersen, 1939; Chafe & Li, 1976; Du Bois, 1980; Heim, 1982, 1983; Chafe, 1994; Givón, 2001, 2005). Those who speak of definiteness mostly in terms of uniqueness or specificity are referring to the function of picking out individual referential members from a set or category of members (e.g., Russell, 1905; Lyons, 1977; Löbner, 1985; Kadmon, 1990; Hawkins, 1991; Abbott, 2001, 2011). Abbott refers to this approach as the "classical" approach. It should be mentioned that although the above two groups *mainly* focus on one or the other approach, they generally attempt to include within that approach all the phenomena that impinge upon definite and indefinite constructions – i.e., issues of both identifiability and specificity, in some form or another. There are those, also, who more explicitly include both aspects by trying to explain both a reference's cognitive accessibility (i.e., identifiability) and its individuation (i.e.,

specificity) within one overarching account (Fodor and Sag, 1982; Lambrecht, 1994; Berezowski, 2001).

An example of an approach that encompasses both identifiability and specificity is Berezowski's (2001:222-223), which assumes that the key function of the definite article is to convey that 1.) a mental representation of a referent exists that is shared in the "cognitive environment" between speaker and listener (i.e., identifiability), and 2.) this referent is unique in that shared cognitive environment (i.e., specificity). He further assumes that the key function of the indefinite article is to indicate exclusive reference, or in other words, to imply that a referent is a part of a set of at least two referents of the same kind (i.e., again, specificity).

Among those scholars who have linked definiteness with identifiability are Givón (2001, 2005) and Riley (2007). They both suggest that as soon as a referential NP is mentioned, listeners create a mental "file" that serves to identify that referent in subsequent discourse. In addition to this sort of identifiability based on previous mention, a number of other types of cognitive "grounding" also serve to identify referents. These can include grounding in the shared context (*Hand me <u>the hammer</u>*), grounding in shared knowledge of lexical items involved (*the President*), and grounding in a mental frame invoked by ideas in the discourse (*He walked into a restaurant and asked to see <u>the menu</u>*).

Two examples of scholars who link definiteness with specificity rather than with general identifiability are Lyons (1977) and Abbott (e.g., 1999, 2011). Lyons, for example, says "We can also distinguish those [referring expressions] which refer to some specific individual (or class of individuals) from those which (granted that they do have

reference) do not refer to a specific individual or class; and these we will call definite and indefinite expressions, respectively" (p. 178). There are a number of well-known cases, however, that demonstrate why this description alone is insufficient for explaining the full range of functions of articles. For example, in a phrase like (1), below, we can imagine someone describing a scene (perhaps looking out the window and telling a friend what she sees) in which it would be difficult to say that *man* does not refer to a specific individual.

(1) I see a man.

For this reason, it seems that definiteness could be linked to a number of considerations, including at least specificity and identifiability by the speaker vs. the listener. Lambrecht (1994:80-81) addresses this issue directly by describing (in)definiteness in terms of both specificity and identifiability, saying that "a specific indefinite NP is one whose referent is identifiable to the speaker but not to the addressee, while a non-specific indefinite NP is one whose referent neither the speaker not the addressee can identify at the time of utterance." By excluding the possibility of identifiability from non-specific NPs, Lambrecht may in effect be saying that the term "non-specific" is equivalent to "non-referential." (This idea will be discussed in more detail in the section on specificity beginning on p. 34.)

Abbott (1993, 1997, 2001, 2006, 2007, 2008, 2011) discusses definiteness at length and argues strongly for a theory of definiteness based on uniqueness, which is what she terms the classical view, similar to theories proposed by Russell (1905), Lobner (1985), Kadmon (1990), and Hawkins (1991). She argues against theories based on familiarity (i.e., a *listener's* familiarity or lack of familiarity with a referent in a

discourse) such as those espoused by Christophersen (1939), Heim (1982, 1983), Chafe (1996), since she finds many real-world data which require exceptions to familiarity-based rules (Abbott, 1999, 2011). However, her arguments appeal to very limited interpretations of familiarity-based accounts, focusing on the limitations of new vs. given explanations (i.e., whether a referent is "new" to the listener vs. familiar to the listener because it has already been mentioned in the discourse) but not including frame based accounts, which might be considered part of familiarity based accounts. Therefore, when such cognitive identifiability processes as shared lexical knowledge and frame activation are taken into account (ideas which are discussed below in sections 2.2.1.3 and 2.2.1.4, starting on p. 21), many of Abbott's arguments against theories of familiarity become less compelling.

2.1.2. Use of the term "definiteness" here

While many of the studies mentioned above focus on definiteness as a bona fide functional phenomenon, my current study focuses instead on what governs the use of English articles and sees them as serving a collection of overlapping semantic and pragmatic functions related to referentiality, identifiability, discourse purpose, and number. Therefore, this study will consider definiteness as a *structural* concept only, an approach that in the literature on definiteness, only few studies explicitly follow. Du Bois (1980) is one who does, suggesting that definiteness should be used as a strictly formal term, and that other semantic/pragmatic terms such as "identifiability" and "specificity" should be used when describing functional features. Likewise, Lambrecht (1994) proposes that definiteness is a formal, grammatical category, while identifiability

is a cognitive category, and that this is an insight which is crucially important when sorting out all the sometimes overlapping and confusing terminology on the subject.

In keeping with tradition, for purposes of this study, English "definite" constructions will be considered as NPs which are marked with the so-called definite article, *the*, along with the other aforementioned constructions generally considered to be definite (proper names, personal pronouns, possessed nouns, demonstratives, etc.). "Indefinite" constructions will be considered as singular NPs which are marked with the so-called indefinite article, *a* (or *an*, a phonological variant) and bare (or zero-marked) plural and uncountable NPs which lack an article, possessive, or demonstrative. Since what functionally constitutes "definiteness" has been so far a controversial and broad matter, by here using the term "definiteness" for a structural rather than functional category, we will sidestep this controversy, thereby freeing us to explore exactly what the functions of the definite and indefinite articles are (along with what functions are achieved by not using either article). This will be the focus of our investigation.

It should also be noted that this study focuses solely on how articles are used in Standard American English. The extent to which article use patterns vary in other varieties of world Englishes must for the time being remain a matter for future research.

2.2. Referentiality: Referential vs. Non-referential

The concept of referentiality can briefly be described as follows. *Referentiality* is a term used in the literature to talk about whether or not a noun phrase (NP) "refers" to an entity in the discourse. If it is intended to "refer," it is considered *referential*; if not, it is *non-referential*. If it is referential, it necessarily means that the NP refers to an entity (or

"object"; Du Bois, 1980) in a discourse that is *identifiable* by either the speaker, the listener, or both (Givón, 2001). Du Bois (1980) adds to this definition the idea that this NP can have continuous referential identity over time in the discourse. This requirement of some kind of identifiability status inherent in referential NPs (i.e., identifiable or non-identifiable) stands in contrast to *non-referential* NPs, for which the concept of identifiability does not apply, since no value of identifiability is assumed or intended by the interlocutors. Since the referentiality and identifiability of noun phrases in English affects how they are marked with articles, these concepts will be discussed in some detail.[4]

Riley (2007) emphasizes in his discussion of referentiality the idea that a referent is a mental representation (c.f., Givón, 2001:459). This observation is helpful since it allows discourse models to track this representation rather than any particular NP utterance. For example, a representation of a discourse entity may remain the same in the minds of interlocutors, even though the interlocutors may employ various synonyms, paraphrasis, or anaphora to refer to the entity throughout their discourse. Referentiality therefore becomes a property of the mental representation rather than of a specific linguistic form. This idea of a referent being a mental representation also allows for the fact that a reference need not correspond to anything in the "Real World" (as logical positivists such as Bertrand Russell, 1905, 1919, used the term). Instead, it can

[4] Since this use of the terms *referentiality* and *identifiability* often overlaps with the way some people use the term *definiteness* (as discussed above in the section on Definiteness on p. 3ff), it might be worthwhile to mention again here my decision to use "definiteness" strictly as a structural term. Since the presence of the "definite" article does not always co-occur with referential NPs, nor its absence with non-referential NPs (the presence of the indefinite article, *a*, also does not always co-occur with non-referential NPs, for that matter), using definite and indefinite as functional terms can inaccurately conflate functional and structural categories. Hence, in this study, definite and indefinite will be used to refer strictly to the structural form of NPs and not to any functional categories.

correspond to an entity in a "Universe of Discourse," which can include such things as imaginary or purported entities (e.g., Givón, 2001:438f).

Furthermore, what determines whether an NP is referential or not hinges neither on a referent's existence in a "Real World" nor on its *possible* "existence" in a Universe of Discourse, but on the speaker's intent, since the same NP can be used either referentially or non-referentially, depending on how the speaker chooses to construe it. Givón (2001) gives the following examples to illustrate this point:

(2) She's looking for a horse; it escaped last Friday.

(3) She's looking for a horse; it had better be white.

In (2), *a horse* is clearly referential, corresponding to an actual entity in the interlocutors' discourse. In (3), however, *a horse* is non-referential; it does not yet refer to any entity in the discourse that is intended to be identifiable *or* non-identifiable to the hearer. However, the non-referential status of an NP such as *a horse* in (3) can change if the speaker chooses to change his construal. For example, it can subsequently be used in a referential way in the discourse if the speaker posits its imaginary existence, as the speaker in example (3) has already begun to do by establishing *"it"* as being white. He could further establish this referential intention by saying something like, "I can see it now, walking towards her in the pasture, head held high, nostrils flaring; I'm sure this horse I'm imagining is exactly the horse she wants." Of course by saying something like this, the speaker would not be creating a real entity in the external "real world," but that is not ever a requirement for a noun to have referential status. The only requirement is that something is treated as a "real" entity in the "world of discourse," or, in other words, that there is a mental representation of an entity in the mind of the interlocutors.

2.2.1. Referential mentions and Identifiability

Since referentiality refers to the idea of whether an interlocutor treats a given noun as referring to an entity that "exists" in the universe of discourse and that is therefore *capable* of being identified in the discourse (e.g., Givón, 2001), *referential uses* are thereby traditionally divided into two (listener-oriented) identifiability types: identifiable and non-identifiable. An *identifiable* referent is one that the speaker treats as being identifiable by both the speaker and the listener, but a *non-identifiable* referent is one that the speaker treats as being only identifiable by the speaker herself at that point in the discourse. In *non-referential uses*, however (as mentioned above), the concepts of identifiability and non-identifiability do not apply. These various situations are further described below.

The data on referentiality and identifiability suggest that the referents of some noun phrases in a discourse share identifiability (and hence by definition, referentiality as well) between interlocutors, and that this identifiability may be due to a number of factors that flow from memory processes. The referents of some noun phrases, however, are not identifiable by the listener from memory since they are just being introduced into a discourse by the speaker, but they still refer to the concept that is being introduced. In addition, some noun phrases are not only not identifiable to the listener, but also intended to not refer to any "real thing" in the discourse. As mentioned above, article usage varies in a pattern that correlates with these referentiality and identifiability considerations.

Riley (2007) reasons that as soon as an utterance about a referent is made, both the speaker and hearer immediately establish a mental "file" (Lambrecht, 1994) of the referent, and therefore after that moment, the referent is identifiable. From this

perspective, it stands to reason that the first time a referent is mentioned that previously has not been identifiable, the identifiability becomes shared at and after the moment of the utterance. As Givón (2001) points out, initial grammatical marking on this NP referent by the indefinite article *a* serves as a signal from the speaker to the listener that a common cognitive "file" should be established.

There are a number of contextual factors by which the referents of a certain NP can be identifiable, however. Some of these factors include proper noun status (typically), shared speech context, shared lexical understanding, roles in contextual frames, as well as the aforementioned topic continuity status (i.e., anaphoric reference, or given information). Each of these will be treated in turn.

2.2.1.1. Identifiability and Proper Names

Proper names are generally considered to be labels for unique and thereby identifiable referents. Items (4) and (5) illustrate this typical usage of proper names.

(4) *Ben Franklin* was a prolific inventor.

(5) I'd love to tour *the White House*.

However, although this characterization of proper names usually applies, it is not always the case (Burge, 1973; Chafe, 1994; Berezowski, 2001; Anderson, 2007; Matushansky, 2008). For example, when a speaker says a sentence such as (6) he and/or the listener might be aware of a number of people named *Bob* to which he might be referring:

(6) *Bob* called. (Chafe, 1994)

Furthermore, a proper name can be formally marked as indefinite to show that the speaker wishes to present someone as non-identifiable to the listener, as in (7):

(7) *A Mr. Palermo* [lst mention], who had lived up here helping his uncles in the old days and had a cabin at the foot of the trail, came by at least twice a month. (Vonnegut, 1975:55; as quoted by Du Bois, 1980:218)

However, even in such cases, since the referent of proper nouns has the *capability* of being identified (and therefore can be described as either identifiable or non-identifiable in the discourse), this indicates that they have a clear referential status. To give them a non-referential meaning, one must resort to special constructions not normally used with proper nouns, such as adding the indefinite article in a clause with irrealis modality, as in (8):

(8) I've never met *an Anastasia*.

Berezowski (2001) argues that proper names are formally definite based on the similarity they have with certain basic referential functions of common nouns marked with the definite article (namely, existence and uniqueness in the "cognitive environment" of speaker and listener). He reasons that because of these similar functions, many languages use definite articles with proper names (e.g., Greek, Catalan, and Hebrew). However, in the case of English, the reason that the definite article is usually not used before proper names is arguably due to what he calls "language economy," or "the propensity of languages to forego encoding of grammatical features evident from the context" (p. 223). Since the English definite article can be used to signal different kinds of referential accessibility – such as that based on memory of previous mention or association with other referents in a discourse – its presence with proper names in some cases might activate unneeded referential cognitive processing. In short, in English, because obvious cases of proper names allow easy identifiability of their referents, the

definite article in those instances would be a redundant and possibly confusing inclusion. Berezowski identifies situations where such identifiability is relatively clear or obvious and where the definite article therefore tends not to be used. For example, the definite article will generally not be used with clear single-referent conceptualizations (*Benjamin Franklin* vs. *the Franklins*), with referents that have clearly defined borders or boundaries (*Greenland* vs. *the Arctic*), or with referents whose names are arbitrary rather than descriptive and hence obviously not common nouns (*Lake Erie* vs. *the Great Salt Lake*). These three situations can be summarized as situations where there is a single, bounded referent with an arbitrary (i.e., not descriptive) name. Where those conditions for proper names do not apply, the definite article will be used.

Berezowski (2001) lists six conditions or situations which will violate the "single, bounded referent with an arbitrary name" rule and which will therefore cause the definite article to be required with proper names. The three situations that violate the "single, bounded referent" part of the rule are 1) proper names with a collective referent construal, 2) proper names with an unbounded referent construal, and 3) proper names with an unclassified referent, while the three that violate the "arbitrary name" requirement are 4) descriptive proper names, 5) prenominal genitive structures, and 6) proper names with descriptive modifiers (p. 226-233).

Since such generalizable features govern the use of articles with proper nouns, Berezowski (2001) presents his model as a cognitive one based on the mental grammar of speakers: "Given the vast number of proper names used in any language (English included), it would be difficult to expect that [article collocations with proper names] are all simply memorized and there is no grammatical algorithm to derive them when needed

(especially if not used before)" (p. 234). For this reason, if these rules are indeed applicable to proper nouns generally, his model should easily find a place somewhere within a semantic map of English article usage.

2.2.1.2. Identifiability from shared speech context

The idea of shared speech context (Givón, 2001:460) (or a referent's "salient presence in the external environment," Chafe, 1994; or its "situational basis," Christophersen, 1939) is a key element of some instances of referentiality, and one that is so obvious that it might be considered a basic form of referentiality from which other types derive. When two interlocutors are in the same room, for example, looking at objects on the same table, the element of identifiability for all objects visible on the table is presupposed, such as when one person says to the other something like the sentence in (9):

(9) Hand me *the hammer*. (Givón, 2005)

Such pragmatic referring is arguably the origin of identifiable referents in human language. It is based on a short-term working memory process that does not need to rely on any longer term episodic discourse memory. Similarly, because such NPs are identifiable perceptually to everyone in the immediate environment of the speakers, they do not require an indefinite introduction when the object (such as *hammer*) is first mentioned (e.g., the indefinite form *a hammer* is not needed, since the definite *the hammer* signals the correct initial identifiability status in the discourse). (This applies at least when the issue of specificity does not come into play – i.e., where there is no choice between a number of similar objects, such as many *hammers*.)

2.2.1.3. Identifiability from shared lexical understanding

Some referents are globally accessible because they are represented as unique and identifiable items in the lexicons of all interlocutors participating in a given discourse. This can be described as identifiability from *shared lexical understanding* and is based on permanent culturally-specific semantic memory. The identifiability imparted by shared lexical understanding is in some cases similar to the shared speech context just mentioned, such as the mention of *the park* in (10):

(10) Are you hanging out at *the park*? (Chafe, 1994)

In the conversation from which this sentence was taken, both speakers shared a certain common experience of a particular park, and therefore their lexical denotation for a definite first mention of park was unambiguous for them. However, unlike the identifiability that comes from the above-mentioned shared speech context, in the case of a shared lexical understanding the thing being referred to does not have to be in the direct perceptible environment of the speakers. Clearer examples of this can be seen in (11)-(13):

(11) *The sun* came out. (The sun is known to all humans.)

(12) *The president* has resigned. (The president is known to all members of a nation.) (Givón, 2005)

(13) Call *the sheriff*! (The sheriff is known to the citizens of a county.) (Givón, 2001)

In these cases, speakers share a common understanding of what the nouns *president* and *sheriff* refer to, even if the objects or persons being referred to are not directly available to be pointed at.

2.2.1.4. Identifiability from Frames

The concept of contextual *frames* has been used by many researchers, following Fillmore (1968), as way to explain how the mention of some referents evoke the thought of other, somehow related referents in both speakers and hearers, and this mental linking of referents has in turn been offered as a way to explain the (otherwise predictable) identifiability associated with certain referential mentions. Givón (2001), for example, offers a model of frame activation that links the identifiability of nouns newly mentioned in the discourse with lexically stored semantic information of nouns previously mentioned in the discourse. The mental representation of these anaphoric references elicits a frame which licenses identifiable reference for these semantically-related new nouns (cf. Chafe, 1994). An example of this phenomenon is seen in (14):

(14) My boy missed school today, he was late for the bus. (Givón, 2001:461)

In this example, *bus* is marked with the definite article because its identifiability is facilitated by the frame evoked by the word *school*. The idea here is that *school* (and the concept of *missing school*) often involves buses, and therefore when the listener hears the phrase *missed school*, she can posit the existence of a particular bus that picks up the boy.

The exact cognitive mechanism by which frames work is still a matter of research, but there have been some attempts to offer a more precisely defined model of their operation. Barsalou (1992), Epstein (1999), and Riley (2007) are notable examples of those who have provided accounts of how frames work.

Baraslou (1992) characterizes frames as being the core cognitive process that allows human knowledge to be represented. If that is true, then it is arguably the case

that not only is the idea of frames important for explaining why some newly-mentioned referents are definite in a discourse, but also that frames provide the overarching structure within which the entire article system (and indeed all grammar) operates.[5]

Basically, Baraslou (1992) proposes that frames are (highly complex and often dynamic) sets of attributes that each have particular values, and that these collections of attribute-value sets have specified relations between them. Furthermore, the values of certain attributes can constrain the sort of values that other attributes have. He offers Fillmore's (1968) classic example of the verbal syntactic knowledge frame that people have of *buying*, as in (15):

(15) The artist (agent) buys paint (theme) at the art store (source) with a credit card (instrument).

Although Fillmore's discussion of frames emphasized the knowledge of speakers that enables verbal syntax to function meaningfully, this example can also be used to highlight the general conceptual frame structure offered by Baraslou (1992). The category *buying* involves attributes of *agent*, *theme*, *source*, and *instrument*, each of which has a value (such as *artist*, *paint*, *art store*, or *credit card*, respectively). These attribute-value sets have specified relations with each other (e.g., the agent *receives* the theme), and these relations can be constrained depending on the particular attribute-value sets involved (e.g., someone could buy paint – but not a new Volvo – at the art store).

This schema can be used to posit a way by which frames function to elicit mental representations of definite, identifiable "new" NPs in discourse. When any given referent

[5] Under Baraslou's (1992) conceptualization, semantic maps – such as the ones proposed in this current study (discussed below on p. 52) – could be considered frames, and frames of different scales and purviews could operate within a semantic map to accomplish various and myriad functions. In this way, frames can be seen as fractal.

is mentioned (e.g., *vacation*), a frame is evoked (e.g., a *vacation* frame). This frame consists of a number of attributes that are naturally associated with this category of reference (e.g., *place*, *cost*, *transportation*), and each of these attributes has a given value or range of possible values, and the values may be constrained by each other. When the speaker mentions any attribute or value that can normally be assumed to exist (or likely exist) as part of the frame, she can assume it is identifiable by virtue of its inclusion in that frame and can therefore mark it structurally as a definite reference. An example can be seen in (16):

(16) We had a great vacation, but *the plane* ride was scary.

In this example, we see that *plane ride* in (16) requires a definite marking (i.e., the definite article, *the*) since the frame evoked by *vacation* includes a *transportation* attribute that must be filled by one of several "likely" values of how the participants got to the vacation spot (these can be "stereotypical" or "conventional" or even merely "possible" values; Epstein, 1999). The speaker can assume that any of these likely values (such as *plane ride*) will be identifiable to the listener because the listener shares the same (or very similar) frame.

Epstein (1999) does not attempt to build a model similar to the one in Baraslou (1992). However, he, like Baraslou, also uses the term "value," but he contrasts it with the term "role" rather than "attribute." It seems reasonable, based on Epstein's usage, that the two terms "role" and "attribute" can be used more or less synonymously.

An example of Epstein's "roles" can be seen in (17):

(17) Like sex, crime can be brief and messy: more about buildup and aftermath than event and arrival. So **the gun** is fired, **the police officer**

dies, and **the diamonds** are stolen. So what happened afterward and how did the relevant players get there, in what kind of car, and did they wear clean underwear? Thus far, the film maker Quentin Tarantino, whose second movie, "Pulp Fiction," won the 1994 Palme d'Or at Cannes, has aimed his writing and directorial focus strictly on crime.... (Epstein, 1999:57; emphasis his)

Epstein uses this example to show that *gun*, *police officer*, and *diamonds* serve as "roles" (but not "values") of the *crime* frame. His argument is that *gun*, *police officer*, and *diamonds* are used as if they are the typical categories of a *crime*, and hence they are "roles" of the frame, not mere "values" (since "values" here would presumably be something like a particular *gun*, a particular *police officer* with a name, and some particular *diamonds*). For this reason, Epstein's "role" does appear to be very similar to Baraslou's "attribute," both serving as terms for the categories that are stereotypically associated with a frame.

Epstein (1999) shows that the evocation of frames does not require an explicit mention of the frame itself, but instead can occur simply by mention of values and/or roles within the frame. He gives the following example of this phenomenon from an article in the L.A. Times in which a writer was bemoaning the loss of the city's professional football teams:

(18) So we lost the Rams and Raiders. Lost our innocence. But hold **the flowers**. Put away **the handkerchiefs**. Stop **the sobbing**. We still have the Rose Bowl, don't we?! (Epstein, 1999:58; emphasis his)

Here we see the frame *funeral* or *mourning* evoked, even though no term like *funeral* or *mourning* was explicitly mentioned, but only terms that are stereotypical "roles" (in Epstein's terms) of the funeral/mourning frame.

Epstein (1999) also notes that frames can be evoked even with a "role" that is not at all stereotypical of a frame. Speakers who do this creatively coopt the linguistic constructions that normally evoke frames in order to "force" novel frame aspects, treating them as if they were actual stereotypes. In item (19), Epstein offers an example from a New York Times article discussing the TV show *Roseanne*:

> (19) When "Roseanne" first appeared in 1988, the Conners were refreshingly realistic; the two working parents worried about money, and everyone yelled all the time ... But seven years involves a lot of plots, and the once-ordinary Conner family has become more socially troubled ... In the last few seasons alone, Roseanne's mother was arrested for driving while drunk ... And last year, Roseanne's grandmother turned up at Thanksgiving dinner with a new husband and revealed that she had had two illegal abortions when she was young. "Gee, I wonder who told **the abortion stories** at the very *first* Thanksgiving," Roseanne said with a redeeming, acerbic twist increasingly rare for her. (Epstein, 1999:60; bolded emphasis his, italics in the original)

Here we see that the use of the definite article suggests that *abortion stories* are part of the *Thanksgiving dinner* frame, even though, of course, they aren't (even, presumably, for the fictional Connor family as portrayed in *Roseanne*). Therefore, as Epstein

(1999:63) notes, this kind of use of the definite article in the broader "frame" construction serves as a "marker of the speaker's intention to create a new role" (or, in Baraslou's term, a "new attribute") for the frame.

Riley (2007) attempts to assign some precision and standardization to the terminology associated with referentiality as a whole, and the concept of frames figures prominently into his overall scheme. Although more detailed in some ways – i.e., described with precise formulaic definitions and illustrated carefully with many examples – his overall treatment is similar to that of Givón (2001) and Baraslou (1992). One addition he makes is the inclusion of referents in "the immediate situation," referents known within "general knowledge, or the larger situation," and "anaphoric reference" (or what we have been calling here "shared speech context" and "shared lexical understanding," and "'given' or 'old' references," respectively) as subspecies of frame reference, which they clearly can be seen as, as the following examples illustrate.

First, (20) and (21) show references to entities in the immediate situation (or "shared speech context") which are part of the cognitive frame triggered by things that are perceptible in the surrounding context:

(20) '*The rain* is heavy, isn't it?'

(21) '*The butler* will show you out.' (Riley, 2007:857; italics mine)

In the case of *the butler* or *the rain*, these referents can be clearly perceived in the environment in which the speaker is located. *The rain* can be perceived directly, which triggers a cognitive frame that allows identifiability. The frame for item (21) might be a large estate house with household servants. A *butler* can be assumed by virtue of perceiving that context, even if he is not directly visible. Indeed, Riley's point is that

everything perceived by a speaker or listener in their immediate environment triggers a cognitive frame of perceptibly and/or stereotypically associated attributes or values, so reference to any of them will be identifiable by virtue of this frame.

Second, (22) shows an example of how general knowledge of the larger situation (or "shared lexical knowledge") can create identifiability by virtue of frame-based association:

(22) Today we elected *the governor of California*. (Riley, 2007:857; italics mine)

In this case, *the governor of California* is part of a general frame of knowledge about the organization of governments in U.S. states. Since everyone in the audience shares this general frame, the speaker can assume the listeners know that California has one and only one *governor* and that this role is therefore identifiable.

Third, (23) illustrates how anaphoric reference can be considered part of a frame.

(23) A huge trailer truck overturned this afternoon on Route 1. The vehicle was rounding a curve at high speed, and tipped over. (Riley, 2007:858)

Here we see *a huge trailer truck* marked with the indefinite article, followed by *the vehicle* marked with the definite article. As he reasons, the first mention of the referent (*a huge trailer truck*) triggers a frame which allows identifiability for any subsequent mention of the same referent (*the vehicle*), just as the frame would also include any associated attributes or values.

What Riley (2007) has done is broaden the definition of frame reference to include a number of referential processes that others have listed separately. What others

have normally called "frame reference," he identifies as a subtype of a more general cognitive frame reference activation process. This subtype he names "bridging or associate frames," since these kinds of frame create a mental bridge or association between a mentioned referent and one or more unmentioned ones.

Even though Riley's point may be sound, lumping together references to shared speech context, references to shared lexical items, and anaphoric reference and treating them all as part of frame reference helps little in the task of organizing a semantic map of English article use, since all these categories must be included in the map either way. Therefore, this current study will treat references to shared speech context, references to shared lexical items, and anaphoric reference each as their own categories and separate from the category of "frame reference," even while acknowledging that they may all indeed share similar underlying cognitive activation processes. When we use the term "frame reference" here, we will follow what Riley means when he refers to the "bridging or associate frames" subspecies under his general frame reference rubric.

In the end, we are left with the idea that when a frame is evoked, the lexical semantics of a noun are used to identify a related noun newly mentioned in the discourse. The perspectives of Baraslou (1992), Epstein (1999), and Riley (2007) are attempts to provide explicit details about how mentions of referents function to evoke frames, whether by explicitly naming them, explicitly naming some of their stereotypical "roles" or "attributes" or "values," or appealing to the cooperative imagination of speaker and listener to creatively invent novel roles/attributes that are then ascribed to the frame.

2.2.1.5. Identifiability and New references vs. Old references

The status *new reference* vs. *old reference* is a dichotomy that many researchers have discussed as having an effect on the formal *definiteness* and functional *identifiability* of *referential noun phrases* (Chafe, 1970, 1994; Du Bois, 1980, Epstein, 1999; Givón, 2001, 2005; Riley, 2007). As Givón (2001, 2005) succinctly describes them, "old" (or "given") references are *grounded in the current discourse*, in the sense that they have already been referred to previously in the discourse: after their first mention, the concept to which they refer has been cognitively activated in the memory of the listener. Therefore, the cognitive activation of the referent in subsequent mentions allows the referent to be identified, and these subsequent references may thereby be marked with *the* to signal their identifiable status.

This identifiable status of noun phrases that refer to previously-mentioned referents contrasts with the *non-identifiable* status of "new" referential noun phrases which are newly introduced into the discourse and which have none of the other, previously discussed types of grounding to make them identifiable to the listener (i.e., no grounding as a *proper name*, in a *shared speech context*, in a *shared unique lexical understanding*, or in a *frame*). This contrast can be seen in (24), in which new references (labeled "1st mention" in (24)) are marked with the indefinite article as non-identifiable but are then marked with the definite article as identifiable on subsequent mentions ("2nd mention").

(24) then *a boy* [1st mention] comes by,… on *a bicycle* [1st mention]; the man is in the tree,… and *the boy* [2nd mention] gets off *the bicycle* [2nd mention] (Du Bois, 1980:206)

Chafe (1994) makes a point of using the term "new" to indicate a referent that is newly mentioned in the current discourse, whether or not that referent is identifiable since, indeed, some new referents are identifiable, as in (25):

(25) ... I talked to Larry last night, (Chafe, 1994)

In this case, even though Larry is known to both interlocutors in the conversation, he is just being introduced in the discourse. Similar cases of referents being identifiable even though they might also be newly mentioned are seen in (26) - (28), below:

(26) *The sun* came out. (*The sun* is known to all humans as a unique referent.)

(27) *The president* has resigned. (*The president* is known to all members of a nation as a unique referent.) (Givón, 2005)

(28) Call *the sheriff*! (*The sheriff* is known to the citizens of a county as a unique referent.) (Givón, 2001)

An additional example (29) illustrates that a referent evoked by a frame can also be newly mentioned in a discourse but also identifiable, where *cover* is only newly mentioned but is identifiable based on its association with *book*:

(29) I bought a book but *the cover* was torn. (Epstein, 1999:54)

In these cases, we see that the simple dichotomy "new reference" (i.e., first mentioned) vs. "old reference" (i.e., subsequent mention) is not enough to account for the patterns of definiteness seen in all noun phrases in a discourse.

The above general use of "new" contrasts with the way that Riley (2007, following Prince, 1992), uses the term. In Riley's (2007) usage, a referent is considered "new" (or "brand new") only if it has not been evoked by a previous frame. Therefore,

for Riley, in item (29), above, the term *cover* would not be considered "new" since its mental representation would presumably have already been evoked by the frame established by the previously-mentioned term *book*. This use of the terminology helps Riley establish the careful definitions he is attempting, and it allows the use of the term *new reference* to serve as a useful indicator of when newly mentioned referents will be non-identifiable (contrasted with other terms like *frame*, which indicates a referent will be identifiable). In this current study, we will follow this approach, and we will further specify that *new reference* stands as a separate category in the list of identifiability-determining situations, in contrast not only with *frame* grounding, but also groundings in *proper names*, *shared speech context*, *shared unique lexical understanding*, and the *current discourse*. Therefore, only when a newly mentioned referent cannot be identified based on one of these other kinds of groundings, it will be considered a *new reference.*

Gundel et al. (1993) propose an implicational Givenness Hierarchy (Figure 1) of six cognitive statuses that correspond to the degree to which a referring expression is assumed to be cognitively activated or relevant to an addressee.

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
{it}	{that, this}	{that N}	{the N}	{indefinite this N}	{a N}

Figure 1. Givenness Hierarchy with English examples of forms corresponding to each status (Gundel et al., 1993).

The lower two (i.e., rightmost) statuses correspond to the indefinite uses, and the highest four to the definite uses (Abbott, 2006). However, the use of the articles alone cannot

indicate which of the higher statuses a referring expression occupies. To determine this, other definite constructions besides articles (e.g., *it, that, this*) must be used. Givón (2001) also situates definite article usage within a broader perspective of memory activation as expressed in various anaphoric grammatical devices. His list of these devices include the following, presented in order from presumed most given to least given (i.e., in the same order as the Gundel et al. list, though not perfectly aligned with it):

> a. zero anaphora, b. unstressed PRO, c. stressed PRO, d. Y-movement, e. Def-noun, f. Def-noun with modifier(s), and g. L-dislocated Def-N.

The phenomena of English article usage do not fit within all of the items included on Gundel et al.'s (1993) list or all the anaphoric devices suggested by Givón (2001), but instead fit within a limited set of the full range of all definite statuses. A full model of definite expressions would therefore require the inclusion of many more functional and structural categories than a model that just accounts for the expression of English articles alone (which the current study is limited to).

Conversely, even though the Gundel et al. (1993) Givenness Hierarchy encompasses a broad range of definite expressions, it alone is not sufficient to account for all the natural language English article data. For example, although they mention that "the status 'referential' is necessary for appropriate use of all definite expressions" (p. 276), there are situations in which that is not actually the case. Du Bois (1980) lists examples in which a definite construction is non-referential, such as the term *the banjo* in (30):

(30) Somebody in Dullingham Junction was playing *the banjo*. (Du Bois, 1980:216)

The non-referential status of *the banjo* in this example becomes more obvious when it is considered that it is (nearly?) synonymous with *a banjo* in (31):

(31) Somebody in Dullingham Junction was playing *a banjo*.

Therefore, although the Givenness Hierarchy may account for most data, a multi-dimensional model may be needed to account for a number of situations where article usage does not conform to an hypothesized "normal" pattern.

2.2.1.6. Specificity

The discussion of definiteness, above, pointed out the often alternate approaches taken by researchers between describing definiteness in terms of identifiability on the one hand and specificity on the other. As mentioned earlier, Lambrecht (1994:80-81) links the two approaches by offering the interpretation that "a specific indefinite NP is one whose referent is identifiable to the speaker but not to the addressee, while a non-specific indefinite NP is one whose referent neither the speaker nor the addressee can identify at the time of utterance" and seems to suggest that the latter may be considered non-referential, since it lacks the ability to be (specifically) identified. It should at his point be noted that Lambrecht here makes a key contribution to the discussion on specificity by making it a property of identifiability. However, his suggestion that the lack of identifiability of non-specific mentions entails their non-referentiality deserves some careful consideration. As I have already argued, referential NPs have the attribute of identifiability; that is to say, they can be either identifiable or non-identifiable, with both

terms referring to the *listener's* assumed ability to identify any given referent. Non-referential NPs, however, do not possess the attribute of identifiability. They cannot possibly be identified as specific referents since they do not pretend to refer. Therefore, in considering whether Lambrecht's suggestion that non-specific NPs are indeed non-referential is true, we must determine which of the following is the case. Either (a) they merely cannot be identified as specific referents by either the speaker or listener, even though they do indeed refer to an individual referent (though one that is currently unknown with precision), or (b) they are not intended to refer to a specific individual referent at all. In the former case, they would be referential; in the latter, non-referential.

Certain data suggest that the answer is the former, that non-specific NPs are indeed referential – but are just non-specific. Consider (32) - (33) (both adapted from Cormack and Kempson, 1991:547):

(32) *A student* in the syntax class is cheating, and I'm going to confront him about it today.

(33) *A student* in the syntax class is cheating, but we don't know which one it is.

In the first example, (32), the identity of the *student* is clearly known to the speaker but is not assumed by the speaker to be identifiable to the listener. This is a classic case of using the indefinite article to introduce a new referent in a discourse. However, in the second example, (33), the particular identity of the *student* is unknown to both the speaker and the listener, a common situation in discussions of specificity but one which is not covered under the terms "old reference" vs. "new reference" in most models of

identifiability. But in both examples, the term *a student* definitely refers to an actual existing entity in the world of the discourse. In no way would it be accurate to say that there is *not* an actual student in this scenario, which would be the requirement for us to properly term the reference "non-referential." Therefore, such data suggest that specificity (which includes specific and non-specific mentions) is a cognitive category that only applies to *referential* mentions, one that we might say involves a third type of "identifiability" that is normally not considered under that rubric – namely, *non-identifiability to the listener and the speaker*.[6] Christophersen (1939) and Du Bois (1980) (who quotes Christophersen on this point) mention this type of non-identifiability to both listener and speaker, saying that is it often marked by use of the word "some," as in (34):

(34) I have read it in some book. (Christophersen, 1939:188)

In this example, the speaker himself likely cannot recall which book he "read it in." Du Bois comments that the contrast between *a* and *some* in such cases is often used precisely to mark a distinction between identifiability and non-identifiability to the *speaker*. While this distinction may prove common, we have seen that it is not always used, and that sometimes *a* can indeed be used where this lack of speaker-based identifiability applies.

Under such a view, one in which specificity is only relevant for referential items which involve this third type of "identifiability" (i.e., non-identifiability to the listener *and* the speaker), we must consider how to categorize "specificity" and "identifiability" as referentiality subtypes. We could either 1) keep the category "specificity" as one that is functionally salient and thereby divide referential mentions into two types – specific and non-specific – and then under the "specific" type we would subsume identifiability

[6] As discussed previously, identifiability is normally seen as a measure of the *listener's* presumed ability to identify a reference in the discourse. The *speaker's* ability to do so is traditionally assumed under both the identifiable and non-identifiable types of "identifiability."

with its two types – identifiable (to the speaker and listener) and non-identifiable (to the listener only). Or, 2) if we did not see a compelling reason to keep the term specificity, we could do away with it and instead of using the label "non-specific," simply adopt a third type of "identifiability" – one that is non-identifiable both to the speaker and to the listener. The two options are diagrammed below in Figure 2 and Figure 3:

Referentiality Types		
Referentiality Status	**Specificity Status**	**Identifiability Status**
Referential	Specific	Identifiable to speaker only
		Non-identifiable to listener only
	Non-specific	
Non-referential		

Figure 2. Keeping the category "Specificity."

Referentiality Types	
Referentiality Status	**Identifiability Status**
Referential	Identifiable to speaker and listener
	Non-identifiable to listener only
	Non-identifiable to speaker and to listener
Non-referential	

Figure 3. Doing away with the category "Specificity" by subsuming "non-specific" as third type of identifiability.

However, we are still left with examples such as the bare-bones (35), in which it is unclear (to the listener) whether the speaker means or does not mean that he knows the specific identity of the student.

(35) A student in the syntax class is cheating.

Is this a case that requires the category "specificity," since the listener is unclear as to whether the term "*a student*" is specific or not? Although the meaning of the utterance is ambiguous to the listener, it is not the case that the speaker is unclear about the situation. Therefore, for a speaker-based (i.e., production-based) model of English article use, the category "specificity" could be made redundant by describing "specific" and "non-specific" in terms of identifiability – whether a referent is identifiable to the speaker or not. However, for a listener-based (or interpretation-based) model of article use, the category "specificity" may be required as one type of the phenomena of ambiguity resolution. Since this study seeks to build a speaker-based model, the category "specificity" appears unnecessary, and therefore for simplicity's sake the third type of identifiability will be included in the model instead (as in Figure 3, above).

2.2.2. Non-referential mentions

As discussed above in section 2.2 on referentiality, some NPs are not intended ever to be identifiable when they are mentioned; that is to say, they are not intended by speakers to have the property of identifiability. These NPs are therefore *non-referential* since they are not intended to refer to any existing entity in the world of the discourse.

Du Bois (1980:209-217) offers a careful treatment of English article use in non-referential mentions, and our analysis will stem from the types that he discusses. His list

consists of the following kinds of uses associated with non-referential noun phrases: 1.) *membership in a class*; 2.) *categorizing predicate nominals*; 3.) *comparatives*; *negation* – which I have divided into two types: 4.) *negation of a verb phrase* and 5.*) negation of a noun phrase* – 6.) *compounds*; 7.) *secondary predicates*; 8.) *performatives*; 9.) *vocatives*; and 10.) *predicate conflation* (or what I here call *noun incorporation*). The first four of these have patterns of article use that conform to the traditional accounts of non-referentiality, but the remaining six do not. Table 1 lists all of these types, together with Du Bois' indications of their patterns of article use, and gives examples of each type.

Table 1. *Types of Non-referential constructions. (All examples and page numbers are from Du Bois, 1980, unless otherwise noted.)*

	Singular	**Plural**	**Uncount -able**	**Type of Non-referential construction**	**Examples**
Categorization:					
1.	a	Ø	Ø	Membership in a class (p. 210f)	▪ Mary's <u>a forester.</u>
2.	a	Ø	Ø	Categorizing predicate nominals (p. 213)	▪ it's <u>a young woman</u>
3.	a	Ø	Ø	Comparatives (p. 214)	▪ who looks like <u>a Mexican-American</u> ▪ they look like <u>Ø bullies</u>
4.	a	Ø	Ø	Negation of a VP	▪ It's not <u>a conversation</u>. (my adapted example)
Negation of NP:					
5.	Ø	Ø	Ø	Negation of a NP (p. 210)	▪ there's no <u>conversation</u>
Compound Modification:					
6.	Ø	Ø	Ø	Compounds (Du Bois 1980, p. 209)	▪ [a [<u>pear</u>] tree] ▪ *[a [<u>a pear</u>] tree] (my example) ▪ *[[<u>a pear</u>] a tree] (my example)

Table 1, continued.

	Singular	**Plural**	**Uncount -able**	**Type of Non-referential construction**	**Examples**
Secondary Predicates:					
7.	Ø	Ø	Ø	Secondary predicates (p. 213)	▪ The gardener of the convent, being chosen Ø muleteer, led out the two mules.
8.	Ø	Ø	Ø	Performatives (p. 212)	▪ I pronounce you Ø man and wife.
Vocatives:					
9.	Ø	Ø	Ø	Vocatives (p. 212)	▪ Ø Buddy, could you spare a quarter? ▪ Hey, Ø man, can't you read the sign?
Noun Incorporation:					
10.	a the Ø	Ø	Ø	Noun incorporation (p. 214ff) [idiosyncratic collocational use of *a/the* - p. 216]	▪ Somebody in Dullingham Junction was *playing the banjo* ▪ …was *playing a banjo* (my adaptation) ▪ …was *playing Ø banjo* (my adaptation) ▪ We *knocked them on the head*. ▪ the guy who is *picking Ø pears* ▪ *…picking a pear or two* (my adaptation ▪ *…Ø pear-picking*. ▪ I only wear *one* [i.e., one contact lens] in my left eye when I'm *wearing my lenses*

2.2.2.1. Categories

Types 1 – 4 in Table 1 seem very similar functionally (they all deal with categorization) as well as structurally (they are all predicate nominals). (For examples of the data in this section, see Table 1.) Therefore they should arguably be grouped together

under Du Bois' label of *categorizing predicate nominals* (type 2), or perhaps labeled simply *category*.

In regard to types 4 and 5, whereas *negation of a verb phrase* is simply a way of excluding something from membership in a category and shares the indefinite pattern with other categorizing constructions, the *negation of a noun phrase* construction follows neither the definite nor indefinite pattern, but instead is marked with the zero form in all grammatical numbers. This is a pattern shared with quantified noun phrases generally (i.e., *two conversations, one conversation, no conversation*) and suggests that the negation word *no* should be treated as a quantifier as well. Therefore it does not need to be given a special referentiality status in a model of article usage.

2.2.2.2. Modifying nouns (i.e., nouns as adjectives)

Compounds (type 6) contain two (or more) prototypical noun roots in which one modifies the other. From a theoretical point of view, since the modifying noun roots never vary in their article use pattern and since they behave distributionally more like adjectives than nouns and can in fact be treated as adjectives, there might be little need to account for them in a model of article usage. However, there might be practical reasons to do so, such as to show a comprehensive treatment of nouns (or noun-like forms), perhaps at the very least for the benefit of non-native speakers who are learning or researching English.

2.2.2.3. Secondary predicates

Du Bois (1980) lists *secondary predicates* and *performatives* (types 7 & 8) as separate categories. However, the example he lists as performatives ("I pronounce you man and wife") is also a clear secondary predicate example and follows the same "Ø article" pattern; for this reason I propose to include them both under the category *secondary predicates*. While Du Bois offers only a cursory treatment of these types of constructions, others such as Anderson (2007) and Matushansky (2008) go into more depth. Their basic pragmatic function is to give names to referents or to formally place them into categories or roles, as in (36):

(36) The queen appointed her lover Ø *treasurer* of the realm.
(Matushansky, 2008:579)

There are also types of secondary predicates that do not follow the Ø article pattern, as in (37) and (38):

(37) The queen appointed her lover *the treasurer* of the realm. (adapted from Matushansky, 2008:579)

(38) The queen appointed her lover *a treasurer* of the realm. (adapted from Matushansky, 2008:579)

Therefore, it appears that secondary predicates often have the option of omitting the article, but an article can be included if a speaker wishes use it to signal one of the typical article functions, such as to *identify* or to *categorize* a referent. Because the article usage pattern for secondary predicates is not completely predictable from the data currently studied, further research is needed to arrive at a full understanding.

2.2.2.4. Vocatives

Vocatives (type 9) are a special case of non-referential usage. Both Du Bois (1980) and Anderson (2007) mention the fact that vocatives have some affinity with referential nouns – and particularly with proper names (Berezowski, 2001) – in that they when they are used, a referent can be identified. However, Du Bois argues that vocatives nonetheless should be considered non-referential (or non-definite, in Anderson's terminology) since their primary function is "not to refer to the addressee but to attract his attention or index his social position" (Du Bois, 1980:212). Anderson further argues against their referentiality by noting that they often cannot be used in argument positions (especially in "default vocative names" such as *Mac* or *Honey*), as seen in (39) and (40):

(39) *Honey*, did the dentist call you back?

(40) *The dentist called *honey* back.

Therefore, for our current purposes here, vocatives will be considered as a type of non-referential mention.

2.2.2.5. Noun Incorporation

Noun incorporation in English (type 10) is typically seen as involving nominalizations, participles or finite verb constructions where the object of the verb is moved from after the verb or verbal item and placed before it instead, thus binding the verb and its object into a tight semantic and syntactic unit (Rice & Prideaux, 1991). Du Bois describes noun incorporation (or *predicate conflation*, as he names it) as involving a noun phrase used together with a verb "to express a unitary concept rather than to refer to an actual object" (Du Bois 1980:214). The verb-noun unit is seen as expressing a single,

whole event, and the noun is therefore not cognitively construed as being referential. Rice & Prideaux (1991) list six constructions (items (43)-(48), below) that correspond to what they call the "unincorporated finite verb counterpart" (42). Their example of an unincorporated finite verb is a non-referential one, so I have added a referential example for comparison in item (41).

(41) *Unincorporated forms - finite verbs with referential objects:*

He lifts/lifted a weight. He lifts/lifted the weight. (my adapted examples)

(42) *Unincorporated forms - finite verbs with non-referential objects:*

He lifts/lifted Ø weights professionally. (Rice & Prideaux, 1991:3)

(43) *Incorporation in finite verbs:*

He Ø weightlifts/weightlifted professionally. (Rice & Prideaux, 1991:3)[7]

(44) *Incorporation in infinitives:*

He used to Ø weightlift professionally. (Rice & Prideaux, 1991:3)[8]

(45) *Incorporation in progressive participles:*

He's Ø weightlifting as part of his training program. (Rice & Prideaux, 1991:3)

(46) *Incorporation in participial adjectives:*

The Ø weightlifting competition is next. (Rice & Prideaux, 1991:3)

[7] Rice & Prideaux (1991:3) list the noun-incorporated verb *weightlifts* as unacceptable in this example. However, two decades later, Google searches in 2012 reveal several instances of this verb in common usage on the Web, as in "Because I'm not someone who weightlifts professionally." (Weight perception - Lonely Planet travel forum. (n.d.). Retrieved September 12, 2012, from http://www.lonelyplanet.com/thorntree/thread.jspa?threadID=1869351)

[8] Again, although Rice & Prideaux (1991) list this example as only questionably acceptable, Google searches in 2012 return a number of examples of the infinitive "to weightlift" in usage on the Web.

(47) *Incorporation in gerunds:*

Ø Weightlifting is a good complement to aerobic exercise. (Rice & Prideaux, 1991:3)

(48) *Incorporation in agentives:*

He's a champion Ø weightlifter. (Rice & Prideaux, 1991:3)

It should be noted that in all of the clear noun-incorporation examples in (43) through (48), the pre-verbal object is marked with Ø, the absence of an article. This Ø serves as a signal of the clear lack of referentiality for these nouns.

Although items (43)-(48) are clear examples of English noun incorporation, it seems to be the case that some typical finite verbs with *post*-verbal non-referential objects also can exhibit some semantic and syntactic features of noun incorporation. This possibility becomes apparent when such cases (as in (42)) are compared with situations where finite verbs have referential objects and are therefore clearly *not* cases of noun incorporation (as in (41)). Du Bois (1980) offers evidence to show that such verb-object constructions can sometimes behave like a noun-incorporated pair. In such constructions, the finite verb is followed by an object that *could* be placed before the verb to form a prototypical noun incorporation, but instead the object is not moved and maintains its post-verb position. The noun incorporation is evidenced semantically by the non-referentiality of the object, which in turn is signaled syntactically by the fact that (in at least some cases) the object is not susceptible to taking a definite article, even when it follows a previous use of that same noun. An example of this phenomenon is seen in (49), where *pears* have already been referred to in the narrative in the phrase *a guy who's*

picking pears [1st mention] and therefore could conceivably be considered to be identifiable in subsequent mentions.

(49) …And.. um… the guy who is picking Ø pears [2nd mention], um… um.. picks the pears and puts them in a.. in um… these baskets that he has… (Du Bois, 1980:214)

However, the underlined *pears* in this second mention, marked with the Ø article pattern, arguably does not refer to actual pears – since if it did, it would be marked with *the* to signal identifiability based on previous mention. Instead, it is only used to speak about the generalized, abstracted action of the noun-incorporated construction "pear-picking." In demonstration of the distinction between the non-referential and referential uses – and thereby noun-incorporated versus non-noun-incorporated verb-object constructions – the speaker afterward marks a referential mention of *pears* (in that they are mentioned as being placed in baskets) by using the definite article:

(50) [He] picks the pears and puts them in a.. in um… these baskets that he has…

The data above exemplifies that nouns in full noun incorporation constructions (i.e., items (43) to (48) above) are marked with the Ø article pattern in English. Furthermore, as we see in item (42), some nouns in verb-object constructions that exhibit semantics typical of noun incorporation also are marked with the Ø article pattern. Therefore, the Ø article pattern might be thought of as the prototypical pattern associated with noun incorporation. However, this pattern does not apply in all cases of verb-object constructions with noun incorporation semantics, as the following examples illustrate.

(51) Somebody in Dullingham Junction was *playing* *the banjo* ((Du Bois, 1980:213, quoting Christophersen, 1939)

(52) Somebody in Dullingham Junction was *playing* *a banjo* (my adaptation)

(53) Somebody in Dullingham Junction was *playing* *Ø banjo* (my adaptation)

In items (51) to (53), we see that all three article patterns evoke very similar semantics. The differences in meaning between the three are extremely subtle and in some contexts could be virtually indistinguishable. Obviously, *the banjo* in item (51) and *a banjo* in item (52) could certainly be referential in some contexts, but they could be non-referential in others, and this non-referential semantics is similar to the full noun incorporation construction "banjo playing." In this case, it may be that the semantics of the idea of "playing banjo" is so firmly conceived as a "unitary concept" (Du Bois, 1980) that it does not highlight the banjo as an actual object but instead highlights the holistic nature of the action expressed by the verb together with the noun. If so, it may be that the particular article used with the noun does not matter: even the presence of *a* or *the* cannot always force a non-noun-incorporated referential meaning in such constructions (depending on the context, of course).

When the whole range of noun incorporation instances is considered, we see that the data in (43) to (48) suggests that for full noun incorporation constructions – i.e., object-verb and object-verbal constructions – the nouns will be marked with the Ø article pattern. For the type of data in (51), (52), and (53) – i.e., constructions that have finite verbs followed by non-referential objects and that have semantics similar to full noun

incorporation constructions – we see that although the Ø article pattern is possible, it is not always the only possible pattern, and that both *a* and *the* also are sometimes possible.

In summary, in positing the categories for inclusion in the non-referential component of the larger referentiality dimension of a semantic map of English articles, I propose to collapse some of Du Bois' descriptive categories together while at the same time dividing the noun incorporation category into two, leaving in the end five distinct non-referential usage types: *categorization*, *secondary predicates, vocatives*, *full (object-verb) noun incorporation,* and *finite verb (verb-object) "noun incorporation"*.

2.2.2.6. Idioms

A final type of non-referential use of nouns are the nouns in idioms. Idioms are stock phrases that do not follow the normal variation in article use patterns because their article-noun correlations are lexically rather than functionally determined. This lexically-based article usage can be seen in examples (54) and (55), below, where the second sentence in each pair is unacceptable on the idiomatic meaning.

(54) It's raining Ø cats and Ø dogs.

*It's raining the cats and the dogs.

(55) He kicked the bucket.

*He kicked a bucket.

Since the nouns in idioms are clearly non-referential (in the standard sense that they do not refer to any existing entity in the world of discourse), idioms must be included in the non-referential section of a model of English articles. However, since the article usage

patterns cannot be predicted but must be retrieved from memory, it will be necessary to specify that fact in the model.

2.3. Grammatical Number

Although there are some problems associated with rigidly classifying English nouns into singular (SG), plural (PL), and uncountable (UC) lexical class categories (e.g., a word such as *cake* can function as a singular, plural, or uncountable noun – *a cake*, Ø *cakes*, Ø *cake*), it is obvious that these categories are nonetheless highly relevant to the patterning of article distribution. Therefore one of the dimensions of any model of English article use must be a grammatical number dimension along which the singular, plural, and uncountable semantic distinctions are arranged. In the indefinite pattern, nouns that express prototypical singular semantics will be marked with *a*, and nouns that express prototypical plural or uncountable semantics will be marked with Ø. In the definite pattern, all types of number are marked with *the*. Table 2 lists these patterns along with examples.

Table 2. *Types of nominal Number constructions.*

Indefinite	**Definite**	**Number**	**Examples**[9]
a	the	Singular (SG)	a thought, a beer, a book, a word, a sentence, a pencil, a keyboard, a tree, a leaf, a root, a car, a plane, a party, a year
Ø	the	Plural (PL)	Ø thoughts, Ø beers, Ø books, Ø words, Ø sentences, Ø pencils, Ø keyboards, Ø trees, Ø leafs, Ø roots, Ø cars, Ø planes, Ø parties, Ø years
Ø	the	Uncountable (UC)	Ø thought, Ø beer, Ø rice, Ø salt, Ø sunshine, Ø weather, Ø air, Ø advice, Ø neglect, Ø storage, Ø access, Ø attire, Ø bewilderment

Even though these three types of Number are the bare minimum required by English morphosyntax, it is not always easy to predict whether a given noun should be categorized as countable (and hence able to be marked with the indefinite article *a* in the singular and able to be pluralized with /-*s*/), vs. uncountable (and hence unable to take *a* or /-*s*/), vs. able to be both countable and uncountable. In fact, when all types of English nouns, including proper nouns, are taken into account, the situation is even more complicated. The data in Table 3 illustrate this complexity by showing that, based on the ways a given noun lexeme is typically construed, it can be used in one of the following combinations: 1. singular, plural, and uncountable, 2. only singular and plural, 3. only

[9] Note that *the* can occur with all the examples in Table 2 as well.

singular and uncountable but with historical plural *-s* ending, 4. only plural (*plurale tantum*), 5. only uncountable, 6. only singular but with historical plural *-s* ending, 7. only singular with *the*, 8. only singular with Ø, 9. only singular with Ø or plural with *the*, 10. only plural with *the*.[10]

Patterns 7-10 in Table 3 all stem from the article usage patterns associated with proper names in English, and therefore as such, they can be considered as distinct from the patterns found with common nouns (i.e., patterns 1-6). The complex patterns found with proper names have been discussed above in section 2.2.1.1 on proper names. Although the common noun patterns in 1-6 in Table 3 are all well known, treatments of English grammatical number usually provide only the traditional three-way "singular vs. plural vs. uncountable" distinctions, a categorization system that is sufficient for describing the morphological marking available for any given common noun that appears in discourse (i.e., the possibility of the presence or absence of the indefinite article and the pluralizing /*-s*/[11]).

[10] This table does not include nouns borrowed into English from other languages that have kept the singular-plural morphology from their original languages, such as *graffito-graffiti, radius-radii, criterion-criteria, etc.* These nouns, while complicating the English number system, follow the same article usage patterns that nouns with regular *-s* plurals do.

[11] Or, again, other pluralizing strategies in the case of some foreign borrowings such as singular-plural forms *graffito-graffiti, radius-radii, criterion-criteria, etc.*

Table 3. Data illustrating the complexity of the English grammatical number system. (Examples are mine. Unacceptable forms are starred; acceptable forms are bolded.)[a]

	Singular	Plural	Uncountable
1. Singular, plural, & uncountable	*Ø cake **a cake** **the cake** **(A cake *is* perfect.)**	**Ø cake-s** **the cakes** **(Ø Cakes *are* perfect.)**	**Ø cake** **the cake** **(Ø Cake *is* perfect)**
2. Only singular and plural	*Ø book **a book** **the book**	**Ø book-s** **the book-s**	*Ø book ?the book
3. Only singular and uncountable but with historical plural *-s* ending	*Ø poetics **a poetics** **the poetics** **(A new poetics *is* needed.)**	*Ø poetics *the poetics (*Ø New poetics *are* needed.)	**Ø poetics** **the poetics** **(Ø Poetics *is* the study of…)**
4. Only plural (*plurale tantum*)	*Ø scissor *a scissor *the scissor (*Ø/A scissor *is* dangerous.)	**Ø scissor-s** **the scissor-s** **(Ø Scissors *are* dangerous.)**	*Ø scissor *the scissor (*Ø Scissor *is* dangerous.)
5. Only uncountable	*Ø butter *a butter *the butter (*A butter *is* fantastic.)	*Ø butter-s *the butter-s (*Ø butters *are* fantastic.)	**Ø butter** **the butter** **(Ø Butter *is* fantastic.)**
6. Only uncountable but with historical plural *-s* ending	*Ø checkers[b] *a checkers *the checkers (*A checkers *is* fun.)	*Ø checkers *the checkers (*Ø Checkers *are* fun.)	**Ø checkers** **the checkers** **(Ø Checkers *is* fun.)**
7. Only singular with *the*	*Ø Hague *a Hague **the Hague**	*Ø Hague-s *the Hague-s	*Ø Hague *the Hague

Table 3, continued.

	Singular	Plural	Uncountable
8. Only singular with Ø	**Ø John** ?a John ?the John	*Ø John-s ?the John-s	*Ø John *the John
9. Only singular with Ø or plural with *the*	Ø 20 (age, year) *a 20 *the 20 **Ø America** ?an America *the America	*Ø 20s **the 20s** ?Ø America-s **the America-s**	*Ø 20 *the 20 *Ø America *the America
10. Only plural with *the*	*Ø Netherland *a Netherland *the Netherland	*Ø Netherland-s **the Netherland-s**	*Ø Netherland *the Netherland

[a]Representative sentence examples are included with some of the nouns in Table 3 to further illustrate their grammatical number based on verb agreement. Where no examples are provided, the reader can follow the exemplified pattern to supply other similar sentence examples.

[b]This refers to the game *checkers* (not an individual checker piece, which follows the "only singular and plural" pattern). The game *checkers*, while keeping its historical plural *-s* ending is grammatically singular, as seen in the verb agreement in the sentence *Checkers <u>is</u> a fun game* (as opposed to **Checkers <u>are</u> a fun game*).

2.4. Discourse Modes

Du Bois (1980:227ff) argues that "discourse mode" is a salient category for explaining patterns of definite and indefinite article use in natural language English. He lists two main modes, which he calls *narrative* and *descriptive*, each of which is further divided according to pragmatic functions that are reflected in syntactic patterns of article use. All in all, there are at least five discourse modes suggested by Du Bois: 1.) *immediacy narrative mode* (so named by me, not by Du Bois), 2.) *(normal) narrative mode*, 3.) *(normal) descriptive mode*, 4.) *defining descriptive mode*, and 5.) *deferred descriptive mode*. These modes can change from clause to clause within a discourse.

2.4.1. Normal narrative mode

Mode 2, the *normal narrative discourse mode*, is the one that serves as the "standard mode" in the sense that it describes the pattern of article use that most researchers seek to describe. In other words, it is the one that describes article use according to all the principles and patterns that we have been discussing to this point in the above long section on Referentiality (section 2.2). It is likely the most common mode – evidenced by the fact that it appears to explain the article patterns researchers notice and investigate most – so we will start with it. Du Bois describes the *normal narrative discourse mode* as the mode which speakers use when they wish to advance the story line. This mode evinces the typical article use pattern as it has been traditionally understood: referential identifiable referents are marked as definite while referential non-identifiable referents and most non-referential mentions are marked as indefinite (with the exception of some idioms and some noun incorporation constructions). Since this mode is exactly equivalent to the description of referentiality and identifiability that researchers typically offer (the one that has been detailed above in section 2.2), one way of thinking about it might be to think of it as the "normal" pattern of article use when a narrative is being carried forward, progressing from one event to another, in English discourse. Examples of this normal narrative mode can be seen in the first and last clauses of (56):

(56) he comes across another ... bicyclist ... bicyclist;
it's a young woman,
... and ... for some reason she catches his attention (Du Bois 1980:227)

2.4.2. Normal descriptive mode

Du Bois points out that while the first and last clauses in (56) are in the normal narrative mode, the second clause is in the normal *descriptive* mode. Speakers use the *normal descriptive mode* when they do not wish to advance the story line but to pause and comment on some aspect of it. It differs from the *normal narrative mode* in the way that articles are used with referents that have frame-activated lexical grounding. Whereas these types of referents are marked with the definite pattern in the *normal narrative mode,* in the *normal descriptive mode* they follow the indefinite pattern. Du Bois suggests that this phenomenon may be at least partly due to the fact that speakers normally shift to the *descriptive mode* from the *narrative mode* when they feel the need to mention something that can't be assumed. In so doing, a mention of the new referent must be marked as new or noteworthy by giving it an indefinite marking. In (56), *a young woman* is newly mentioned and is a non-identifiable referent since it does not receive identifiability from a frame, shared speech situation, etc., and as such it receives an indefinite marking. Although the clause containing *a young woman* is in the normal descriptive mode, it so happens that even if the clause were in the normal *narrative* mode, it would still receive an indefinite marking, since both modes code new references with the indefinite pattern.

However, the article usage pattern between the modes is seen to be different when *frame* activation is involved. In the normal narrative mode, referents with identifiability from frame-based grounding receive a definite marking. But in the normal descriptive mode, such referents receive an indefinite marking. For an example of this difference, consider items (57) and (58), below. In the *normal narrative mode* in (57), the boy who is wearing a hat has already been introduced, but no mention of his *hat* has yet been made.

However, the mere existence of the boy in the narrative has provided sufficient enough context to trigger a frame-activation that allows the first mention of *hat* to be marked as definite with the definite personal pronoun *his*.

(57) and he rides off, and passes um… a girl… on the road… and-- his hat [1st mention] falls off (Du Bois, 1980:243)

However, (58) provides an example of a *normal descriptive mode* use of the indefinite pattern marking an initial mention of *shirt*. The previous mention of *man* is not sufficient in this context to allow a frame-activated definite marking on this particular *shirt*, possibly because the speaker wants to specify information that cannot be assumed by the listener (such as the color of the shirt), and therefore the speaker switches to *descriptive mode* and uses the indefinite article normally associate with that mode.

(58) A--nd you see a middle-aged .. u--m .. Chicano man,… who's wearing .. a-- .. navy blue shirt (Du Bois, 1980:243)

Du Bois (1980:227-228) also notes that the appearance of certain words in a clause appears to be correlated with both the *descriptive* and *narrative* modes. Verbs such as *be, have, look like, wear, have on, be dressed in,* as well as the preposition *with* ("in the sense of 'having an attribute'") generally correspond with the *descriptive mode*. Likewise, concrete verbs such as *pick, fall,* and *scatter* and more abstract verbs such as *fulfill, grow, ponder* generally correspond with the *narrative mode*. However, these correlations do not perfectly predict article use, since examples like (59) and (61) can be cited in which the typical definiteness or indefiniteness associated with these verbs do not apply.

(59) He looks like the guy we saw last week at the orchard.

In (59), although the verbal phrase *looks like* may represent the descriptive mode, *the guy we saw last week at the orchard* is marked as definite, even though many other instances of arguments following that same verb might be categorizing ones such as (60):

(60) he looks like a uh .. Chicano American, (Du Bois 1980:227)

Furthermore, in (61) with the mention of the non-referential NP *no pear,* we can see that the verb *fall* does not always predict definite *narrative mode* referential NPs, even if it generally might:

(61) We sat under the trees and watched pears fall all day, but at that moment, no pear fell.

Du Bois for his part says that these lexical correspondences are only general and that discourse modes have "considerable" rather than absolute influence on the use of particular articles.

2.4.3. Immediacy narrative mode

The *immediacy narrative discourse mode* (mode 1) may be used to instill a sense of "immediacy" or vividness into a narrative when a new referent is first mentioned. It differs from the *normal narrative* pattern in the way that articles are used with non-identifiable referents. In the *normal narrative* pattern, new references are marked as indefinite, but in the *immediacy narrative pattern*, new references are marked as definite. Example (62) illustrates the use of the *immediacy narrative* mode. Normally, the first mention of *man* would be indefinite (*a man*), but here it is definite (*the man*), possibly to give a sense of immediacy in the introduction of this new character.

(62) the first… thing I noticed.. was.. the sound of the man [1st mention] picking… pears. …And of course there was a… a man [2nd mention] there standing on a ladder in a pear tree,

2.4.4. Defining descriptive mode

The *defining descriptive mode* (mode 4) and the *deferred descriptive discourse mode* (mode 5) share the same article distribution pattern: they both use the indefinite pattern (*a*, in the singular) to mark referents that have a current discourse grounding (i.e., "old" referents, as described above in section 2.2.1.5). This differs from the definite pattern (*the*) seen in the other discourse modes.

The *defining descriptive mode* (mode 4) is used when a definition is required in a discourse. When a word is being defined after it has already been mentioned, the subsequent defining words obviously refer to that same word and concept. In other discourse modes, subsequent references have a definite marking. But in the *defining descriptive mode*, the defining noun phrases that refer to the same concept or referent as a previous noun phrase receive an indefinite marking instead. In item (63), we see two examples of this pattern. First, *Kutchin* is mentioned, but when it is defined, the subsequent phrase *an Athabaskan language* which refers to it receives an indefinite marking. Second, *phonemes* is mentioned, but when it is defined, the subsequent phrase *distinct consonantal elements* which refers to it also receives an indefinite marking.

(63) Ø Kutchin [1st mention], an Athabaskan language of Alaska [2nd mention], possesses no less than 55 consonantal "phonemes" [1st mention],

Ø distinct consonantal elements of the total phonetic pattern [2nd mention].

(Du Bois 1980:231, quoting Sapir, 1929:140)

2.4.5. Deferred descriptive mode

Finally, the *deferred descriptive mode* (mode 5) has the same pattern of article use as the *defining descriptive mode*, but the pragmatic function is somewhat different. The *deferred descriptive mode* is used when the momentum of the narrative causes the speaker to make an initial mention of something without giving an adequate explanation of it. The explanation is deferred for a short time, perhaps only for a phrase or two; however, when the speaker finally gives attention to it in order to "officially" introduce it in the discourse, it receives the same indefinite marking that is normally given for initial mentions, even though it technically has identifiability grounded in previous mention in the current discourse (i.e., it's an "old referent"). An example is seen in (64) (repeated here from above), where even though the second mention of *baskets* has current discourse grounding, it is not coded as definite but as indefinite.

(64) [The Pear Man] picks pears,… puts them in.. his apron,… climbs down the ladder,… and empties the pears.. into… big.. Ø baskets [1st mention]. …tsk There's like Ø three baskets [2nd mention] sitting there

(Du Bois, 1980:229)

Item (65) not only demonstrates the use of the *immediacy narrative mode* (described above) but also gives another example of the *deferred descriptive mode*, since the second mention of *man* gives the deferred explanation of who he is.

(65) the first… thing I noticed.. was.. the sound of the man [1st mention] picking… pears. …And of course there was a… a man [2nd mention] there standing on a ladder in a pear tree,

Table 4 lists all of Du Bois' (1980) five discourse modes and shows how the article distribution patterns differ with each mode. Although many identifiability types were described above in section 2.2.1, only three of those types differ in terms of the discourse modes (at least as far as I now know), so only those three identifiability types are listed here for comparison purposes (frame, current discourse [i.e., old reference], and new reference).

Table 4. *Du Bois' (1980) five Discourse Modes and their associated article distribution patterns. (Note: SG = singular; PL= plural; UC = uncountable.)*

	Discourse Modes				
	Immediacy Discourse Mode	**Normal Discourse Mode**	**Normal Descriptive Discourse Mode**	**Deferred Descriptive Discourse Mode**	**Defining Discourse Mode**
Identifiability Type	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**
Identifiable - Frame	the, the, the	the, the, the	a, Ø, Ø	the, the, the	the, the, the
Identifiable - Current discourse ("Old" referenc)	the, the, the	the, the, the	the, the, the	a, Ø, Ø	a, Ø, Ø
Non-identifiable - New reference	the, the, the	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø

2.4.6. Headline mode

Du Bois mentions that discourse modes other than the ones he delineated will likely need to be posited. One such mode is a *headline* discourse mode that is used for written headlines and signs, since the article use patterns are often quite distinct for these types of discourse. Consider the contrast between the headline in (66) and the news article content that follows the headline in (67):

(66) Palestinian prisoner ends 102-day hunger strike (Al-Akhbar, July 23, 2012)

(67) A Palestinian prisoner has ended his 102-day hunger strike after Israel agreed to a deal to release him early. (Al-Akhbar, July 23, 2012)

Where in normal narrative discourse, we might expect the inclusion of the indefinite article (*A Palestinian prisoner*) upon the first mention of the referent (as indeed we find in item (67)), the headline discourse mode does not supply it, leaving only the bare NP *Ø Palestinian prisoner*. Similarly, article usage patterns in signs sometimes differ from patterns in other types of discourse, as in (68), where a definite article might be expected (in keeping with the standard practice in narrative discourse, for example) but is sometimes not provided:

(68) Keep off Ø grass.

Because of their unique patterns of article use and their unique pragmatic functions, we might consider headlines and signs as comprising an additional discourse mode (or perhaps even two modes, depending on their degree of similarity).

2.5. Semantic Maps

It is not a trivial task to display in a simple and clear way all of the relatively large number of semantic and pragmatic features that are implicated in the operation of a complex system like the English article system. Semantic maps, however, offer one way to do so. In fact, it is this sort of complicated arrangement of features that semantic maps were originally created to display. The first clearly described use of semantic maps for broad linguistic application is credited to Anderson (1982, 1986), who employed them to investigate and diagram complex packages of ideas such as the perfect and evidentials both within and across languages. The way subsequent researchers have conceptualized and utilized semantic maps has changed relatively little since Anderson first delineated their use, a fact which gives evidence to their intuitive and practical utility.

Anderson received his inspiration for semantic maps from the work on cross-cultural color categorization done by Berlin and Kay (1969), who, by mapping the perceived color boundaries of color names on a chart whose two axes were wavelength (i.e., hue) and brightness, were able to visually display and compare how different languages organized the naming of similar colors. When colors terms are mapped in this fashion, the scopes or areas of perceptual color that they describe are displayed on a "map" showing a range of hue and brightness, where similar perceptual hues/brightness are near one other and one color fades gradually into another. In similar fashion, when adopting this method for other semantic domains, Anderson arranges concepts with similar meanings and functions adjacent to one other on one axis or dimension of a map. One or more additional dimensions are displayed on another axis. Linguistic expressions that have related meanings or functions generally end up being distributed next to each

other within the map, thus showing their close semantic or pragmatic relationship. However, sometimes this kind of distribution or arrangement proves to be difficult or impossible when constructing a map, depending on the particulars of each linguistic structure that is being included in the map.

2.5.1. Conceptual space

Following the framework provided by Berlin and Kay (1969) and Anderson (1982, 1986), the structure of the linguistic models offered by linguists such as Pederson (1991a, 1991b), Croft (2001), Haspelmath (2003), and De Haan (2004, 2010) also consists of conceptual dimensions along which functional or semantic categories are labeled. Where two of these dimensions intersect, the resulting plane constitutes a "conceptual space" over which forms or constructions are placed according to how they correspond to or express the functional and semantic categories along the dimensions. A conceptual space therefore provides a realm in which a number of semantic or pragmatic functions could potentially be distributed. Since the specific expression and range of these functions can vary across languages, any given semantic map of constructions for a particular language will likely differ from those of other languages. However, since the underlying conceptual space is claimed to be universal, it should remain the same (Croft, 2001). For this reason, the goal of creating (or more properly, representing or modeling) a conceptual space onto which any number of particular semantic maps can be placed is to lay out dimensions (domains) and attributes (functions) that are presumably universal across speakers of human languages (Croft, 2001; Haspelmath, 2003).

For a schematic example of a two dimensional conceptual space, consider Figure 4. In this model (Croft, 2001:92), the complex semantic categories of *objects, properties,* and *actions* are aligned on the vertical dimension of the chart, and the functional speech-act categories of *making reference, making modification,* and *doing the job of predication* are arranged along the horizontal dimension. The possible functions made possible by the two dimensions are listed in the conceptual space defined by the dimensions.

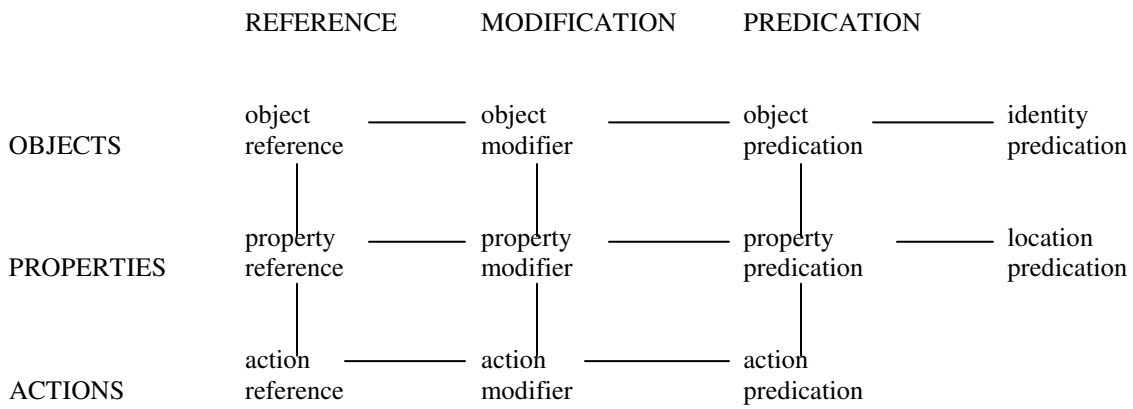

Figure 4. Conceptual space relevant to parts of speech (Croft, 2001:92).

2.5.2. Semantic maps

Although the term "semantic map" might at first seem to imply that it is a tool used only for dealing with "semantics," in reality it can be used more broadly to analyze the expression of any semantic/pragmatic function. A semantic map is a diagram that

illustrates exactly how the structures of a particular language (or set of languages) uniquely express the possibilities represented in the overall conceptual space defined by any given set of functions. For example, a language may have two morphemes that each express certain non-overlapping areas of the conceptual space, or a (different) language might use three morphemes that each express non-overlapping areas of the same space.

To further illustrate this point, we could imagine two different languages – Languages A and B – that express the functions in Figure 4 differently (see Figure 5, below). Language A (represented by ovals in Figure 5) might have one affix used to mark reference to properties and a different affix to mark predication of objects and properties, while Language B (represented by rectangles) might have three affixes: one for marking modification and predication of objects; one for marking reference, modification, and predication of properties; and one for marking reference and modification of actions. Some possible functions in the conceptual space (represented by their lack of encapsulating shapes) might be expressed with unmarked lexemes or with more complex multi-word or idiomatic constructions.

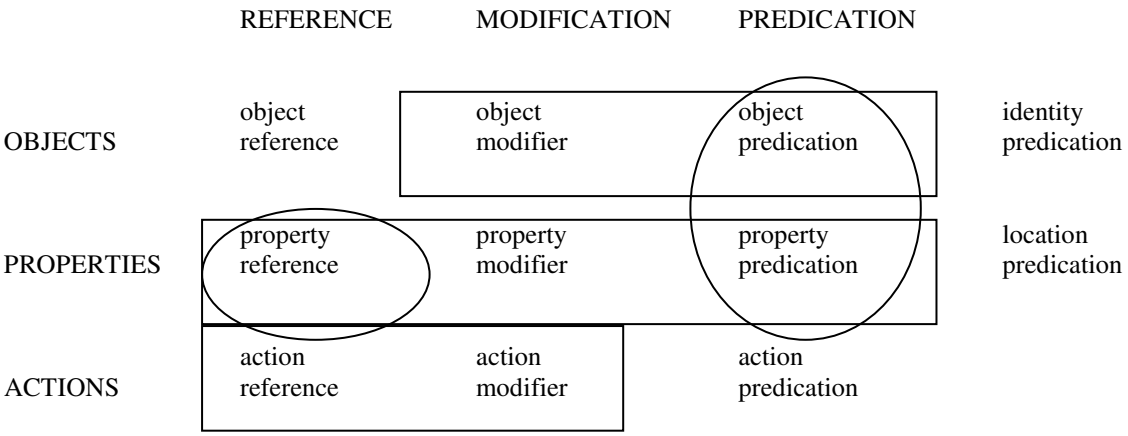

Figure 5. Semantic map illustrating how two hypothetical languages (represented by ovals vs. rectangles) might differently express the possibilities of the underlying conceptual space (adapted from Croft, 2001:91).

Semantic maps can also provide a way to compare different constructions in a single language in terms of how they overlap semantically among the various functional dimensions of the conceptual space model. For an example of this useful feature, consider Figure 6, in which Croft (2001) schematizes how certain Japanese constructions coincide and overlap in their ability to mark various Japanese lexemes (only the English glosses for the words are listed on the map). Although there are many details represented in this map, the important thing to notice for our purposes here is that it offers a way to quickly conceptualize the relationships between the various constructions listed on it and to visualize that the seven words listed must be divided into at least six lexical classes or

construction types, based on the syntactic distribution evidence of constructions distributed over the conceptual space defined by just two dimensions.[12]

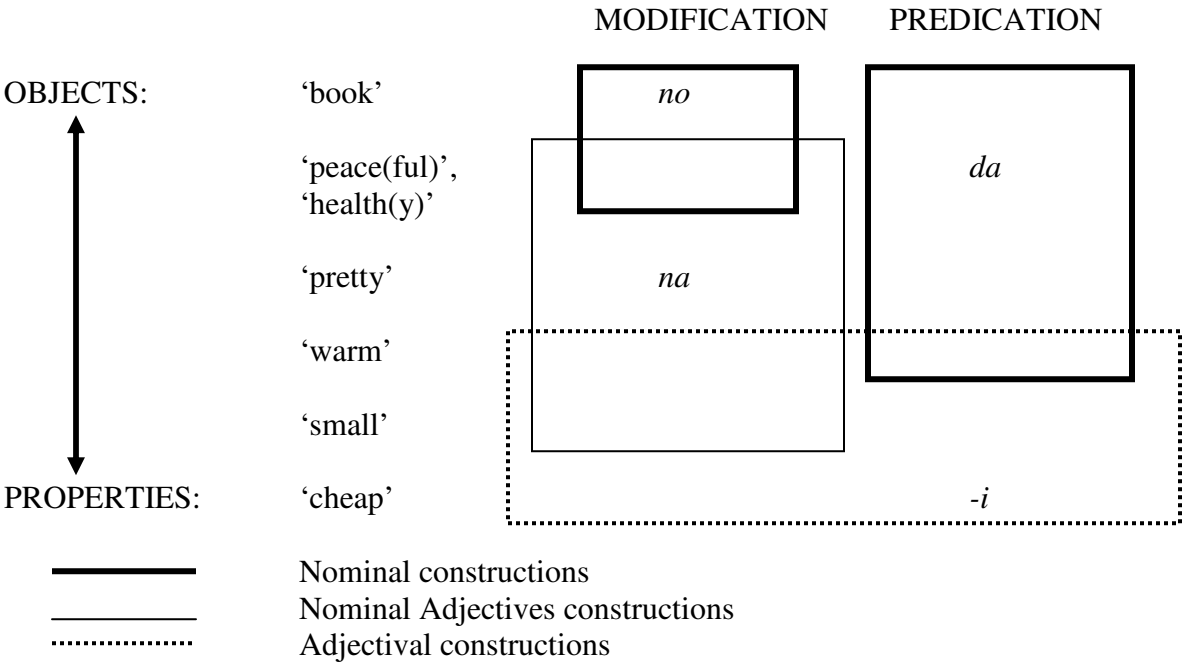

Figure 6. Semantic map for the Japanese Nominal, Nominal Adjective, and Adjective Constructions (Croft, 2001:95).

It should be noted that drawing a semantic map with the right arrangement of features (or "attributes") and axes (or "dimensions") is a matter of intuition, deduction, and experimentation. An initial arrangement can be guessed at, but as more data is

[12] The six lexical classes are based on the modification and predication morphemes combinations allowed with each of the lexemes listed on vertical axis: 1) *book* (*no, da*); 2) *peace(ful)/health(y)* (*no, na, da*); 3) *pretty* (*na, da*); 4) *warm* (*na, -i, da*); 5) *small* (*-i, da*); and 6) *cheap* (*-i*).

considered, the ordering of the features may need to be rearranged (to the extent that it is possible) so that 1) similar meanings and functions can be kept together along each dimension and 2) the places on the map where a certain linguistic form appears is a contiguous area within the conceptual space. However, this is not always possible, since neither the diachronic pathways of semantic and functional change nor the synchronic meanings and functions are always entirely transparent (Anderson, 1986; Croft, 2001). A perfect map might account for all of these historical and cognitive characteristics by accurately supplying all of the relevant features and dimensions represented in the minds of the speakers of a language throughout time. Such a map might have many dimensions, and perhaps the dimensions and dimensional attributes of the map might each be weighted differently and have differing connection strengths between one another (based on frequency distributions of the functions and constructions represented, for example), and therefore it might be difficult to represent visually, even with advanced computer animations. Ever more complete and complex maps may be possible in the future as methods for examining and measuring semantic and pragmatic representations and processes in the human mind become more advanced, but even now experimental and historical research tools can in many cases provide data sufficient to propose thorough, multidimensional maps.

2.5.3. Implicational hierarchies

The structure of implicational hierarchies (e.g., the Relativization Noun Phrase Accessibility Hierarchy of Keenan and Comrie, 1977, or the Givenness Hierarchy of Gundel et al., 1993) makes them well suited for use as dimensions for a conceptual space

since including a construction on a map at a lower level of such a hierarchical dimension predicts that same construction would also be used at the higher levels, thus preserving contiguous areas of the map being filled without gaps (a desideratum when creating semantic maps). In fact, because of this aspect of their structure, accessibility hierarchies in essence already represent one dimension of a semantic map.

2.5.4. Functional dimensions and attributes

The above sections (e.g., Figure 4 - Figure 6) provide some examples of the types of functions that might be included as attributes on the dimensions of maps of conceptual spaces. However, of course, any number of other functions might be included in maps. The important thing is to try to posit cognitively or functionally real dimensions (e.g., *referentiality*) with similarly real attributes that are arranged in an order that reflects their functional relationships (i.e., with the most related functions placed nearer than less related functions).

A number of semantic maps have been published in the literature. The three examples below provide a sampling of some types of dimensions and attributes that can be represented in maps of conceptual space. Anderson (1982), in attempting to explicate cross-linguistic expression of grammatical perfect tense-aspect-mode-evidential constructions, plots a conceptual space with a tense/aspect dimension consisting of attributes with such (rather non-intuitive) names as “current relevance result” and “current relevance anterior,” and a voice/mood dimension consisting of attributes such as “current relevance experience” and “current relevance new situation.” Pederson (1991a) posits a map of semantic space for “change of event” constructions that has a

"transitivity" dimension with gradual range of attributes from high to low transitivity and a "subject responsibility" dimension with a range of attributes from high to low responsibility. De Haan (2010) offers maps trying to account for the distribution of terms expressing modality and evidentiality; his maps have an epistemic modality dimension consisting of "strong" and "evaluative" attributes and an evidentiality dimension consisting of "assertive" and "predictive" attributes.

In this study, a semantic map will be proposed that describes the use of English articles. This map will include three dimensions: a referentiality dimension (with a number of referential/identifiability and non-referential functional attributes), a discourse mode dimension (with a number of functional discourse types as attributes), and a grammatical number dimension (with singular, plural, and uncountable attributes). The next chapter will test various aspects of the most complex two dimensions – referentiality and discourse mode.

CHAPTER III

PROPOSING AND COMPARING VARIOUS SEMANTIC MAP MODELS

3.1. Introduction

3.1.1. Overview

This chapter contributes to the development of a full description of English article usage by testing various semantic maps. Four competing semantic maps are proposed that describe many of the form-function mappings of English article use, organizing them along two cognitive domains or dimensions[13] – Referentiality Type (how a noun is accessed in the memory of the interlocutors) and Discourse Mode (why the speaker mentions a noun in a discourse, such as to narrate, describe, clarify, etc.). These four semantic maps include an Initial vs. Subsequent Mention Model (an overly simplistic model), a No Discourse Modes (Referentiality Only) Model, a 2 Discourse Modes Model, and a 3 Discourse Modes Model. These models are tested and compared with an approach that uses data collected specifically for this purpose from an elicited conversational discourse. Analysis shows that the 2 Discourse Modes Model offers the most accurate performance, followed closely by the No Discourse Modes (Referentiality

[13] The Number dimension that must be included in a full model of English article use will not be tested here since accounting for article distribution patterns that vary by grammatical number is a trivial task.

Only) Model, and that both of these models perform on par with human predictors. These findings support the importance of including both a relatively complex referentiality system and a discourse modes system in any model of English article usage. The methodological approach employed here, with its provisional but incomplete semantic-map-based model, allows for the development and testing of additional or alternative semantic map features that might expand and improve predictive ability regarding article use.

3.1.2. Extracting testable models from the literature on English articles

The models of English Article Use proposed and tested here will incorporate variations of two pragmatic dimensions that are implicated in syntactic article distribution patterns in English: a Referentiality dimension and a Discourse Mode dimension. A third dimension that must be included in a full model of English article use – Grammatical Number – will not be tested because its operation is relatively straightforward and uncontroversial.

A full model of English article use will be proposed in Chapter IV. That model will be based partly on the findings of the experiment in this chapter, but it will also include dimensional attributes that are not included in the four models tested here. The four models in this current chapter are designed to include and test only certain attributes of the overall comprehensive model. Not all attributes in the full model could be tested in the paradigm employed here, and there are a number of reasons why only certain attributes were selected for testing.

The first reason why not all attributes were tested is that some attributes are well supported in the other studies, while others are not. Those that are not well established require the sort of careful evaluation that experimental performance testing can provide. In our model of English article use, the Discourse Mode dimension is the most unique and the least well explored in the literature. It is based on ideas proposed in Du Bois (1980) but which have received little attention since then, and therefore it is necessary to test whether or not the inclusion of a Discourse Mode dimension will improve the performance of a predictive model of article use. Our experimental paradigm will allow us to explore that question and will further allow us to compare various versions of models that include Discourse Mode to see which version performs the best.

Some of the attributes on our full model are not controversial, and therefore it is less urgent to test these attributes at this time. These relatively uncontroversial attributes include especially the attributes on the Grammatical Number dimension (the Singular, Plural, and Uncountable attributes), and so in order to simplify the experimental design by reducing the number of factors, only Singular NPs will be examined in the performance testing. Other well established attributes include the Non-identifiable "New Reference" attribute and the Current Discourse (also known as "Old Reference") attribute on the Referentiality dimension: all accounts in the literature note these two functions as being key functions of English articles. However, since they participate as only two attributes in a larger set of many Referentiality attributes, they must remain in our tested models for two reasons: 1) so that we can compare the effectiveness of the other attributes in the Referentiality dimension, and, more importantly 2) so we can compare

the performance of various Discourse Mode models across these many Referentiality attributes.

A second reason why not all attributes were tested involves the types of predictions that can be made about them and the availability of data that can test those predictions. Since our primary goal is to test competing Discourse Mode models, it is important to select attributes that make clear, testable predictions about article use across the proposed discourse modes. Furthermore, we must select attributes for which we can likely collect sufficient data for making comparisons within the limitations of our experimental method. The methodological paradigm we employed was chosen specifically to foster the elicitation of sufficient quantities of data in the following Discourse Modes that we wished to test: the Normal Narrative mode (also called simply the "Normal Mode" in one of our models), the Normal Descriptive Mode, the Deferred Descriptive Mode, and the Reintroducing Mode. In addition, it was determined that the following Referentiality attributes conformed to the "testable predictions" and "sufficient data" criteria: Shared Lexical Understanding, Share Speech Situation, Frame, Current Discourse, New Reference, and Categorization. Therefore, the above Discourse Mode and Referentiality attributes were selected for inclusion in our tested models.

A final reason to test the performance of certain attributes is because of their possible utility in assisting learners of English in overcoming problems in acquiring usage of articles. Since the attributes listed in the previous paragraph account for the majority of types and tokens of predictable usages of English articles, it is likely that mastering these functions would contribute to English learners' success in mastering articles, and testing them to increase the confidence of our understanding of their

operation is therefore a worthwhile practical goal. Certain other attributes are less frequent (Secondary Predicates, Finite Verb "Noun Incorporation"), less predictable and therefore requiring rote memorization (Idioms), highly predictable and simple in their form since they consistently use the Ø article pattern (Secondary predicates, Vocatives, Compound Modification [i.e., Adjectives], and Full Noun Incorporation), or partially predictable but under their own different, complex semantic map (Proper Names). For those reasons, including those attributes on our tested models would not give us as much useful information from a theoretical point of view, and it would for that reason also not provide much additional help to English language learners. Any attributes that are not included in our tested models but which must be included in our full model will, however, be discussed in Chapter IV.

Four semantic-map-based models will be proposed for testing in this chapter. These four semantic maps include the following: an Initial vs. Subsequent Mention Model (an overly simplistic model), a No Discourse Modes (Referentiality Only) Model, a 2 Discourse Modes Model, and a 3 Discourse Modes Model. These models will be explained below.

3.1.2.1. Models based on Referentiality Types

3.1.2.1.1. The proposed Initial vs. Subsequent Mention Model.

The first testable model proposed here will be a simple one based on only two Referentiality types: Initial vs. Subsequent Mention Model (these types are also known as Given [or Old] vs. New reference, or Current Discourse Grounding vs. New Reference). A detailed description of these two types can be found in Chapter II.

Although no one proposes that these two considerations are all that is required to predict article use, they are commonly the only factors described in terms of referentiality in pedagogical grammar books (e.g., Azar, 1999; Murphy, 2007), so they warrant testing as a stand-alone model, even if for no other reason than as a metric by which to measure the performance of more complex models.

3.1.2.1.2. The proposed No Discourse Modes (Referentiality Only) Model.

A second testable model will incorporate a version of the above two Referentiality types (Old vs. New Reference) but will somewhat change their definitions and implementation in order to also include a number of other Referentiality types described in Chapter II. The syntactic evidence suggests that most of the Referential (i.e., not Non-referential) types discussed in Chapter II are motivated by distinct distribution patterns of the English articles within the Referentiality dimension and/or across Discourse Modes. (That is to say, different functions can produce different article patterns for those types. The evidence for this is presented later in the discussion on Discourse Modes.) Article distribution patterns in the available data do not support the division between Shared Lexical Understanding (type 2) and Shared Speech Situation (type 3). However, because these two categories do seem pragmatically distinct and because of the typological considerations presented by Givón (2001), I will include them provisionally as separate attributes in the Referentiality dimension of the models.

Two of the Referential types discussed in Chapter II will be excluded from performance testing here, even though they must be included in a full model of English article use. First, Proper Names will be excluded because the article distribution patterns

that they exhibit are quite different from those of common nouns, and accounting for them will require a separate semantic map and a separate testing methodology which must at this point remain a goal of future research (a provisional semantic map for their use will, however, be proposed in Chapter IV). In addition to Proper Names, Non-specific mentions will be excluded from testing, for two reasons. First, since Non-specific mentions are relatively rare (at least as defined in Chapter II), it was thought to be unlikely that enough could be collected in the current experiment paradigm. Gathering experimental data with sufficient Non-specific mentions would likely require a special methodological design that is incompatible with the design used to collect Discourse Modes data in this experiment. Second, even if some Non-specific data were to be fortuitously obtained via the current elicitation methodology, no unique predictions about article use patterns are made for them, since their article use pattern are identical to those with New Reference identifiability status. Again, separating those two Referentiality types in a way that would yield meaningful results will likely require a dedicated experimental design.

The various Referential Identifiability types proposed for inclusion in the testable models are summarized in Table 5, along with one typical distribution pattern of the articles and examples of each type. (Other distribution patterns will be discussed below under the section on Discourse Modes.)

Table 5. *Types of Identifiable and Non-identifiable Referential uses that will be included in the testable models.*

SG	**PL**	**UC**	**Type of Cognitive Reference Activation (i.e., Type of Referential Grounding)**	**Examples**
the	the	the	Shared Lexical Understanding	The sun came out. (Givón, 2005)
the	the	the	Shared Speech Situation	Hand me the hammer.
the	the	the	Frame activation	She walked into a restaurant and asked the waiter for the menu. (Givón, 2005)
the	the	the	Current Discourse	the a boy [1st mention] comes by,… on a bicycle; the man is the tree,… and the boy [2nd mention] gets off the bicycle (Du Bois, 1980:206)
a	Ø	Ø	No current grounding (i.e., "New reference" that is non-identifiable to listener)	the a boy [1st mention] comes by,… on a bicycle; the man is the tree,… and the boy [2nd mention] gets off the bicycle (Du Bois, 1980:206)

We now turn to a discussion of Non-referential uses. As mentioned in the literature review in Chapter II, in positing the categories or attributes for inclusion in the Non-referential component of the larger Referentiality dimension of a full semantic map of English articles, I proposed to exclude one of Du Bois' descriptive categories, to collapse other categories together, and to split one category into two, in the end leaving

five distinct uses of Non-referential NPs: *categorization*, *secondary predicates, vocatives*, *full (object-verb) noun incorporation,* and *finite verb (verb-object) "noun incorporation"*.

However, because at this point we have no clear predictions for article use in Secondary Predicates, for the Finite Verb (verb-object) "Noun Incorporation," or for the Idioms types, they will be excluded from performance testing in our proposed models. Furthermore, Vocatives, and Full (object-verb) Noun Incorporation all consistently and uncontroversially follow the Ø article pattern and are therefore unimportant to our goal of testing Discourse Modes, so they, too, will be excluded from the performance testing. In addition to these concerns, it was thought that all of these Non-referential types were likely to be only rarely expressed in our elicited data since they are generally rare to begin with and since our experimental design does not specifically attempt to generate their occurrence, so even if we were to attempted to include them in our performance tests, our efforts might be stymied by scarcity of data. (All of these types will, however, be coded and counted in the experimental data for sake of comprehensiveness.) This leaves us with just the non-referential Categorization type for inclusion in our tested models. It will remain since it is the only type that does not predictably follow the Ø article pattern. It also is the most frequent Non-referential type, so it was thought that enough Categorization tokens would likely be collected to warrant its inclusion. Our proposed models' full Referentiality dimension, including both Referential/Identifiability and Non-referential attributes is expressed in Table 6. This matrix of Referentiality/Identifiability types represents in full our proposed No Discourse Modes Model (which could also be described as a "Referentiality Only Model," i.e., a model without reference to more than one "Discourse Mode"). This model will be tested

against a text corpus in the Experiment section. (The 2 Discourse Modes Model and 3 Discourse Modes Model, described below, will incorporate this same scale of Referentiality/Identifiability types, but will add to it another dimension called Discourse Mode which will vary the predictions about the article distribution patterns for each Referentiality/Identifiability type as the Discourse Mode varies.)

Table 6. *The proposed Referentiality dimension of English article use.*

Referentiality Dimension		
Referentiality Type	**Identifiability Type**	**Identifiability Source**
Referential	Identifiable	Shared lexis
		Shared speech situation
		Frame
		Current discourse
	Non-identifiable ("New reference")	
Non-referential Categorization		

3.1.2.2. Models including Discourse Modes

As detailed in Chapter II, Du Bois (1980) argues that Discourse Mode is an important category for explaining natural language English article use, and he lists two main modes, Narrative and Descriptive, each of which is further divided according to

pragmatic functions that are reflected in syntactic patterns of article use. All in all, there are at least five Discourse Modes suggested by Du Bois: 1.) Immediacy Narrative Mode (so named by me, not by Du Bois), 2.) Normal Narrative Mode, 3.) Normal Descriptive Mode, 4.) Defining Descriptive Mode, and 5.) Deferred Descriptive Mode. This chapter will describe an experiment designed to test the effect of three of these modes on article use – the Normal Narrative, the Normal Descriptive, and the Deferred Descriptive. The Immediacy Mode and the Defining Descriptive Mode will not be tested here, since they likely have limited use relative to the other three.

3.1.2.2.1. The proposed 3 Discourse Modes Model

The 3 Discourse Modes Model proposed here for testing is composed of these three Discourse Modes: Normal Narrative, Normal Descriptive, and Deferred Descriptive. These three modes were described in Chapter II, but a recap of their operation will be offered here for the convenience of the reader. First, the Normal Narrative Discourse Mode evinces the typical article use pattern as it has been traditionally understood: Referential Identifiable referents are marked as definite while Referential Non-identifiable referents and most Non-referential referents are marked as indefinite. This mode is exactly equivalent to the Referentiality/Identifiability scale that has been detailed above. One way of thinking about that Referentiality/Identifiability scale is to think of it as the "Normal" pattern of article use when a narrative is being carried forward in English discourse.

Second, Du Bois' Normal *Descriptive* Mode differs from the Normal Narrative Mode in the way that articles are used with referents that have Frame-activated Grounding. Whereas these types of referents are marked with the definite pattern in the

Normal Narrative Mode, in the Normal Descriptive Mode they follow the indefinite pattern. Du Bois suggests that this phenomenon may be at least partly due to the fact that speakers normally shift to the Descriptive mode from the Narrative mode when they feel the need to mention something that can't be assumed, and in so doing, the new information must be marked as new or noteworthy by giving it an indefinite marking, even if it might usually or otherwise be Frame-activated and therefore given a definite marking. For example, consider the following items. In the Normal Narrative Mode in (57), the boy who is wearing a hat has already been introduced, but no mention of his *hat* has yet been made. However, the mere existence of the boy in the narrative has provided sufficient enough context to trigger a Frame-activation that allows the first mention of *hat* to be marked as definite with the definite personal pronoun *his*.

(69) and he rides off, and passes um… a girl… on the road… and-- his hat [1st mention] falls off (Du Bois, 1980:243)

However, (58) provides an example of a Normal Descriptive Mode use of the indefinite pattern marking an initial mention of *shirt*. The mention of *man* is not sufficient in this context to allow a Frame-activated definite marking on this particular *shirt*, possibly because the speaker wants to specify information (navy blue) that cannot be assumed by the listener's understanding of the frame, and therefore the speaker switches to Descriptive Mode and uses the indefinite article normally associate with that mode.

(70) A--nd you see a middle-aged .. u--m .. Chicano man,… who's wearing .. a-- .. navy blue shirt (Du Bois, 1980:243)

Third, the Deferred Descriptive Mode is used when the momentum of the narrative causes the speaker to make an initial mention of something without giving an

adequate explanation of it. The explanation is deferred for a short time, perhaps only for a phrase or two, but when the speaker finally gives attention to it in order to "officially" introduce it in the discourse, it receives the same indefinite marking that is normally given for initial mentions, even though it does have Current Discourse Grounding. An example is seen in (71), where even though the second mention of *baskets* has Current Discourse Grounding, it is not marked as definite but as indefinite.

(71) [The Pear Man] picks pears,… puts them in.. his apron,… climbs down the ladder,… and empties the pears.. into… Ø big.. baskets [1st mention]. …tsk There's like <u>Ø three baskets</u> [2nd mention] sitting there (Du Bois, 1980:229)

In summary, the 3 Discourse Modes Model will comprise these three Discourse Modes suggested by Du Bois (1980), and its ability to predict accurate article use will be compared with that of the other models proposed in this study (see Table 7 on p. 87, below).

3.1.2.2.2. The proposed 2 Discourse Modes Model

In order to compare the performance of the Discourse Modes of the 3 Discourse Modes Model, another model containing only 2 Discourse Modes will also be tested. The 2 Discourse Modes Model will consist of these two modes: a Normal Discourse Mode and a new mode that I propose, a "Reintroducing" Discourse Mode. Under this system, Du Bois' Normal Narrative Mode will be preserved, but will be labeled simply as the Normal Mode, and this Normal Mode will be considered the default mode for all narrative *and* descriptive pragmatic tasks in discourse. As explained above, it is

proposed that for most discourse tasks, speakers employ a single set of article usage patterns that varies only according to the Referentiality/Identifiability type (and the Grammatical Number) of the noun phrases being used. Speakers vary this Referentiality/Identifiability pattern only for special pragmatic purposes, such as those described in detail by Du Bois (1980). However, rather than needing two separate descriptive modes (i.e., Du Bois' Normal and Deferred Descriptive Modes), it may be possible that a single Reintroducing Mode can serve in their place.

The purpose of the Reintroducing Mode to allow a speaker to clarify a real, perceived, potential, or posited misunderstanding or lack of understanding concerning a discourse referent. When this function is needed, the speaker simply *acts* as if that referent has not been properly introduced and therefore "reintroduces" it, using the indefinite pattern, even if it would otherwise (under the Normal Mode pattern) be marked as definite. The article usage pattern in the Reintroducing Mode differs from that in the Normal Mode only in the cases where the referentiality/identifiability type is either Frame or Current Discourse grounding. Although it is *possible* to use the indefinite pattern to "reintroduce" a referent that has Shared Lexical Understanding or Shared Speech Situation grounding, it is unlikely that a speaker would do so in most discourse contexts[14]; therefore, even in the Reintroducing Mode, when references have Shared

[14] To illustrate why this is unlikely, consider the following conversation, where, even if there is an apparent breakdown in communication, the speaker does not switch to the Reintroducing Mode because she is likely to assume that the listener cannot hear rather than that he does not know the meaning or identity of *the president*:

A. *The president* was shot.
B. Who?
A. *The president. / *A president.*
B. Who?
A. *The president! / *A president!*

Lexical Understanding grounding or Shared Speech Situation grounding, they will normally retain the definite pattern to mark their clear identifiability.

In the 3 Discourse Modes model, the term "three baskets" in item (71), above, would be classified as being in the Deferred Descriptive Mode, but in the 2 Discourse Modes Model, it would be considered as being in the Reintroducing Mode. Under the 2 Discourse Modes Model, although it could be considered as having a Current Discourse grounded Referentiality status, and therefore normally receive a definite *the* marking, since the speaker feels that the baskets have not been properly introduced and therefore may cause rhetorical if not referential confusion, he or she switches to Reintroducing Mode for that noun phrase. Note that the sentence begins with "There's like…," which signals this switch.

Additionally, in the 3 Discourse Modes model, the "shirt" in item (58), above, would be classified as being in the Normal Descriptive Mode. However, in the 2 Discourse Modes Model, it would be categorized as being in the Reintroducing Mode. In this 2 Discourse Modes Model, even though the man could be assumed to be wearing a shirt, and therefore the shirt might have Frame-activated grounding and therefore be expected to have definite marking, the speaker wishes to introduce the shirt properly, since the Frame activation does not suffice for his rhetorical or information-conveying needs. Therefore, the speaker switches discourse modes to the Reintroducing Mode for that noun phrase.

By combining the two descriptive modes suggested by Du Bois into one Reintroducing Mode, we hypothesize that simplifying the model in this way will nonetheless account for the same range of data. Whether this hypothesis is true or not

can best be assessed by comparing the two versions – the 2 Discourse Modes Model and the 3 Discourse Modes Model – in a side-by-side performance test (see Table 7, below, for a schematization of all the models to be tested).

However, in order to see if either model involving a multiplicity of Discourse Modes offers any benefit over having no separate Discourse Modes, the No Discourse Modes (Referentiality Only) Model will also be tested (as described above). This model will employ only the predictions made by the Normal Narrative Mode, which is in effect the same as testing the Referentiality/Identifiability system alone.

3.1.2.3. The Models Schematized

In summary, the proposed models for English article use contain either two or three dimensions: All four models contain Number and Referentiality/Identifiability dimensions, and two of those models contain an additional Discourse Mode dimension. Table 7, below, schematizes the models. In the models with Discourse Modes, the Discourse Mode dimension is arranged horizontally, and in order to squeeze three dimensions into a two-dimensional rendering, within each Discourse Mode column is nestled the Number dimension so that "SG, PL, UC" (i.e., Singular, Plural, Uncountable) can be seen repeated in each column. In models without Discourse Modes, the Number dimension is simply arranged horizontally. Although all three Number types are presented in this schematic, only Singular nouns will be included in the performance tests (for reasons described above). In all models, the Referentiality Type dimension is arranged vertically.

Table 7. *The 4 Models of English Article Use Proposed for Testing.*

Referentiality Dimension			Initial vs. Subsequent Mention Model	No Discourse Modes (Referentiality Only) Model	2 Discourse Modes Model		3 Discourse Modes Model		
					Normal Discourse Mode	**Reintroducing Discourse Mode**	**Normal Narrative Discourse Mode**	**Normal Descriptive Discourse Mode**	**Deferred Descriptive Discourse Mode**
Referentiality Type	**Identifiability Type**	**Identifiability Source**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**	**SG, PL, UC**
Referential	Identifiable	Shared lexis		the, the, the	the, the, the	the, the, the	the, the, the	the, the, the	the, the, the
		Shared speech situation		the, the, the	the, the, the	the, the, the	the, the, the	the, the, the	the, the, the
		Frame		the, the, the	the, the, the	a, Ø, Ø	the, the, the	a, Ø, Ø	the, the, the
		Current discourse	the, the, the	the, the, the	the, the, the	a, Ø, Ø	the, the, the	the, the, the	a, Ø, Ø
	Non-identifiable ("New reference")		a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø
Non-referential Categorization				a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø	a, Ø, Ø

3.2. Methodology for testing competing models

3.2.1. Experimental design

In order to test different models of English article usage by comparing their predictions to actual article use and to the level of correct predictions by native speakers, an experiment was designed with the goal of eliciting a variety of Discourse Modes and Referentiality/Identifiability Types in reasonably naturalistic conversations. Four versions of the model were proposed. Data was collected from 9 pairs of speakers having conversations about a video that one of them watched while the other asked prepared questions about it. The questions they were instructed to ask were designed to elicit responses that would include sufficient variety of Discourse Modes and Referentiality/Identifiability Types so that the four models could be adequately tested. The data was collected, transcribed, and coded. As a point of reference, human

participants employed their intuitions on the use of articles in a sizeable sample of the corpus. After the data was analyzed, the results suggested that two of the four models performed on the level of human predictors, with the other two models performing less well.

3.2.2. Participants

Eleven female and 7 male participants took part in the recording sessions. All were university undergraduate students who were native speakers of standard American English. Of these 18 participants, 5 females and 4 males functioned as the video watchers – i.e., the ones whose speech was analyzed for this experiment. In addition, another 4 native speaker participants – 3 females and 1 male – acted as "human predictors" (described below). Of these human predictors, 1 was a university undergraduate, 1 was a university graduate, and 2 were graduate students.

3.2.3. Data collection

Pairs of participants were recorded as they had conversations about a 3½ minute video ("Mr. Bean Reclaims His Trousers"). The video depicts a character ("Mr. Bean") who, upon donning his trousers in a public dressing room, realizes that he has put on the wrong pair of trousers because someone has mistakenly taken his. He goes in search of his own trousers, and after a series of strange antics, succeeds in swiping them back from the man who had taken them. The video was chosen because it contained no words that might prime article use in the experiment participants and because it contained a number of distinct scenes with some referents that appeared in multiple scenes and some that

were introduced in scenes along the way, conditions conducive to eliciting the types of data that were being sought.

In each pair of participants, one participant was assigned to watch the video and the other was assigned to ask questions about it after each scene. The two sat facing each other across a table on which sat a video screen. The participants could see each other, but the questioner could not see or hear the video. The video contained 5 scenes, and the experimenter, who sat at one end of the table with a computer, an additional screen, and a computer mouse, played a scene for the watcher and then paused the video to give the questioner time to ask the watcher questions about the events, characters, and setting in that scene. When the questioner decided he or she was finished asking questions, the experimenter played the next scene, pausing at the end of the scene and again giving the participants time to discuss the scene. This process was repeated until the entire video had been watched and discussed.

Before the experiment began, the following instructions were given to the watcher and questioner.

Video watcher's instructions:

Your job is to watch the 3 ½ minute video and tell your partner about it. You partner will not be able to see the screen, so she or he will rely on you to describe the video. Your partner will be able to ask you questions, so feel free to interact normally as you answer the questions.

The experimenter will pause the video at various points so that you will have a chance to take time to tell your partner about what's going on up

to that point. Try to wait for your partner's questions, and give most of your description during these pauses.

At the end of the video, your partner will be given a simple quiz about the information she or he learned from you about the video. If you partner answers accurately, both of you will receive a (very, very) small prize!

In keeping with the experimental design, the questioner's instructions were designed with the purpose of having the questioner elicit responses that would include the various Referentiality/Identifiability Types and Discourse Modes being investigated.

Questioner's instructions:

Your partner will watch a video that you will not be able to see. Your job is to get information from your partner. To get the information, please listen to your partner's descriptions and interact with your partner, asking questions as needed. The experimenter will pause the video from time to time so that the two of you will have a chance to discuss it. Feel free to talk to your partner as much as you want, but try to interact mostly (or only) when the experimenter pauses the video.

The specific information you need is in three types, so please ask the following questions to get these three types of information: (1) what

happened, (2) what the setting looked like and (3) what the people looked like.

Therefore, during each pause, please ask the following (feel free to paraphrase):

1.) Can you describe what's happening so far?

2.) Can you describe the place where it's happening? Or: Can you describe what the setting looks like?

3.) Can you describe the people who are in the scene? Or: Can you describe what the people in the scene look like?

And then, if you can, repeat the questions for extra clarity (even if it seems unnecessary):

4.) Can you describe what's happening again?

5.) Can you describe the place again?

6.) Can you describe the people again?

Focus on asking those questions (again, feeling free to paraphrase) and on getting those types of information – i.e., (1) what happened, (2) what the setting looked like and (3) what the people looked like. Please directly request those three specific pieces of information throughout the video to make sure you clearly understand the answers to those questions.

You will be given a simple quiz at the end of the video. If you are able to accurately describe the three types of information – i.e., (1) what happened, (2) what the setting looked like and (3) what the people looked like – both you and your partner will get a (very, very) small prize!

The simple quiz and small prize at the end of the session were intended to serve as extra incentives to motivate the participants to fully engage in the task.

3.2.4. Exclusion of data from analysis

In all, conversations from 9 pairs of participants were recorded using a digital audio recorder and transcribed using the Transcriber software application (Barras, 2002).

Of the noun phrases spoken by the video watchers in the transcribed texts, all that could potentially have an article were identified as being marked by the speaker with one of the three patterns available in English articles: *a*, *the*, or Ø (i.e., NPs without *a* or *the* were explicitly marked with Ø). (Since *an* is an allomorph of *a*, all instances of *an* were treated as *a*.) Many noun phrases (NPs), however, were excluded from the analysis because they were found in constructions that precluded any possible use of either *a*, *the*, or both. Such constructions include those in which the NP was marked with the demonstratives *this, these, that, those*; the possessive pronouns *my, your, his, her, its, their*; the interrogative pronouns *what, whose*; the quantifiers *some, any, several, all, every, no* and the numeral quantifiers *one, two, three,* etc.; and any NP that was modified by a possessive NP (e.g., *the guy's* _____).

A number of other factors also caused items to be excluded. The first was based on grammatical Number. In order to simplify the analysis by removing a factor that was deemed to have little theoretical importance to the models, all Plural and Uncountable NPs were excluded, leaving only Singular NPs to be included in the analysis.

In addition, any NP or article was excluded that was involved in any obvious or suspected repair or repetition that resulted in a noun-less article (e.g., the first two

instances of *the* in *the, the, the stall door*), incomplete or unclear noun (*the pa-, his pants*), or other ambiguity (*there's a red floor, Ø painted floor…*). Also excluded were any articles that were unclear in the recording, especially those few cases were *a* could be mistaken for *uh* or vice versa; also, in some cases in very rapid speech, the phonological shortening of *the* was apparently so extreme that it was difficult to determine if it was really there or not, and therefore such items were excluded. Another reason for exclusion was coding ambiguities: in all cases where any coding categorization was indeterminate, those items were excluded. For example, if the Discourse Mode could not be clearly determined for some reason, the item was discarded.

NPs were also excluded from the analysis if they were involved in any of three constructions which themselves alter the normal article distribution patterns: possessive prepositional phrase "of" constructions, relative clauses (or "that" constructions), and coordinating conjunction constructions (elsewhere here also called "and" constructions, but also including other coordinating conjunctions such as *or* and *but*).[15] Regarding the "of" constructions, NPs following the preposition "of" (i.e., NPs inside the prepositional phrase) were excluded from analysis, but the possessed NPs (i.e., the NPs before "of") were not excluded (e.g., in the phrase *the beginning of the scene*, the NP *the beginning* was retained, but *the scene* was excluded). Regarding the "that" and "and" constructions,

[15] "Of" constructions and "and" constructions alter the "normal" article distribution patterns in NPs because the definiteness status bestowed by the article before the first NP can be extended to the second NP as well (e.g., "*the* pair of white underwear" where "*the* pair of *the* white underwear" might otherwise be expected, or "*a* locker room or dressing room" where "*a* locker room or *a* dressing room" might also appear). Relative clause "that" constructions can alter the "normal" article patterns since they can sometimes license a definite marking on the phrase-external head noun (e.g., *"the* sink that's in *the* bathroom" where "*a* sink that's in *the* bathroom" might otherwise appear). Although this definite marking on such head nouns can probably be predicted in the models presented here by assigning them a "Frame" Referentiality Type, all relative clauses were nonetheless excluded in the analysis an effort to avoid a possible confound.

all NPs were excluded anywhere in the constructions (i.e., whether they came before or after the "that" or "and"; see Footnote 15 for examples).

Furthermore, as mentioned above, all NPs that were coded as Proper Names or Noun Incorporation were excluded from analysis since no prediction regarding article use could be made about them with the models being tested.

It should be noted that although many NPs were excluded from the analysis for the above mentioned reasons, they were still fully coded wherever possible, and the coding results (including numbers and types of excluded NPs) are included below in the Results section.

3.2.5. Coding

3.2.5.1. Removal of articles from the text

Before the transcribed texts of the recorded conversations were coded, all articles were removed from the text to prevent the coder (me) from using any information from the articles themselves when making decisions about the coding categories of the models, forcing the coder to rely only on information expressed in other linguistic structures in the context. The articles (including Ø) were replaced with blanks, and the coder then coded the following fields for each resulting article-less NP: Sequence of Mention, Number, Referentiality/Identifiability Type, and Discourse Mode for all proposed models.

3.2.5.2. Sequence of Mentions

NPs were coded according to the sequence of their mention in each text, since sequence of mentions is important for determining Current Discourse grounding vs. Non-identifiable ("New Reference") referentiality types. When an NP was mentioned for the first time in a text, it was coded as 1. All subsequent mentions were coded as 2. In cases where the same NP was repeated, determining 1st versus subsequent mention was straightforward. However, in cases were NPs with similar but possibly different meanings were employed, it was more difficult to determine if the second NP was co-referential with the first one. In such cases, a judgment was made by the coder as to whether the NPs were co-referential or not.

Another case where determining sequence of mention was not straightforward involved mentions of groups of individuals versus individual members of groups. In those situations, the idea of sets was used to determine the uniqueness vs. co-referentiality of NPs. If an NP was a set of four people, for example, that NP was treated as a separate and unique set of referents from an NP in which one person was mentioned, even if that one person had been a member of the previous set. In other words, for purposes of determining sequence of mentions, the one person was considered a set (of one), and the four-person group was considered as a separate set (of four). Thus whenever any referential *set* first appeared, it was treated as 1st mention, and only if that exact referential set appeared again was it coded as subsequent mention. An example of this kind of coding decision based on unique sets is seen in (72).

(72) Uh, there was Ø four men [1st mention] in the hallway and there was this, he had Ø particular fun with this one black guy [1st mention]

looking at the um wall. And he actually like pulled out his pants to see if it was his, and it, the guy [2nd mention] got disturbed and he walked away. So, there [inaudible], yeah, Ø four men [2nd mention]. (lines 7.47-55)

One more case where special coding consideration was required was when a noun was mentioned a second time but with new information in the NP that wasn't mentioned the first time. Since the second NP contained new information, it was treated as a 1st mention. In other words, the coding was based on information in the whole NP, not just the individual head nouns. An example of this kind of coding decision is seen in (73).

(73) OK, so in the video there's a man who is preparing, he's getting dressed and he's getting dressed uh in Ø formal dress clothes and he seems to be doing it rather awkwardly. He uh, it shows him first with his dress shirt on and a tie [1st mention], and he's about to put on his pants [1st mention] and he's doing it again very awkwardly, and he finally gets his pants [2nd mention] up and he looks down and he realizes that he has forgotten his shoes, and his dress socks are all the way rolled down, and he has no shoes on. He's wearing a white dress shirt [1st mention, since "white" brings entirely new information] with a red tie [1st mention, since "red" brings entirely new information] and just Ø black slacks [1st mention, since "black" brings entirely new information]…. (lines 8.3-15)

In any of the above situations involving sequence of mentions, if a particular case was indeterminate, it was coded as "unsure" and excluded from analysis.

3.2.5.3. Referentiality Type

NPs were coded according to one of the eight posited Referentiality Types: Proper Name, Shared Lexis, Shared Speech Situation, Frame, Current Discourse, Non-identifiable ("New Reference"), Categorization, and Noun Incorporation.

Proper Name. The four models do not make predictions about the article usage pattern with proper names, treating them instead as having their article pattern determined by their inherent lexically represented form; therefore, all Proper Names were coded as such so that they could be excluded from the analysis. The most common Proper Name by far in our data was "Mr. Bean."

Shared Lexis. NPs were coded as having Shared Lexis cognitive activation when they were judged to be identifiable to speakers based on general shared meaning outside the bounds of the speakers' immediately shared environment or discourse. The only NPs in the corpus coded as Shared Lexis were "the 30s" and "the 40s" (referring to the decades), as in item (74).

(74) or like a locker room ya know from probably like the 30s or the 40s' type style. (line 4.027-4.030)

Shared Speech Situation. NPs were coded as having Shared Speech Situation cognitive activation when they were identifiable based on the fact that the speakers shared something in their common environment or task, such as things inherent to videos (e.g., "clip," "scene," "beginning," "setting"). An example of an NP that was coded as having Shared Speech Situation is *main scene* in item (75):

(75) Um, it-- the main scene is in a… is just in a bathroom. (lines 8.151-8.153)

Frame. An NP was coded as having Frame-based cognitive activation when it was judged by the coder that "most people" would assume that the NP's existence in the narrative was extremely likely, predictable, or assumable based on other referents mentioned. Examples of such NPs include "the bottom half" of something, "the building" when it occurred after mention of components of a building such as a room or hallway, "the door" when it occurred after mentioned of a room, and "the restroom" when mentioned in the context of a school.

Current Discourse. In general, an NP was coded as having Current Discourse cognitive activation when it had been previously referred to in a particular discourse. The Sequence of Mentions coding result was used to distinguish between the Current Discourse vs. Non-identifiable ("New Reference") status of an NP, but only after the NP was judged not to be one of the other Referentiality/Identifiability Types. An example of NPs that were coded as having these two different statuses can be seen in the two mentions of the word *room* in item (76), where the first mention of *room* was coded as being Non-identifiable and the second mention was coded as having Current Discourse identifiability:

(76) OK, so there's a guy in a room [1st mention]…. (line 2.004)

OK, so he walked out of <u>the room</u> [2nd mention] (line 2.042)

Non-identifiable ("New Reference"). As when coding Current Discourse, *if all other Referentiality/Identifiability Types besides Current Discourse vs. Non-identifiable ("New Reference") were excluded*, then the Sequence of Mentions coding result was used to determine the Referentiality/Identifiability Type. When in those cases an NP was first mentioned, it was coded as being Non-identifiable ("New Reference"). Item (77)

illustrates NPs that were coded as being of the Non-identifiable ("New Reference") type, where *white shirt* and *tie* were both first mentions and where neither could be assigned another sort of identifiability such as identifiability from a Frame.

(77) and the main, main person is um wearing a white shirt [1st mention] and a tie [1st mention] (lines 2.018-2.020)

Categorization. An NP was coded as having a Categorization cognitive activation when it was used in a non-referential sense (i.e., not referring to a unique, identifiable entity) to classify something as belonging to a particular category. Examples of such Categories primarily included occupations ("he's like a comedian," "he kinda looks like a police officer") and rooms or structures ("it wasn't an apartment," "it reminds me of an older styled dressing room") (lines 9.004, 4.232, 2.043, and 4.026, respectively).

Noun Incorporation. Since, as with Proper Names, the current model does not attempt to make predictions about the complicated article usage patterns with NPs found in Noun Incorporation constructions, all NPs that appeared to be involved in Noun Incorporation were coded as such and excluded from the analysis. The most common such NP in the corpus was "going to the bathroom."

3.2.5.4. Discourse Mode

NPs were coded according to the Discourse Mode they occurred in, but two different Discourse Mode models were coded for so that they could be compared.

The 3 Discourse Mode Model, based on the questioner's questions. The coding of Discourse Modes for the 3 Discourse Mode Model was based primarily on the type of

questions that the questioner asked. As mentioned above, the questioner had been instructed to ask different types of questions in the hopes that these questions would elicit answers in three posited Discourse Modes: Normal Narrative, Normal Descriptive, and Deferred Descriptive. Under this system, if the questioner asked the video watcher to describe the action (e.g., "Can you describe what's happened so far?"), all NPs in the video watcher's responses that appeared to answer that question were coded as being in the Normal Narrative Discourse Mode. An example of NPs that were coded as being in the Normal Narrative mode can be seen in item (78), where *video, main character, first person, last scene,* and *bathroom*

(78) Questioner 8: And what is going on in the video?

Watcher 8: In the video, the main character he waits until the first person that was in the last scene leaves the bathroom (lines 8.221-8.226)

If the questioner asked the video watcher to describe the place or the people in the scene (e.g., "Can you describe what the people look like?"), then all the NPs in the response that appeared to answer that question were coded as being in the Normal Descriptive mode. An example of this can be seen in item (79), where *hallway*, *bulletin boards,wall,* and *tile floor* were all coded as being in the Normal Descriptive mode.

(79) Questioner 9: Um, can you describe the place where it's happening?

Watcher 9: Uh, it looks like just some hallway. There's like Ø bulletin boards on the wall, and Ø tile floor. (lines 9.043-9.047)

Likewise, if the questioner asked the video watcher to describe the place or the people *again* (e.g., "Can you describe the place again?"), then the NPs in the response that appeared to answer that question were coded as being in the Deferred Descriptive mode. In item (80), *hallway, tile floor, bulletin boards,* and *yellow wall painting* were all coded as being in the Deferred Descriptive mode:

(80) Questioner 9: Can you describe the place again?

Watcher 9: Yeah, it's a hallway, it's got a tile floor, Ø bulletin boards, uh yeah, Ø yellow wall painting. (lines 9.059-9.063)

Sometimes, however, even though the questioner asked a question that would be expected to elicit a certain type of response, the video watcher would give an answer that did not match the question type. For example, the questioner might ask "Can you describe the people in the video?" but rather than giving a description of what the people looked like (a Normal Descriptive response), the watcher would describe their actions (a Normal Narrative response). Or in some cases the speaker would begin with a description of the people but would soon thereafter switch to narrating their actions, even though the questioner had not asked about their actions. An example of a case where the video watcher's answer did not match the Normal Descriptive mode that the questioner's question was expected to elicit can be seen in item (81), where the questioner asked for a description but the watcher responded by narrating the action.

(81) Questioner 8: Alright, um, can you describe the people in the video?

Watcher 8: Um, there's one person that comes out of a bathroom and the main character follows him into the bath- er, follows into the bathroom as

the man leaves the bathroom, and he then is looking at a man who is going to the bathroom in a urinal, and he's coming up behind him, again to compare his pant length to the other man (lines 8.133-8.145)

In such cases, if the overall pattern of the responses was judged by the coder (me) to be a clearly identifiable Discourse Mode, the NPs in that section of discourse were coded according to the coder's judgments (judgments which were based solely on contextual information, since all articles had been removed from the text). In the case of item (81), for example, all the underlined NPs were all coded as being in the Normal Narrative mode rather than the "expected" Normal Descriptive mode. In addition to these guidelines, if the Discourse Mode appeared ambiguous, confusing, or unclear in any way, the relevant NP was marked as "unsure" and was excluded from the analysis.

The 2 Discourse Mode Model, based on line-by-line judgments. The coding of the 2 Discourse Mode Model was based on individual judgments about each NP in the corpus. This model involved two proposed Discourse Mode types: Normal and Reintroducing. Under this system, the Normal mode was considered the default mode, and all NPs that were not judged to be the Reintroducing mode were coded as Normal.

The Reintroducing mode was reserved for cases in which the speaker "reintroduced" an NP that had already been introduced earlier in the discourse. These were cases where a previously-mentioned referent was not only repeated (since most NPs were repeated many times in the discourse), but repeated in a way that was similar to the way an NP would be introduced the first time it was mentioned. An NP might be reintroduced in this way because the speaker thinks the questioner did not hear, remember, or understand the information the first time it was presented (maybe because

the questioner asks a question expressing lack of understanding), or perhaps because the speaker wants to re-enact the introduction of an NP for some rhetorical strategy, such as setting up the next part of his story. An example of NPs that were coded as exemplifying Reintroducing Discourse Mode can be seen in (82), where *some guy, another guy,* and *pants* had all been mentioned before and were here coded as Reintroducing.

(82) Questioner 2: OK. And what's happening again in the dressing room?

Watcher 2: Um, they're getting dressed,

Questioner 2: OK.

Watcher 2: and, yeah, some guy stole another guy's pants. (lines 4.031-036)

Another example can be seen in items (83) and (84). In item (83), the watcher is giving normal narration and descriptions, so all the NPs are coded as being in the Normal mode. In (84), however, the questioner asks for clarification, and the watcher switches to Reintroducing mode in an effort to comply with the clarification request; therefore, the NP *school hallway* was coded as being in the Reintroducing mode.

(83) Questioner 2: Alright, so, um, can you tell me little bit a about what's going on so far?

Watcher 2: OK, so he walked out of the room where he was and it wasn't an apartment, it's like a school. So now he's in like a hallway where I guess like Ø other teachers are? (lines 2.040-2.044)

(84) Questioner 2: OK. Um, you kind of explained my second question, so it's like in a school?

Watcher 2: Yeah. It's like a school hallway. (lines 2.059-2.060)

Item (85), below, shows another example, but one that highlights the difference between the 3 Discourse Modes Model and the 2 Discourse Mode Model. In the 3 Discourse Modes Model, there is no mode that can account for a switch to an indefinite marking on NPs to signal a "deferred" or repeated narrative in cases where the identifiability of the NPs has previously been established. Du Bois (1980) suggested a *deferred descriptive* mode to account for such a switch for NPs involved in descriptions of characters or scenes, but he did not posit a similar *deferred narrative* mode to account for the switch in NPs involved in narrative clauses, as we see in (85), where the watcher is clarifying the details of the narrative action. The Reintroducing mode in the 2 Discourse Mode Model, however, does make such allowances. In the 3 Discourse Modes Model, the NPs *some guy, another guy,* and *pants* in item (85), below, were all coded as being in the Normal Narrative mode. In the 2 Discourse Modes Model, however, any NP involved in a clarification can receive an indefinite marking to signal this switch to Reintroducing mode; therefore, the NPs *some guy, another guy,* and *pants* in (85) were all coded as being in the Reintroducing mode for the 2 Discourse Mode Model.

(85) Questioner 4: OK. And what's happening again in the dressing room?

Watcher 4: Um, they're getting dressed,

Questioner 4: OK

Watcher 4: and, yeah, some guy stole another guy's pants. (lines 4.031-4.037)

No Discourse Modes (Referentiality Only) Model, based on Normal Narrative mode. The No Discourse Modes (Referentiality Only) Model was a model simply based on only the Referentiality types that were described in the "Normal Narrative" mode that was proposed for the 3 Discourse Modes Model discussed above. Therefore, under this No Discourse Mode Model, all NPs were coded as if they were in the Normal Narrative mode of the 3 Discourse Modes Model. The purpose of proposing a system with no Discourse Mode dimension was to be able to compare it to other systems to see if there was any advantage in proposing a multiplicity of Discourse Modes.

In addition to the three models that contained a complex Referentiality dimension with 6 different Referentiality/Identifiability types to be tested, a simplified model consisting of only 2 Referentiality/Identifiability types – Initial vs. Subsequent Mention – was also tested in order to measure the advantage gained in utilizing the additional 4 Referentiality/Identifiability types. The coding field of Sequence of Mentions, which coded all NPs according to whether they were a first or subsequent mention in a given text, was used to determine the Initial vs. Subsequent Mention assignment. All NPs with first mention status were coded as Non-identifiable (also commonly known as "New reference"), while those with subsequent mention status were coded as having Current Discourse identifiability grounding (also commonly known as "Given" or "Old" Reference).

3.2.6. Predictions for each model

After all NPs were coded, the article for each NP was predicted for each of the four proposed versions of the model (see Table 7 on p. 89, above). The actual articles that were used by the speakers in the discourse were then compared to each model's predicted articles so that each model's accuracy could be assessed.

3.2.7. Human predictors

Four participants were recruited to read a sample of the transcriptions of the recordings in which all the articles had been removed. The participants manually filled in the missing articles (including Ø) line-by-line according to their own best judgments. Their results were compared with the results from the three models proposed here in order to see if any of the models could perform at the level of the intuitions of a native speaker of the language. The participants filled in articles for a 2-text sample of the corpus, which represented 23 percent, or 176, of the 759 total non-excluded Singular NPs in the corpus.

3.3. Results

The raw coding results are listed below in Table 8 and Table 9. Table 8 lists the counts and percentages of articles that were included in, and excluded from, the analysis, as discussed above in the Methodology section.

Table 9 shows the number of Singular NPs (since only Singular NPs were used in the analysis) that were coded in each coding field. The coding fields consist of Initial vs. Subsequent Mention, the basis for the Initial vs. Subsequent Mention Model; 8

Referentiality/Identifiability Types, the sole basis for the No Discourse Modes (Referentiality Only) Model and the partial basis for the remaining two Models (although recall that the Proper Names and Noun Incorporation types are excluded from the analysis since no predictions regarding article use can be made for them); the Discourse Modes for the 2 Discourse Modes Model; and the Discourse Modes for 3 Discourse Modes Model.

Table 8. *Total article tokens in the corpus.*

Inclusion status	All NPs	Singular (SG) NPs	Plural (PL) NPs	Uncountable (UC) NPs	Unspecified or missing NPs
Articles considered in analysis	1005 (71.1%)	852 (60.3%)	126 (8.9%)	27 (1.9%)	0 (0.0%)
a	224 (15.8%)	223 (15.8%)	0 (0.0%)	1 (0.1%)	0 (0.0%)
the	585 (41.4%)	516 (36.5%)	62 (4.4%)	3 (0.2%)	0 (0.0%)
Ø	196 (13.9%)	109 (7.7%)	64 (4.5%)	23 (1.6%)	0 (0.0%)
Articles excluded from analysis	409 (28.9%)	221 (15.6%)	82 (5.8%)	16 (1.1%)	90 (6.4%)
Excluded due to unsure coding	93 (6.6%)	47 (3.7%)	33 (2.3%)	4 (0.3%)	9 (0.6%)
Excluded due to speaker repairs	104 (7.4%)	21 (1.5%)	2 (0.1%)	0 (0.0%)	81 (5.7%)
Excluded in “of” constructions	116 (8.2%)	85 (6.0%)	26 (1.8%)	5 (0.4%)	0 (0.0%)
Excluded in “that” constructions	44 (3.1%)	34 (2.3%)	9 (0.6%)	1 (0.1%)	0 (0.0%)
Excluded in “and” constructions	52 (3.7%)	34 (2.4%)	12 (0.8%)	6 (0.4%)	0 (0.1%)
Total articles in corpus	1414 (100%)	1073 (75.9%)	208 (14.7%)	43 (3.0%)	90 (6.4%)

Note: Another 14 potential tokens, not included in these totals, were excluded from analysis due to auditory ambiguity regarding whether the utterance was an article or not (e.g., “a” or “uh”). In addition, although NPs that were coded as Proper Names or Noun Incorporation will not be included in the analysis of the three proposed Discourse Model systems, they are included for purposes of the article counts in this table.

Table 9. *Coding results for all coding fields for Singular NPs*

Coding field	*n*	
Initial vs. Subsequent Mention		
Initial (Non-identifiable)	294	(34.5%)
Subsequent (Current Discourse)	558	(65.5%)
Total	852	(100.0%)
Referentiality Types		
Proper name	89	(10.4%)
Shared lexis	0	(0.0%)
Shared situation	21	(2.5%)
Frame	75	(8.8%)
Current	447	(52.5%)
Non-identifiable	169	(19.8%)
Non-specific	0	(0.0%)
Category	47	(5.5%)
Secondary Predicates	0	(0.0%)
Vocatives	0	(0.0%)
Full Noun Incorporation	0	(0.0%)
Finite Verb "Noun Incorp."	4	(0.5%)
Total	852	(100.0%)
Discourse Modes for 2 Discourse Modes Model		
Normal	818	(96.0%)
Reintroducing	34	(4.0%)
Total	852	(100.0%)

Table 9, continued.

Coding field	*n*
Discourse Modes for 3 Discourse Modes Model	
Normal Narrative	558 (65.5%)
Normal Descriptive	205 (24.1%)
Deferred Descriptive	89 (10.4%)
Total	852 (100.0%)

Note: Although all Proper Names and Noun Incorporation tokens are included here for sake of the overall NP counts, they were excluded from the analysis since none of the models make predictive claims about their article patterns.

Table 10 summarizes the performance of the various models that were tested; more detailed numbers can be found in the subsequent tables. In all cases, the results are reported as *miss rates*, or in other words the number (and percentage) of times that the model predicted the *wrong article* (or in other words, the number of times the prediction did not match the actual articles used by the native speaker participants in the audio recordings). For the two models that utilize Discourse Modes, a chi square test was performed to test whether the discourse modes differed from each other in terms of article use. Only Singular NPs were included in the test, and only *a* and *the* were tested, since the frequency of Ø marked NPs was too low in some of the modes to justify their inclusion in the tests (i.e., N = 1 for the Reintroducing Mode, and N = 3 for the Normal Narrative Mode). The results of the tests for each of these models showed that their Discourse Modes differed from one another in terms of their predictions for article distribution, $\chi^2(1, N = 739) = 41.94$, $p < .001$ for the 2 Discourse Modes Model, and $\chi^2(2, N = 671) = 42.68$, $p < .001$ for the 3 Discourse Modes Model.

The impoverished Initial vs. Subsequent Mention Model was tested to compare the efficacy of having a detailed system of Referential/Identifiability Types over the possibility of using a simple one. In this system, only Non-identifiable ("New Reference") (based simply on Initial Mentions) vs. Current Discourse grounding (based simply on Subsequent Mentions) was used as a basis for predicting articles, with Non-identifiable ("New Reference") predicting *a* and Current Discourse grounding predicting *the* to mark Singular NPs. This system performed worst of all, highlighting the importance of developing a fuller model of Referentiality/Identifiability.

The 2 Discourse Modes Model performed the best in the tests, with a slight 2% better miss rate than the No Discourse Modes (Referentiality Only) Model and a 6% better miss rate compared with the 3 Discourse Modes Model. Both the No Discourse Modes (Referentiality Only) Model and the 2 Discourse Modes Model were in the performance range that the four human predictors achieved on the 23% sample of text on which they did their predictions. In any event, the range of the humans' performance seems no better than the two higher performing models.

Table 10. *Miss rates of 4 Models and Human Predictors Compared (Singular NPs only)*

3 Discourse Modes Model	2 Discourse Modes Model	No Discourse Modes (Referentiality Only) Model	Initial vs. Subsequent Mention Model	Human Predictors avg*
89/759 (12%)	47/759 (6%)	63/759 (8%)	165/759 (22%)	11.75/176 (7%)

**Note:* The human predictors' average is based on their performance on a sample size of approximately one quarter of the entire corpus.

Table 11 gives greater detail about the 3 Discourse Modes Model, showing the overall performance of that system along with the miss rate that each of the individual Discourse Modes in the system achieved. The Normal Narrative Mode performed the best of the three modes by far, but it received a boost in performance by virtue of the other two modes filtering out some mismatches that would otherwise have occurred (as seen by comparison to No Discourse Modes (Referentiality Only) Model, in which the Normal Narrative Mode applied, unfiltered, to all NPs). The implementation of the Deferred Descriptive Mode in this model reduced rather than improved the overall accuracy.

Table 11. *Miss rates for 3 Discourse Modes Model (Singular NPs only)*

All 3 Discourse Modes Model Overall	Normal Narrative Mode Only	Normal Descriptive Mode Only	Deferred Descriptive Mode Only
89/759 (12%)	16/415 (4%)	29/189 (15%)	40/86 (47%)

Note: A chi square test indicated that the 3 Discourse Modes differed from one another in terms of their article distribution, $\chi^2(2, N = 671) = 42.68$, $p < .001$.

Table 12 shows the overall performance of the 2 Discourse Modes Model as well as its two Discourse Modes individually. This system benefited slightly from the filtering effect of the Reintroducing Mode, which removed some mismatches that the Normal

Mode would otherwise have suffered (again, as seen by comparison with the No Discourse Modes (Referentiality Only) Model, which was a system with only one mode identical to the Normal Mode of the 2 Discourse Modes Model).

Table 12. *Miss rates for the 2 Discourse Modes Model (Singular NPs only)*

2 Discourse Modes Model Overall	Normal Mode Only	Reintroducing Mode Only
47/759 (6%)	41/728 (6%)	6/31 (19%)

Note: A chi square test indicated that the 2 Discourse Modes differed from one another in terms of their article distribution, $\chi^2(1, N = 739) = 41.94$, $p < .001$.

Table 10 (above) shows the performance of the No Discourse Modes (Referentiality Only) Model, which is in effect a simplified "single Discourse Mode" model which is identical to the "Normal Mode" of the 2 Discourse Modes Model and the "Normal Narrative Mode" of the 3 Discourse Modes Model. In other words, it is a system which is in effect lacking in a Discourse Mode dimension (since there is no choice of Discourse Mode) and consisting of only Number and Referentiality dimensions. In these sets of data, since only Singular NPs were analyzed, the Referentiality/Identifiability system alone (i.e., the No Discourse Modes (Referentiality Only) Model) can be seen to perform virtually as well as the best Discourse Mode system (the 2 Discourse Modes Model). Additionally, as comparing the No Discourse Modes (Referentiality Only) Model and the Initial vs. Subsequent Mention Model demonstrates (again in Table 10, above), incorporating a multi-type Referentiality/Identifiability

system greatly improves article-predicting ability over a very simple two-type Referentiality/Identifiability system consisting of simply Initial Mention (Non-identifiable) vs. Subsequent Mention (Current Discourse grounding) types.

Table 13 shows the performance of 4 human predictors whose instructions were to use their intuition to fill in the missing articles in a portion of text containing 23% of the analyzable Singular NPs of the entire corpus.

Table 13. *Performance of Human Predictors on Sample of 23% of the Corpus*

	Participant A		Participant B		Participant C		Participant D		Average	
Miss rate*	10/176	(6%)	14/176	(8%)	13/176	(7%)	10/176	(6%)	11.75/176	(7%)

Note: The human predictors' average is based on their performance on a sample size of 23 percent of the total non-excluded NPs in the corpus.

3.4. Discussion

The results from the experiment indicate that the 2 Discourse Modes Model performed best (with a 6% miss rate), followed closely by the No Discourse Modes (Referentiality Only) Model (with an 8% miss rate). The 3 Discourse Modes Model performed third best (with a 12% miss rate), and the Initial vs. Subsequent Mention Model performed worst (with a 22% miss rate).

The best performing model (with only a 6% miss rate, which is in the range of the human predictors) was the 2 Discourse Modes Model, the model that employs a Normal and a Reintroducing Discourse Mode and the full Referentiality/Identifiability dimension.

The Referentiality/Identifiability article distribution pattern that appears in the Normal Discourse Mode was the primary contributing factor to the overall good performance of the 2 Discourse Modes Model. In fact, the No Discourse Modes (Referentiality Only) Model – employing only what is in essence the Normal Discourse Mode of the 2 Discourse Modes Model – performed (with only an 8% miss rate) almost as well as the 2 Discourse Mode Model, suggesting that there is just a slight benefit to adding a Reintroducing Mode to the model. The No Discourse Modes (Referentiality Only) Model was based primarily on a series of Referentiality/Identifiability Types from Givón's (2001, 2005) treatment of referent cognitive accessibility. Overall, this approach to referentiality/identifiability, when supplemented with some non-referential types inspired by Du Bois (1980), accounts well for the conversation data considered here. However, the 2% performance boost derived from adding the Reintroducing Discourse Mode, though small, may indeed represent an important improvement, considering that it was derived from only a very small number of tokens that were actually coded as being in the Reintroducing Mode – 31 out of 759 tokens. In other words, being able to recode just 4% of the tokens into the Reintroducing Mode resulted in a 2% overall improvement in performance.

The worst performing model, as expected, was the Initial vs. Subsequent Mention Model (with a 22% miss rate). The main purpose for testing this model, with its drastically delimited 2-type Referentiality/Identifiability system (i.e., Non-identifiable "New Reference" vs. Current Discourse grounding), was to provide a basis for comparison with the more complex Referentiality/Identifiability system contained in the other models. Since the No Discourse Modes (Referentiality Only) Model (with its 8%

miss rate) consisted of the most common article distribution pattern for the full Referentiality/Identifiability system – as derived from Givón (2001, 2005) and Du Bois (1980) – it offers the clearest comparison with the Initial vs. Subsequent Mention Model. The No Discourse Modes (Referentiality Only) Model's improvement in performance over the Initial vs. Subsequent Mention Model demonstrates the definite advantage of including other Referentiality/Identifiability types. Simply put, articles are not used merely to differentiate Non-identifiable "New Reference" vs. Current Discourse grounding but have more complex functions in discourse that require more complex models of Referentiality/Identifiability.

The second-worst-performing model was the 3 Discourse Modes Model (with a 12% miss rate). There are a number of reasons why this model may have performed relatively poorly. First and foremost, the specifications for this model's three Discourse Modes were garnered from Du Bois (1980), but implementing the model's Normal Descriptive Mode and Deferred Descriptive Mode proved difficult for at least two reasons. First, although Du Bois (1980) offers discussions of how the Normal Descriptive Mode and the Deferred Descriptive Mode work, coding them on a case-by-case basis in a corpus proved difficult since the examples and descriptions in Du Bois (1980) did not sufficiently translate into a unambiguous, implementable definition of these modes. Second, in an effort to provide an additional, independent criterion for coding NPs into the proper 3 Discourse Modes, the experiment was designed to have questioners ask questions that would cue the video watchers to produce speech that would fall into these 3 Modes. For coding purposes (as described above in the Methodology section), the main criterion used to code each NP into one of the 3

Discourse Modes was the type of question that the questioner asked. This methodology, however, was not always sufficient to force speakers to produce discourse that matched the question's intended Discourse Mode. Therefore, the relatively poor performance of the 3 Discourse Modes Model may be due in large part to the difficulties in implementing Du Bois' (1980) descriptions of two of the model's three Modes.

Furthermore, the 3 Discourse Modes Model may have performed poorly because some of its intended Discourse Mode functions might have been taken over by parts of the *Referentiality/Identifiability system*. In other words, some of the Referentiality/Identifiability Types (as specified here) may have in effect fulfilled the function that Du Bois (1980) originally intended for the Discourse Modes to do. First, the "Category" Referentiality type, as here implemented, in some cases correctly predicts indefinite marking on NPs in the Normal Narrative Mode that might otherwise require the Deferred Descriptive Mode to correctly predict. This situation mitigates the need for the Deferred Descriptive Mode. For example, in one case, a speaker responded to a question in the Deferred Descriptive Mode by saying "It's… a school…," using the indefinite marking even though *school* had been mentioned before. The Deferred Descriptive Mode would correctly predict this indefinite marking if the noun had been coded as having a Current Referentiality type. But in this case the noun had been coded as having a Category Referentiality type, which predicts the indefinite marking in all Discourse Modes for all models anyway, thus removing the advantage in this case that the Deferred Descriptive Mode might have otherwise conferred. Second, the model's definition of the Non-identifiable ("New Reference") Referentiality/Identifiability type also in some cases lessened the need for a Deferred Descriptive Mode. As mentioned

above, under the full, multiple-type Referentiality/Identifiability system as defined here, a referent could be considered to be Non-identifiable ("New Reference") if the NP carried *any* type of new information, even if the head noun itself had already been mentioned in the discourse. For example, if *a shirt* had been mentioned earlier in the discourse but was mentioned again as being *a white shirt*, by virtue of the new information (i.e., *white*), this mention of the NP was marked as being Non-identifiable ("New Reference"). This partially removed the necessity of specifying a separate Discourse Mode to account for changing the article to *a* from *the* in such cases.

The final and perhaps most important reason that the 3 Discourse Modes Model offered no advantage over the 2 Discourse Mode Model is that all 3 Discourse Modes in the 3 Discourse Modes Model may be unneeded since the primary functions of two of them really can be described by the "reintroducing" function that is accomplished by the single Reintroducing Mode. The basic reason for positing the Normal Descriptive and Deferred Descriptive Modes is to specify when a Frame reference (in the case of the Normal Descriptive Mode) and when a Current Discourse reference (in the case of the Deferred Descriptive Mode) should not receive their "normal" definite marking but should instead receive an indefinite marking. In contrast to this, the Reintroducing mode basically specifies that for any given referent (whether Frame or Current Discourse), if the speaker wants to "reintroduce" them, she can do so by marking them with the indefinite marking. Thus, the need for two separate Discourse Modes is eliminated by combining them into one.

Even though the 3 Discourse Modes Model did not perform well in this test, it could be the case that if it were implemented differently, its accuracy could be improved.

As mentioned above, one of the main problems with the 3 Discourse Modes Model was implementing the coding of the Deferred Descriptive Mode. If the coding criteria were altered so that less emphasis was given to using the questioner's questions and more emphasis were given to a line-by-line reading within the larger context as a basis for determining the discourse mode, the accuracy of the Deferred Descriptive Mode might be improved. In addition, the data elicitation design could also be altered in a way that might elicit fewer "false" deferred descriptions. One way to do this might be to have the questioner watch for descriptions of referents that were only mentioned once in the narration in passing and without any descriptive commentary. Whenever the questioner noticed such a reference, she could then ask the speaker to describe that referent. This might elicit "Deferred Descriptive" descriptions that are more aligned with Du Bois' (1980) original explication of that discourse mode. In this way, it might be possible to "rescue" Du Bois' (1980) discourse modes and reassess them in a new comparison with the 2 Discourse Modes Model that was proposed here.

Although the results of this study demonstrated the advantages of certain models of English article use over other models, one additional important contribution of this experiment to the field is the experimental paradigm employed in the testing of the models. This experimental paradigm was itself tested here, and its successful implementation has demonstrated that it had the power and sensitivity to discriminate between the models that were tested. Based on the utility it demonstrated here, it is likely that this approach can be applied generally for testing not only models of English articles but also pragmatic models generally. As current features are refined or new features are added to the model of English articles, the paradigm can be used to assess each stage and

version of model development. But beyond that, it could foreseeably be adapted and applied to assessing the predictive performance of models of functional aspects of any language.

3.5. Conclusion

This experiment tested various models of English articles by utilizing a data collection paradigm that was designed for gathering naturalistic discourse specifically suited for the needs of the tests. It demonstrates the feasibility of testing and improving such models experimentally, and it offers a methodological framework for doing so. The methodology was not perfectly implemented, since the questions that were designed to elicit clear responses in various Discourse Modes did not consistently produce that outcome. More careful question design and participant preparation may alleviate such problems in the future. However, even with these problems, the data collected did provide enough variety of responses to allow suggestive results regarding the comparison of the Discourse Mode and Referentiality/Identifiability Type variables.

This study revealed two main findings. First, a full Referentiality/Identifiability system is important for describing the usage of English articles. Although part of the pattern of article use can be described with reference to the Non-identifiable ("New Reference") type (which receives an indefinite marking) and the Current Discourse type (which receives a definite marking), a number of additional Referentiality/Identifiability types is required to describe the full pattern. This study tested a model that relied only on Referentiality/Identifiability Types (Shared Lexical Understanding grounding, Shared Situation grounding, Frame Activated grounding, Current Discourse grounding, Non-

identifiable ("New Reference"), and Non-referential Categorization), and this relatively complex model performed substantially better than a simple model consisting of only Initial mentions (as Non-identifiable "New References") vs. Subsequent mentions (with Current Discourse grounding) , demonstrating the necessity of a more complex Referentiality/Identifiability system.

The second main finding is that adding a dimension of Discourse Modes may be necessary for an accurate model of article usage, even if the additional Modes apply to only a small percentage of NPs in a given discourse. In the case of the models tested here, adding a Discourse Mode dimension that consisted of 2 Discourse Modes resulted in a 2% performance boost over having no Discourse Modes (or in other words, over having a Referentiality system only). Although this improvement may seem small, it came from reanalyzing just 4% of the NPs in the corpus as being in a separate Mode (viz., the Reintroducing Mode) from the Normal "default" Mode.[16] In other words, the 2 Discourse Modes model offered a 50% improvement (i.e., 2%/4%) over the No Discourse Modes model for the relevant NPs. Without including Discourse Modes in a model of English article usage and instead relying on a Referentiality/Identifiability system alone, there will likely be a small percentage of NPs whose article patterns will never be predictable. While the 2 Discourse Modes Model tested here is certainly not fully sufficient to describe all article usage in English (additional modes will be proposed in Chapter IV), it demonstrates the importance of including Discourse Modes in article usage models.

[16] To be clear, the entire corpus was recoded for the 2 Discourse Mode model, but only 4% of the analyzed NPs in the corpus ended up being in the Reintroducing Mode.

For the data tested here, both of the best-performing models had predictive accuracy in the range of the human predictors. In fact, the performance of the best model (the 2 Discourse Modes Model) was as good as the two best-performing human predictors. While it is possible that this model may not prove as accurate if applied to other types of data – such as data from other genres – its success in this test suggests that this general approach to model making and testing is a useful one for efforts to describe article usage. It may also prove useful for investigating other semantic-pragmatic questions, such as how issues of referentiality/identifiability and discourse modes are expressed cross-linguistically.

The model proposed here is only a portion of a larger model that must be developed in order to account for the full range of article use in English. With the addition of a more comprehensive treatment of the Grammatical Number system, the expansion and refinement of the Referentiality/Identifiability types, and the development of the Discourse Mode dimension, a model of the whole article system may be possible. Chapter IV will discuss these additional necessary features and propose a comprehensive model.

CHAPTER IV

TOWARDS A COMPREHENSIVE SEMANTIC MAP: ADDITIONAL ATTRIBUTES

4.1. Theoretical background revisited: implications for a semantic map for applied linguistics

In the previous chapter, four models of English article use were proposed and compared. The most accurate model contained a Discourse Modes dimension with two attributes (Normal and Reintroducing), and it provided more accurate predictions regarding English article use than did the two models that did not include Discourse Modes, suggesting that Discourse Mode is an important and necessary dimension in accounting fully for the distribution of articles. In addition, all of the three most accurate models used the same somewhat complex Referentiality dimension, a dimension that contained six attributes (Shared Lexical Understanding, Shared Speech Situation, Frame grounding, Current Discourse grounding, Non-identifiable ("New Reference"), Non-referential Categorization). One of the three models that utilized these six Referentiality/Identifiability attributes was a model that contained *only* this Referentiality/Identifiability dimension (i.e., it had no Discourse Mode dimension), and so it served as a clear basis for comparison with the fourth model, a model that contained only a simplified Referentiality/Identifiability dimension with only two attributes (Non-identifiability ["New Reference"] vs. Current Discourse grounding; or, in other words, for this fourth model, Initial vs. Subsequent Mention). The more complex 6- attribute

Referentiality/Identifiability dimension provided a large (14%) improvement in accuracy over the simpler 2- attribute model, suggesting that a complex Referentiality/Identifiability dimension is needed to fully account for article usage.

However, all of the models in the previous chapter were designed with the intent of including only attributes that were testable with the methodological paradigm that was used and the type of data that was able to be elicited. Therefore, even though the tested models supported the idea that relatively complex, multi-nodal dimensions are needed, there are certainly additional attributes that were not included in the tested models that nevertheless must be included in a full model of English article usage. Based on information in the literature on articles (see Chapter II), we can propose with some degree of confidence what at least some of those additional attributes likely should be. This chapter will offer a discussion of these additional attributes.

One reason for creating a semantic-map-based model of English articles is because doing so provides way to systematically incorporate and unify all of the functional "attributes" that the relevant literature has implicated in article use. An additional reason for creating such a semantic-map-based model is to provide a clear overview of the article system for students and teachers of English as a second language. The article system is notoriously difficult for English language learners to master, and the task is not made easier by the fact that most English teachers do not themselves consciously understand the many parts of the system and therefore cannot explain it to students. English language textbooks also are unable to clearly, fully, and succinctly explain the system. Even though some do provide an extensive list of rules and a catalog of examples of article usage, they ultimately fail to present the system in its entirety. A

model of English article use therefore has the potential to provide a much-improved sense of clarity in the field of English-language teaching and learning. But to do so, the model must be as comprehensive in its scope and as intuitive in its application as possible. The additional attributes presented in this chapter are necessary contributions toward the goal of creating such a model.

4.1.1. Grammatical Number

As discussed in the literature review in Chapter II, the morphosyntax of the English number system provides for three grammatical numbers: singular, plural, and uncountable. Any utterance of a noun in discourse must be expressed as one of these structural types. However, precisely which type a given noun will be expressed as cannot be predicted without knowledge of that noun's lexical representation. In other words, different nouns (in their typical construals) allow different possibilities for grammatical number. A list of these varying mixes of possibilities (with examples) appears in Table 3 of Chapter II. For review, these 10 possibilities are as follows: 1. singular, plural, and uncountable, 2. only singular and plural, 3. only singular and uncountable but with historical plural *-s* ending, 4. only plural (*plurale tantum*), 5. only uncountable, 6. only singular but with historical plural *-s* ending, 7. only singular with *the*, 8. only singular with Ø, 9. only singular with Ø or plural with *the*, 10. only plural with *the*.[17]

As mentioned in Chapter II, while patterns 1-6 are associated with common nouns, patterns 7-10 all stem from the article usage associated with proper names in

[17] As mentioned in Chapter II, this list does not include nouns borrowed into English from other languages that have kept the singular-plural morphology from their original languages, such as *graffito-graffiti, radius-radii, criterion-criteria, etc.* These nouns, while complicating the English number system, follow the same article usage patterns that nouns with regular *-s* plurals do.

English. The complex patterns found with proper names were discussed in Chapter II in the section on proper names, and they will be handled with their own semantic map section in an overall model of English articles (see section 4.1.2.1, below).

Although the common noun grammatical number patterns (1-6) are well known, treatments of English grammatical number usually provide only the traditional three-way "singular vs. plural vs. uncountable" distinctions, a categorization system that is sufficient for describing the morphological marking available for any given common noun that appears in discourse (i.e., the possibility of the presence or absence of the indefinite article and the pluralizing /-*s*/). Since a model of English article use does not need to predict grammatical number but takes it as an input, accounting for or predicting the complexities of common noun grammatical number combinatorial possibilities is not necessary. In fact, for the purposes of predicting article use patterns, since we must predict only whether a noun is structurally indefinite or definite, all that we need to know about the grammatical number of any given common noun token is whether or not it is singular, since plural, *plurale tantum*, and uncountable nouns all have the same structural marking regarding articles (i.e., in the indefinite pattern, singular nouns are marked with *a*, while plural, *plurale tantum*, and uncountable nouns are all marked with Ø).

Therefore, for this study, *plurale tantum* nouns are grouped with the other plural nouns, and we use the traditional three-way "singular vs. plural vs. uncountable" system of dealing with English grammatical number. While we could specify in our model only two number categories – i.e., singular vs. non-singular – we will include the plural and uncountable categories as well simply to keep with the traditional three-way number distinction. This three-way system is more than sufficient for dealing with the indefinite

vs. definite article patterns, and it has the added advantage of being broadly familiar and therefore more accessible, especially to English language students and teachers who may benefit from the model.

4.1.2. Referentiality/Identifiability Types

In Chapter III, a number of referentiality/identifiability types were proposed and tested. They included four referential and identifiable types (Shared Lexical grounding, Shared Speech Situation grounding, Frame grounding, Current Discourse grounding), one referential and non-identifiable type (Non-identifiable ["New Reference"]), and one non-referential type (Categorization). Another six referentiality types (Proper Names, Secondary Predicates, Vocatives, Finite Verb (verb-object) "Noun Incorporation," Full (object-verb) Noun Incorporation, and Idioms) were coded in the data only so that they could be excluded from the analysis since the model and/or data was insufficient to adequately test them. In this section, these six excluded referentiality types will be discussed for inclusion in a fuller model of English article use.

4.1.2.1. Proper Names

As just mentioned, the models offered in Chapter III dealt with proper names simply by ignoring them: no explanation was attempted that could account for the inclusion of articles with proper names. However, obviously, any comprehensive account of English articles must somehow describe how articles are used with proper names – when they are marked with *the* and when they are marked with Ø or *a*. The options for (synchronically) determining this are 1.) to relegate proper names primarily to

the lexicon, where article usage patterns are determined by memory representations associated with each proper name, 2.) to find rule-like patterns based on semantic or other properties of proper names that can predict the various distribution of articles used with the full variety of proper names, or 3.) to use some combination of the above two options. Berezowski (2001) suggests that since speakers know how to use articles with hundreds of different proper names, and since they automatically know how to use them even with new proper names that they have never been exposed to before, there must be discernable patterns that speakers have access to in their mental grammars. He presents an account of the distribution of *the* and Ø that is based on the semantics of proper nouns (*a* is rarely used with proper names), and although it is not universally accurate, it does account for most instances. It should be noted that his account applies only to referential and identifiable proper names, even though there are also obvious cases of non-identifiable and non-referential proper name usage as well (see Chapter II). Recall that he argues that in English, proper names that are easily identifiable *as* proper names because of their form (e.g., "Steve") do not need the definite article to mark them as identifiable. In fact, he argues, the inclusion of the definite article would actually prompt unneeded cognitive processing since listeners would start searching for other non-identifiable instances of that reference in order to find and differentiate the unique, marked referent.

According to Berezowski (2001), proper names that are easily identifiable have the following constellation of properties: they designate single, unbounded referents that have an arbitrary name. These proper names will not be marked with any article (i.e., they will be "marked" with Ø). The referent of proper names will not be so clearly identifiable based on their form if they violate any of those properties, and such nouns

will therefore be marked with *the* to overtly specify their identifiability. Berezowski lists six ways a proper name can violate the set of easily-identifiable properties: it can have collective referent construal, it can have an unbounded referent construal, it can be an unclassified referent, it can have a descriptive name, it can have a prenominal genitive structure, and/or it can have a descriptive modifier (see Table 14 on p. 129; see also section 2.2.1.1, p. 17ff).

Since such clearly and broadly-applicable patterns of article use with proper names are discernable, it seems worthwhile to use these patterns to create a preliminary semantic map for English article use with proper names, a map which can later be tested and refined. Such a map could operate on its own or could be included as part of a broader map of article use with all nouns, both proper and common. Berezowski's (2001) account can be used to create the part of the map addressing referential, identifiable references that are marked with *the* and Ø. Non-identifiable and non-referential attributes can then be added that follow the patterns of common nouns, with references marked with *a* in the singular and Ø in the plural or uncountable. Based on these criteria, a semantic map of English article use with proper names is presented in Table 14. This table is simply a way of broadly organizing Berezowski's scheme into identifiability types and placing them on a provisional semantic map. Future work will be necessary to test and refine this map.

Table 14. *A proposed semantic map of English article use with proper names, based on Berezowski's (2001) account of identifiable proper names, plus my summary of non-identifiable and non- referential uses.*

Semantic Map of Article Use with Proper Names				**Examples**		
Proper Name Identifiability Type	**Semantics Affecting Article Use**	**SG, PL, UC(?)[a]**		**SG**	**PL**	**UC(?)**
Identifiable	Single, bounded referent, with arbitrary name	Ø, Ø, Ø		Ø Buckingham Palace, Ø Steve, Ø Mr. Bean, Ø Ben Franklin	Ø Niagara Falls	Ø Fiji Water (product name)
	Collective referent construal	the, the, the		the Cascade Range	the Franklins, the Cascades	
	Unbounded referent construal	the, the, the		the Arctic		the Old West (?)
	Unclassified referent	the, the, the		the Kremlin		
	Descriptive name	the, the, the		the White House		the Miami Heat (??) (basketball team)
	Prenominal genitive structure	the, the, the		the Cape of Good Hope	the Pirates of Penzance (?)	
	Has descriptive modifier(s)	the, the, the		the young Shakespeare		
Non-identifiable		a, Ø, Ø		a Bob that I met once	I met Ø Bobs in accounting. Five of them.	
Non-Referential		a, Ø, Ø		Strangely, he's never met a Bob.	Ø Bobs don't exist in that world.	

[a] SG=singular; PL=plural; UC=uncountable. The question mark after "UC" indicates a lack of certainty about the existence or frequency of uncountable proper nouns.

4.1.2.2. Non-specificity

In the discussion of specificity in Chapter II, an argument was presented that described specificity in terms of referentiality and identifiability. Within the kinds of references that are normally included under the scope of "specificity," some are fully specified in the sense that they are both referential and clearly identifiable to the listener. These are traditionally labeled as "specific" and/or "identifiable" references. Other references are specific in the sense that they are referential and, even though the listener cannot identify them when they are first mentioned, the speaker clearly knows their identity. These are traditionally called "specific" and/or "non-identifiable" references (since they are non-identifiable to the *listener*). A third type of reference is referential in the sense that it refers to a particular referent in the world of discourse, but neither the listener nor the speaker can identify exactly who or what that referent is when it is mentioned in the discourse. This kind of reference is usually called "non-specific." Finally, a fourth type of reference is "non-specific" and non-referential in the sense that it is not meant to refer to any actual referent in the world of discourse. This kind of reference is also sometimes (hence confusingly) called "non-specific."

Regarding this fourth type of reference, it seems unnecessary to use the term "non-specific" to account for it. Other types of non-referential mentions such as those used for a "category," "noun incorporation," or "idiom" should be able to cover all non-referential occurrences that might otherwise be labeled "non-specific." We must however still account for the third type of reference, one which is non-specific in the sense that neither the listener nor speaker can identify the precise referent even though its reference is mentioned by the speaker.

With this goal in mind, Chapter II proposed that the term "specificity" (which covers "specific" and "non-specific" mentions) is unnecessary as long as an additional type of non-identifiable mention was included in the set of "identifiability" types, and it was also mentioned that adding a new attribute to the Identifiability dimension in this way has the benefit of making our model simpler and more elegant than creating a new "Specificity" dimension. This new identifiability type would be one that covers cases where a referent is non-identifiable to both listener *and* speaker. Under this scheme, therefore, identifiability would include three types: identifiable to speaker and listener, non-identifiable to listener only, and non-identifiable to both listener and speaker. The diagram representing this scheme in Chapter II is reproduced here, in Figure 7.

Referentiality Types	
Referentiality Status	**Identifiability Status**
Referential	Identifiable to speaker and listener
	Non-identifiable to listener only
	Non-identifiable to speaker and to listener
Non-referential	

Figure 7. Doing away with the category "Specificity" by subsuming "non-specific" as a third type of identifiability.

This way of dealing with the phenomena normally regarded under the rubric of "specificity" makes the term "specificity" redundant by describing it in terms of referentiality and identifiability. In our full model of English articles, therefore, this additional identifiability attribute labeled "non-identifiable to speaker and listener" will be included in the list of Referentiality/Identifiability Types (along with Frame, Current

Discourse, etc.) in an attempt to account for all instances of so-called "non-specific" mentions.

4.1.2.3. Noun Incorporation

The models that were tested in Chapter III excluded nouns that were seen as functioning in noun incorporation constructions, but a full model of English article usage must account for them. In the summary of noun incorporation in Chapter II, we saw that full noun incorporation constructions in English involve "object-verb" or "object-verbal" syntax. That is to say, they involve moving the object of a verb or verbal item from its typical post-verbal position to a pre-verbal position. These noun-verbal constructions evoke and highlight a "unitary concept" that the whole construction is naming in a way that somehow de-emphasizes the unique, separate semantics of the individual verb and/or noun. Following Rice & Prideaux (1991), we listed six construction types that exhibit this kind of full noun incorporation. They include the following:

1) *Incorporation in finite verbs:*

   He weightlifts/weightlifted professionally. (Rice & Prideaux, 1991:3)

2) *Incorporation in infinitives:*

   He used to weightlift professionally. (Rice & Prideaux, 1991:3)

3) *Incorporation in progressive participles:*

   He's weightlifting as part of his training program. (Rice & Prideaux, 1991:3)

4) *Incorporation in participial adjectives:*

   The weightlifting competition is next. (Rice & Prideaux, 1991:3)

5) *Incorporation in gerunds:*

Weightlifting is a good complement to aerobic exercise. (Rice & Prideaux, 1991:3)

6) *Incorporation in agentives:*

He's a champion weightlifter. (Rice & Prideaux, 1991:3)

These six construction types all have the same article usage pattern: all the nouns inside the construction will be marked with Ø.

In addition, we also discussed a number of examples of finite verbs followed by non-referential objects that exhibit semantics similar to these full noun incorporation constructions. The examples from Chapter II are repeated here:

(86) He lifts/lifted Ø weights professionally. (Rice & Prideaux, 1991:3)

(87) …And.. um… the guy who is picking Ø pears [2nd mention], um… um.. picks the pears and puts them in a.. in um… these baskets that he has… (Du Bois, 1980:214)

(88) Somebody in Dullingham Junction was playing the banjo ((Du Bois, 1980:213, quoting Christophersen, 1939)

(89) Somebody in Dullingham Junction was playing a banjo (my adaptation)

(90) Somebody in Dullingham Junction was playing Ø banjo (my adaptation)

In these constructions, although the objects retain their standard post-verb position, the verb-noun semantics are highly integrated in a way that evokes a "unitary concept" similar to the noun-verb semantics found in full noun incorporation (i.e., similar to

weightlifting, *pear picking*, and *banjo playing*). This semantics is in many cases supported by the lack of an article on the noun, just as in the full, noun-verb noun incorporation constructions. However, all three article patterns – Ø, *a*, and *the* – are possible in some cases, as examples (88), (89), and (90) illustrate.

Based on the above data, the article usage pattern found with full (i.e., prototypical) object-verbal noun incorporation constructions is seen to be the Ø article pattern. Therefore, our full model of English article use will include a Noun Incorporation attribute in the non-referential section of the Referentiality dimension, and this attribute will specify Ø as the article pattern for all Discourse Modes and all grammatical Numbers. However, we also must consider some finite verb-object constructions within the noun incorporation framework, since they also participate in semantic incorporation and some elements of syntactic incorporation (such as altered article use patterns in some cases). Since in these verb-object constructions, the article use pattern is not always predictable, and since the reason governing why speakers choose one article pattern over another in these constructions is not fully understood at this point, our model will remain incomplete on this point. It will accurately predict the prototypical object-verbal noun incorporation article usage patterns as well as some but not all of the verb-object "non-prototypical" cases. Determining what governs speakers' choices for these non-prototypical cases will remain a goal of future research.

4.1.2.4. Secondary Predicates

In Chapter II it was pointed out that Du Bois (1980) lists secondary predicates and performatives as examples of non-referential mentions that use the Ø article pattern. The

function of both secondary predicates and performatives is to give names to referents or to formally place them into categories or roles. Since performatives are also structurally a type of secondary predicate and since they share the same function as secondary predicates, I have listed them both together here simply under the rubric *secondary predicates*. The data in Chapter II demonstrates that secondary predicates can and often do omit any articles before their nouns but that this is not always the case, since an article can be optionally included to accomplish a specific function of articles, such as to identify or to categorize a referent. Therefore, our full model of English article use will include a *secondary predicates* attribute in the non-referential section of the Referentiality dimension. This attribute will specify that nouns in secondary predicates usually use the Ø article pattern, a pattern that might otherwise not be predicted in the model if secondary predicates were not treated as a separate attribute with this unique specification. However, the attribute will also specify that other article use patterns are available, if the speaker chooses to utilize the functions of *the* or *a*, such as using *the* to indicate identifiability based on a frame or current discourse grounding, or using *a* to highlight singular semantics.

4.1.2.5. Vocatives

As discussed in Chapter II, vocatives are a special case of non-referential mentions. Although they resemble referential mentions in some ways – particularly in that when they are used, the referent which they index can recognize that he or she is being addressed – in other ways they differ from referential mentions. First, their primary purpose is not to identify a referent or to trace that identity through a discourse,

which *is* the primary function of referential mentions (Du Bois, 1980). Second, while referential nouns can be used as arguments, vocatives cannot be (Anderson, 2007). Their non-referentiality status might also be reflected in their use of the Ø article pattern, since if identifiability status were a key function of vocatives, one might expect the definite and indefinite articles to be options. For these reasons, vocatives will be placed in our model as an attribute in the non-referential section of our Referentiality dimension.

4.1.2.6. Compound Modification

Nouns that are involved in compound modification (i.e., modification of other nouns) behave like adjectives. The only article option that is available to them is Ø, as seen in (91)-(93) (repeated from Table 1 in Chapter II):

(91) [a [pear] tree]

(92) *[a [a pear] tree]

(93) *[[a pear] a tree]

Although they can be treated as adjectives, there are some practical reasons to include them with the nouns in a full model of English articles. First, this adjectival function and the accompanying adjectival morpho-syntactical forms of these nouns (e.g., the Ø article pattern and the lack of plural forms) are not always obvious insights to learners of English as a second language. Seeing their adjectival function clearly labeled on a model might serve to clarify their function and structure – or at least their necessary lack of articles. Second, although they can be treated as adjectives, they also could be listed as nouns since the relevant lexemes are used as nouns in other (and most) contexts. Including them and their Ø article use in a model of English articles might serve the

simple purpose of comprehensiveness. For these reasons, they will be listed as an attribute on the non-referential section of the Referentiality dimension of our model.

4.1.3. Discourse Modes

The models in Chapter III tested only a few different discourse modes. One model tested a "Normal" and a "Reintroducing" mode, modes that were inspired from ideas in Du Bois (1980) but which I adapted to this model. Another model tested three discourse modes that were very closely based on modes proposed by Du Bois (1980): his "Normal Narrative Mode," his "Normal Descriptive Mode," and his "Deferred Descriptive Mode." All of these modes were described in detail in Chapters II and III. Since the results in Chapter III indicated that the model containing the "Normal" vs. "Reintroducing" discourse modes was the most accurate model tested, those two discourse modes – rather than the three modes in the other model – will be the ones included in our full model of English article use. In addition to those modes, two other modes that were not tested should arguably be included in the full model: the Immediacy Discourse Mode and a Headline Discourse Mode.

4.1.3.1. Immediacy Discourse Mode

As detailed in Chapter II, the Immediacy Discourse Mode (my label, but described by Du Bois 1980), is used to instill a sense of "immediacy" or vividness in a narrative by acting as if referents are identifiable even when they are first mentioned and thereby are actually *not* identifiable. This mode is therefore different from the Normal

mode in that it uses the definite pattern for new references where the Normal mode uses the indefinite pattern.

4.1.3.2. Headline Discourse Mode

The other mode that will be included is a Headline Discourse Mode. This is a special mode used in headlines and public signs. It is noted for its frequent use of the Ø article pattern where another pattern would be expected in other modes or genres. However, even though it is more common in headlines than in other types of discourse, the Ø pattern is by no means used in all headline cases. Mårdh (1980) lists a number of factors that can influence whether articles are dropped or retained in headlines. These factors include the following:

- typographic considerations (the need to maximize, minimize, or balance the space a headline takes up on the page),
- the type of headline (ungrammatical zero is less common in connotative headlines, which give only one sensational aspect of the news story in order to arouse curiosity, than in summary headlines),
- speech quotations (ungrammatical zero is less common in a direct or approximate quotations within a headline),
- contextual considerations (linguistic context, pictures that accompany the headline, and common knowledge of referents [e.g., the president, the queen] affect article use),

- position of noun phrases (a 1st noun with an article will often trigger an article on the 2nd noun, a 1st noun with Ø article will optionally allow an article on the 2nd noun),
- the communicative value of the articles (articles are used to distinguish coordinated vs. non-coordinated noun phrases, to distinguish literal vs. non-literal meanings, and to distinguish word class [e.g., noun vs. verb] in ambiguous cases),
- degree of common familiarity of proper names (well-established proper names such as *the White House* tend to retain their article pattern more than less-well-established ones), and
- acronyms (articles are usually not used with acronyms or initialisms).

Because the use of articles in headlines depends on so many factors and is in fact variable, creating an accurate model of headline article use demands a study unto itself. Although performing such a project is beyond the scope of this current study, it is clear that headlines are a special discourse mode with unique article usage patterns and therefore should be at least mentioned in our full model of English article usage. Since the most common article pattern in headlines is the Ø pattern, our full model will include a Headline Discourse Mode that indicates the Ø pattern as being prevalent. It will also include a note stating that nouns in Headline Mode may also retain their Normal Discourse Mode articles in many cases. Since Mårdh (1980) offers only a list of factors that can affect article use in headlines but not a comprehensive explanation of the wide variations in their application, a full accounting of article use in Headline Mode remains a goal of future study.

4.1.3.3. Reintroducing Mode and the excluded discourse modes

Our model provides a refinement of the system of discourse modes that was proposed in Du Bois (1980). In so doing, it proposes a new mode as well as excludes some of Du Bois' original modes. As mentioned above, Du Bois' Normal Descriptive and Deferred Descriptive Modes are not included in the model because the testing of the models in Chapter III revealed that the functions of these two modes can be accomplished by the functions of the Reintroducing Mode proposed and tested in Chapter III. In addition, it appears that another of Du Bois' modes – the Defining Descriptive Mode – can also be handled by the Reintroducing Mode and therefore does not appear in our final, full model. A reiteration of the function of the Reintroducing Mode will help to explain why it proves so adept at replacing three other modes.

The basic function of the Reintroducing Discourse Mode is to allow speakers to clarify a real, perceived, possible, or posited misunderstanding or lack of understanding concerning a discourse referent. To accomplish this clarifying goal, the speaker uses the indefinite article pattern, where the definite article pattern might otherwise be used, when a referent has either Frame or Current Discourse referentiality grounding. Basically, by using the Reintroducing Mode, speakers treat Frame or Current Discourse referents (which "should" normally already be identifiable under the Normal Mode), as if they were instead Non-identifiable ("New References").

In the Normal Mode (i.e., the Normal Narrative Mode), referents that have either Frame or Current Discourse grounding are marked with the definite article pattern. However, in the Normal Descriptive Mode described by Du Bois (1980), referents that have Frame grounding are marked with the indefinite pattern. Likewise, in both the

Deferred and Defining Discourse Modes, referents that have Current Discourse grounding also are marked with the indefinite pattern. Therefore, structurally, the Reintroducing Mode produces the article patterns that otherwise require the invocation of one of these other three modes.

In addition, the Reintroducing Mode also accommodates the functions that these other three modes were posited to accomplish. First, the Normal Descriptive Mode is designed to handle situations where a speaker pauses in a narrative to give a description of a scene, object, character, etc. As just mentioned, the article usage pattern for this mode is identical to that of the Normal (Narrative) Mode except that when a referent has Frame grounding in the Normal Descriptive Mode, it assumes the indefinite pattern. The postulated purpose of this switch is to allow the speaker to signal that something that "should" normally be identifiable based on Frame grounding in fact is *not* identifiable for some reason and instead requires it to be treated as if it were a New Reference. This purpose is also accomplished by the Reintroducing Mode.

Second, the Deferred Descriptive Mode is used when the momentum of the narrative causes the speaker to make an initial mention of something without giving an adequate explanation of it. She therefore switches to this mode in order to "go back" and "officially" introduce it with a technically second mention, but using the indefinite pattern where normally (in the Normal Narrative Mode) it would receive the definite pattern. Again, this function is also accomplished by the Reintroducing Mode.

Third, although the Defining Descriptive Mode was not tested in the models in Chapter III due to lack of data in the elicited corpus, it appears that it, too, can be replaced by the Reintroducing Mode. As stated in Chapter II, the Defining Descriptive

Mode is used when a definition is required in a discourse. When a word is being defined after it has already been mentioned, the subsequent defining words obviously refer to that same word and concept. However, whereas in the Normal Narrative and Normal Descriptive Modes subsequent references have a definite marking, in the Defining Descriptive Mode, these defining noun phrases receive an indefinite marking instead. If the speaker is thought of as mentioning a new term and then pausing or "going back" to formally define or "introduce" it with another noun phrase, the function of the Defining Descriptive Mode can be seen to be functionally equivalent to the Reintroducing Mode.

The Reintroducing Mode, however, has an advantage over the Defining Mode in that it can optionally apply. Whereas the "Defining" Mode, as such, would seem more or less to *require* the speaker to use it whenever a definition is given in discourse, the Reintroducing Mode allows the speaker the *choice* of whether or not to "reintroduce" a term and treat it as if it is new. This speaker-based choice can therefore produce either the indefinite or the definite article pattern on these subsequent defining noun phrases, a capability that can account for variations in data such as those seen in items (94) and (95):

(94) Ø Kutchin [1st mention], *an* Athabaskan language of Alaska [2nd mention], possesses no less than 55 consonantal "phonemes" [1st mention], Ø distinct consonantal elements of the total phonetic pattern [2nd mention]. (Du Bois 1980:231, quoting Sapir, 1929:140)

(95) Ø Kutchin [1st mention], *the* Athabaskan language of Alaska [2nd mention], possesses no less than 55 consonantal "phonemes" [1st mention],

the distinct consonantal elements of the total phonetic pattern [2nd mention]. (my adaptation of Du Bois 1980:231, quoting Sapir, 1929:140)

Whereas the Defining Descriptive Mode (as a mode specifically designed to handle "definitions") would seemingly necessarily require the indefinite pattern on the 2nd mentions in items (94) and (95), the Reintroducing Mode could be either chosen or not chosen, depending on the speaker's assessment of the listener's level of familiarity with the idea involved. The Reintroducing Mode could thus allow the indefinite pattern on 2nd mentions seen in (94) (if the Reintroducing Mode is chosen because the speaker thinks the listener is unfamiliar with the idea that there are Athabaskan languages in Alaska, for example). And it could also allow the definite pattern seen in (95) (if the Reintroducing Mode is not chosen and therefore the Normal Mode continues to apply because the speaker thinks the listener *is* familiar with the idea that there are Athabaskan languages in Alaska, for example). Since the use of the Reintroducing Mode can thereby not only account for the article use patterns produced by the Defining Descriptive Mode but also allow for alternate patterns that the Defining Descriptive Mode would incorrectly not allow, the Defining Descriptive Mode – along with the Normal Descriptive Mode and the Deferred Descriptive Mode – will not be included in our full model of English article use. All three will be replaced by the Reintroducing Mode.

4.1.4. Constructions stored in the Lexicon

As mentioned in Chapter II, a final type of non-referential nouns are the nouns in *idioms*. The article usage patterns on idioms are lexically determined rather than functionally determined, and as such, they are unpredictable, and they must be retrieved

from memory. The nouns in idioms are non-referential since they do not refer to any "real" entity in the world of discourse, so we will include them in our full model of English article usage. However, since the articles in idioms are based on memorized patterns, no predictions for article use will be made on the model, and a note will be included specifying that the patterns are lexically determined.

4.2. A proposed comprehensive semantic-map-based model of English articles

Our full model of English article use is presented below in Figure 8. As has been discussed throughout this dissertation, this model is based on a semantic map approach to referentiality (including identifiability), grammatical number, and discourse modes. Although the particular attributes on the dimensions of these maps are specific to English, many of them may also be included on a cross-linguistic (universal) conceptual space. For example, on any map of universal conceptual space of identifiability, we would expect to find listener-oriented identifiability based on grounding in the shared speech situation, in conceptual frames, and in the current discourse. Furthermore, the particular dimensions that have been assembled together to create this semantic map are also specific to English articles. Maps of articles in other languages may consist of other constellations of dimensions. For example, even if maps of articles for other languages were to include a referentiality dimension, discourse mode and/or grammatical number may or may not be implicated functional elements for those systems.

As in the models tested in Chapter III, this full model consists of three dimensions. Graphically, the Discourse Mode dimension is arranged horizontally with four discourse modes: the Headline, Immediacy, Normal, and Reintroducing Discourse Modes. In

order to squeeze three dimensions into a two-dimensional rendering, within each Discourse Mode column is nestled the Grammatical Number dimension so that "SG, PL, UC" (i.e., Singular, Plural, Uncountable) can be seen repeated in each column. The Referentiality dimension is arranged vertically. The Referentiality Types consist of Referential and Non-referential types, and the Referential types consist of various Identifiability types. Each of these referential/identifiability types and non-referential types are further divided into various descriptive attribution labels (e.g., Proper name, Shared lexis…, Idioms).

In order to use the model, a noun phrase in a spoken or written discourse must be located on the model according to its current Discourse Mode, Grammatical Number, and Referentiality/Identifiability Type. Once the proper attribute for each of these three dimensions is determined, the predicted article for that reference can be located at the intersection of the dimensions in the chart. The strengths and weaknesses of this model and its applications will be discussed in the following chapter.

Figure 8 (next page). A proposed comprehensive model of English articles. (Note: SG=singular; PL=plural; UC=uncountable.

Referentiality Dimension			Discourse Modes Dimension			
Referentiality Type	**Identifiability Type**		**Headline Discourse Mode**[1]	**Immediacy Discourse Mode**	**Normal Discourse Mode**	**Reintroducing Discourse Mode**
			Grammatical Number Dimseenion			
Referential	Identifiable to Listener & Speaker	**Source of Listener's Identifiability**	SG, PL, UC	SG, PL, UC	SG, PL, UC	SG, PL, UC
		Proper name[2]	Ø/the, Ø/the, Ø/the	Ø/the, Ø/the, Ø/the	Ø/the, Ø/the, Ø/the	Ø/the, Ø/the, Ø/the
		Shared lexical understanding	Ø, Ø, Ø	the, the, the	the, the, the	the, the, the[3]
		Shared speech situation	Ø, Ø, Ø	the, the, the	the, the, the	the, the, the[3]
		Frame	Ø, Ø, Ø	the, the, the	the, the, the	a, Ø, Ø
		Current discourse	Ø, Ø, Ø	the, the, the	the, the, the	a, Ø, Ø
	Identifiable to Speaker only ("New reference")		Ø, Ø, Ø	the, the, the	a, Ø, Ø	a, Ø, Ø
	Identifiable to neither Speaker nor Listener ("Non-specific reference")		Ø, Ø, Ø	the, the, the	a, Ø, Ø	a, Ø, Ø
Non-referential	Non-referential constructions					
	Categorization		Ø, Ø, Ø	a, Ø, Ø (the, the, Ø)[4]	a, Ø, Ø (the, the, Ø)[4]	a, Ø, Ø (the, the, Ø)[4]
	General non-referential expressions -- includes constructions such as Compound Modification (i.e., Adjectives), Secondary predicates[5], Vocatives, and Full (object-verb) Noun Incorporation		Ø, Ø, Ø	Ø, Ø, Ø	Ø, Ø, Ø	Ø, Ø, Ø
	Finite Verb (verb-object) "Noun Incorporation"		a/the/Ø, Ø/the, Ø/the	a/the/Ø, Ø/the, Ø/the	a/the/Ø, Ø/the, Ø/the	a/the/Ø, Ø/the, Ø/the
	Idioms		a/the/Ø, Ø/the, Ø/the	a/the/Ø, Ø/the, Ø/the	a/the/Ø, Ø/the, Ø/the	a/the/Ø, Ø/the, Ø/the

[1]Although the Ø article pattern is the most common pattern in the Headline Discourse Mode, nouns in this mode may also retain their Normal Discourse Mode articles in many cases. A full accounting of article use in Headline Mode remains a goal of future study.

[2]See semantic map for Proper Names in Table 14 on p. 129 for details on when to use Ø vs. *the*.

[3]In the Reintroducing Discourse Mode, it is possible to use the indefinite "*a, Ø, Ø*" pattern to "reintroduce" a referent that has Shared Lexical or Shared Speech Situation grounding, but it is probably highly unlikely that a speaker would do so in most discourse contexts.

[4]Categorization can also sometimes be achieved through the definite pattern, as in the case of the generic *The dog is a vigilant animal* (Christophersen, 1939) or *The great apes are incredibly adaptable*.

[5]Secondary predicates have the option of using Ø where this might otherwise not be an option. However, either *a* or *the* may also be used if the speaker wishes to utilize one of their pragmatic functions.

CHAPTER V

CONCLUSION

5.1. Summary of the model

This dissertation has described a model of the functions of the Standard American English article system. This model consists of a large, 3-dimensional semantic map that is composed of the intersection of three parameters or "dimensions": referentiality, discourse mode, and grammatical number. These three "dimensions" are the main functional domains found in the literature to be implicit in the use of English articles.

Within each dimension are a several "attributes" that describe specific functions. The *referentiality* dimension describes different ways that a noun can be referring or non-referring in a discourse, and it is therefore divided into two subparts: a *referential* section and a *non-referential* section. The referential section is further subdivided into three *identifiability* types, each containing one or more attributes: the *identifiable to both listener and speaker* section (with 5 sub- attributes: *proper name, shared lexis, shared speech situation, frame,* and *current discourse*), the *identifiable to the speaker only ("new reference")* attribute, and the *identifiable to neither speaker nor listener (non-specific reference)* attribute. The *non-referential* section is divided into 4 attributes (*categorization, general non-referential expressions* [which include constructions such as secondary predicates, vocatives, compound modification [i.e., adjectives], full [object-verb] noun incorporation], *finite verb [verb-object] "noun incorporation",* and *idioms.*)

The *discourse mode* dimension consists of 4 attributes: the *headline, immediacy, normal,* and *reintroducing* discourse modes. These discourse modes represent different creative, tactical ways that a speaker can employ a noun phrase to advance the goals of a discourse, such as to give a sense of "immediacy" to a story so as to place a reader directly into a new scene, as if she were already in the middle of the action (the immediacy mode); to clarify an obvious or potential misunderstanding (the reintroducing mode); or simply to advance a narrative in the standard, usual manner (the normal mode).

The *grammatical number* dimension consists of 3 attributes: *singular, plural,* and *uncountable*. Although describing the English article system really only requires knowledge of whether a noun is singular or non-singular, the plural and uncountable categories (which have identical article usage patterns) are included in the model simply to accommodate the traditional tripartite number distinction that most interested parties are familiar with, including especially teachers and learners of English as a second language.

5.2. Reflection on the methodology

Parts of the overall model were tested using a methodological paradigm that allows for an experimenter to influence (and therefore loosely control for) the discourse goals in the minds of speakers. Influencing the discourse goals of the speaker was attained by having a collaborator ask the speaker to answer specific questions, such as at various times to describe an action, to describe a scene, and to re-describe those actions or scenes. It was necessary for the analysis to surmise what was in the mind of the speaker since the model depends on such information. Eliciting an answer from a specific

question was chosen as an efficient and direct way to achieve this purpose of creating specific types of speaker goals. For example, in order to test the discourse modes proposed by Du Bois (1980), an experiment should manipulate whether the speaker tries to speak about the action of the story or to describe the physical setting of a scene. Only with this information could an experimenter confidently categorize the use of a noun as being in the "normal discourse mode" vs. the "descriptive discourse mode."

In the end, this methodology appeared relatively successful in facilitating the goals of knowing what the speaker was thinking. A weakness, however, lay in fact that speakers did not always clearly answer the question that was asked of them. Although sometimes they *did* clearly answer the questions, other times they would seem to ignore the question altogether and proceed with the previous momentum of their story, or they would begin to answer the question but quickly resume discussion of a previous topic or move on to a new topic. Although this problem did result in some ambiguity in the data, the drawbacks of the methodology were largely offset by its ability to elicit very naturalistic discourse from the speakers. This resulting naturalistic data speaks well for the model's ability to handle prediction of article usage in actual real-world data as well.

Furthermore, even with this addition of ambiguity in some instances, another benefit of the methodology was that it *did* produce much data that clearly matched the sought-after type. It is certainly the case that this methodology did produce much of the kind of data that it was designed to produce – i.e., data that corresponded to each of the tested discourse types. Indeed, it is likely that much more data of the desired type was obtained with this method than could have been obtained using more traditional methods such as mining pre-existing corpora. In fact, adequately testing these models with a pre-

existing corpus would have been very difficult since the intention of the speaker could not have been known with much confidence and would have had to be based primarily on the experimenter's making uncertain inferences from what the speaker said. In effect, the methodology used here produced a small but targeted (and thereby uniquely useful) corpus of data to allow testing of specific variables in our model.

The data collection and the results obtained by use of this paradigm demonstrate its utility and indicate that it could be usefully adapted to obtain data to test other linguistic models where the knowledge or intent of the speaker must serve as independent variables.

5.3. Evaluation of the semantic maps in this study

As discussed in detail above, the model presented here is essentially a large semantic map that is composed of three "dimensions," one for each of the three functional categories that are implicit in English article usage (*referentiality, discourse mode,* and *grammatical number*). Since semantic maps organize and make testable hypotheses about form-function mappings, one of the benefits of using them is that they can be adapted for use cross-linguistically as a tool for typological research. Any given semantic map for any set of functions in any particular language will likely contain only a portion of the broader conceptual space that represents the sum total of possible functions that can obtain in human cognition. The semantic maps presented here are an attempt to describe only those functions expressed by articles in Standard American English. A short discussion of typological applicability of the various parts of the overall semantic map is therefore in order.

First, in our model the English number system involves number and countability, treating these two concepts as a single issue. Although this kind of semantic map works for English, it may be prove typologically useful to posit two separate but interacting conceptual spaces for grammatical countability and grammatical number, since conceptually they can be accounted for separately from one another and since neither is expressed in all languages. In English, for example, we could posit a countability semantic map (with countable and uncountable attributes), and a number semantic map (with singular and plural attributes) that is nested within the countable attribute. No matter how a typologically universal conceptual space for these functions is designed, the *number* categories in our model (*singular, plural,* and *uncountable*) would certainly require a place in it (along with, of course, whatever other number categories that might be expressed in other language, such as *dual, trial, paucal*, etc.).

Second, in the *referentiality* dimension, as the literature attests, the general division between *referential* and *non-referential* mentions finds broad application cross-linguistically, since humans everywhere find need to make these distinctions. Likewise, the *identifiability* portion of the referentiality dimension contains seven "attributes" or functional categories that will likely be distinguished and grammatically expressed across many languages, since tracking discourse referents is a general need in human communication. Furthermore, the *Non-referential* attributes of *Categorization* and *General Non-referential Expressions* are likely to be applicable cross linguistically. However, even though noun incorporation can be seen as a function that serves to highlight the unification of a noun and verb into one concept, the *Finite Verb (verb-object) "Noun Incorporation"* portion of the model does not yet explain why speakers

sometimes unify a noun and verb more "strongly" in the expressed form by dropping the articles but other times retain the articles. Understanding and explaining this is a goal of future research.

Third, the *discourse modes* section of our model comprises some "attributes" or functions that may indeed belong in a typologically relevant map of conceptual space. In particular, the normal discourse mode and the reintroducing discourse mode represent two tactics that are likely broadly employed cross-linguistically to accomplish discourse goals: telling a narrative and clearing up potential or actual misunderstandings. Another discourse mode – the so-called "immediacy" discourse mode – also might prove relevant in some other languages, since it represents a tactic for drawing a listener in to a story early on, something that human storytellers generally might wish to do and might therefore have special constructions for accomplishing. The final discourse mode in our model – the "headline" discourse mode – is perhaps or more rare relevance. Its purpose is to provide a very short title introducing the contents of a larger text, a common-enough necessity. However, whether using special grammatical constructions to do this is common in world languages (and if so, in what circumstances or contexts they are used) is a matter for future research.

One additional concern raised by our model is whether or not positing a separate dimension or semantic map for *discourse modes* is justified, since it is conceivable that one could propose additional *identifiability* types to account for the same phenomena. As mentioned before, the idea of including discourse modes as an important feature of the English article system originated with Du Bois (1980), who proposed them in order to account for variations in patterns of article use that otherwise seemed unmotivated. For

example, the addition of discourse modes to the overall model handles cases where a speaker clearly knows a referent is identifiable to the listener but acts (based on her use of articles) as though it is not (as in example (97), below), and, conversely, cases where a speaker clearly knows a referent is *not* identifiable to the listener but acts (again, based on her use of articles) as though it *is* (as in example (96), below).

(96) the first… thing I noticed.. was.. the sound of <u>*the man*</u> [1st mention] picking… pears.

(97) …And of course there was a… *a man* [2nd mention] there standing on a ladder in a pear tree,

On the one hand, these article use patterns might instead be accounted for by positing more identifiability types. However, motivating them as real cases of *identifying referents* seems difficult, since the referents' identifiability status is, in fact, already known. On the other hand, the idea that speakers use different types of *discourse modes* can be more easily motivated by pragmatic functions within a discourse, such as a speaker's desire to clarify a referent's identity (by "reintroducing" it, as in example (97)) or to give a sense of "immediacy" to a narrative (as in example (96)). These kinds of pragmatic functions seem much more like different ways to accomplish discourse goals rather than merely additional means to identify referents. For these reasons, the idea of including discourse mode as an additional dimension in our overall model of seems justified.

5.4. Application of the model

There are a number of ways that the model presented here can be applied. First, it can be used as a reference tool for researchers when they are exploring theoretical issues related to English articles and to language teachers and learners when they are trying to understand article use. Second, various aspects of the model can be incorporated in language curricula to facilitate the relatively early acquisition of articles by English language learners. These two types of applications of the model will be discussed in turn.

5.4.1. As a reference tool for research and language learning

This is speaker-based model of English article use. As such, in order apply the model to a text, one has to know much of what is in the mind of the speaker. That is to say, one has to know what semantic notion she is trying to communicate and what pragmatic intention she is trying to accomplish when she mentions a referent. Therefore, the model can be applied by speakers themselves when they are trying to decide which articles to use in their own discourse. Obviously native speakers have no need to refer to a published model for "assistance" in order to use articles. However, the model might prove quite useful to non-native speakers of English when they are crafting a discourse. As long as they have a clear idea about the number, referentiality type, and discourse mode they intend when they use a noun, they can check the model when they are in doubt about which article to use. Furthermore, even though native speakers do not need it when composing their own speech or writing, they might also benefit from the model when they are trying to understand the theory and use of articles for such purposes as teaching English or researching discourse grammar and pragmatics.

5.4.2. As a guide for incorporating instruction on articles in ESL course curricula

As just discussed, to the extent that non-native speakers of English understand the dimensions and attributes of the model, they can refer to the model when they are in doubt about which article to use in a given situation. However, in order for them to use articles in a native-like manner without constantly referring to the model, they must acquire competence in the areas that the model uses as inputs: Referentiality/Identifiability, Discourse Modes, and Grammatical Number. Acquiring completely native-like competence in all these areas is a lofty goal and a very difficult task, but one way that such acquisition might be facilitated is by incorporating the most often-used attributes of the model in ESL course curricula.

5.4.2.1. Incorporating attributes of the Referentiality/Identifiability dimension

In the results from the experiment described in Chapter III, it was found that the most often-occurring attributes in the Identifiability portion of the Referentiality dimension were the following (starting with the most frequent): Current Discourse, Non-identifiable ("New Reference"), Proper Names, Frame, and Shared Speech Situation. The use of these five sources of Identifiability (or lack of identifiability, in the case of the Non-identifiable type) could be taught in graduated fashion throughout the multi-level courses of an ESL program, both in textbooks and in classroom curricula.

5.4.2.1.1. Non-identifiable ("New Reference") vs. Current Discourse

The two most frequent attributes from the model as measured in our data happen to coincide with the two most frequent that are already most commonly taught: Non-

identifiable ("New Reference") vs. identifiable from the Current Discourse. In our results, the most simplistic model – which used only a simplified implementation of these two attributes (the Initial vs. Subsequent Mention model) – still was able to accurately predict 78% of the article use in the corpus. If English language learners mastered only the simple distinction in article use between initial vs. subsequent mention of referents, they would be able to use articles correctly in the majority of cases. There is no reason that this kind of instruction about articles and referent identifiability should not be continued; indeed, it might be beneficial to begin teaching it even earlier than it often is, perhaps even from the very earliest levels of English learning (in the first or second level of an intensive English program, for example) so that learners can early on begin to acquire discourse competence regarding the basic and most frequent use of articles.

5.4.2.1.2. Proper Names

The next most common Identifiability attribute in our model was Proper Names. English learners should be instructed early on to memorize the article use pattern for proper names that they learn. Although we presented a semantic map that attempts to account for the presence or absence of *the* with proper names (Table 14 on p. 129), it is too complicated to be very useful to early-stage learners. As long as learners are aware that some proper names always use *the* while other never use *the*, they can focus on simply memorizing the articles with the proper names that they learn.

5.4.2.1.3. Frames

Frames, which was the fourth most-common Identifiability attribute in our model, provide a very important additional source of referent identifiability in discourse. As such, they should be taught sometime after learners have begun to acquire some level of mastery of the Initial vs. Subsequent Mention distinction, perhaps in the second or third level of courses in an intensive English program. The basic idea of how frames work is perhaps simple enough – that the mention of a referent by a speaker can prime the listener to accept a number of related ideas as identifiable within the broad context of the original referent. For example, the mention of *restaurant* can allow the listener to accept *waiter* and *menu* as identifiable within that frame, as in item (98) (repeated from Chapters I and III):

(98) She walked into a restaurant and asked the waiter for the menu.

(Givón, 2005)

However, even though the basic concept of frames is somewhat simple, implementing it in a native-like way requires vast cultural knowledge. For example, a speaker (and listener) must know all the referents that can be assumed by native speakers to be identifiable in the *restaurant* frame. Acquiring such cultural knowledge for all the possible referents in a language is a very long-term project for any learner, and therefore ESL learners should not be expected to master article use for Frame-based identifiability quickly. However, exposing them to the basic operation of Frame-based identifiability relatively early in their language learning experience can give them an awareness of how this aspect of the article system works so that they can begin to notice how cultural

frames are constructed. This early awareness should allow them to acquire a better mastery of frames and article use than if they never received such instruction.

5.4.2.1.4. Shared Speech Situation

Although Shared Speech Situation is only the fifth most common Identifiability attribute in our model, it is such a salient and basic source of identifiability that it could also be introduced early on in English language instruction. The fact that it occurs in real-life shared contexts also should make it easy to incorporate Shared Speech Situation article instruction into lesson plans that are accessible to early learners, possibly teaching simple discourse competence along with vocabulary in task-based lessons. The following items illustrate interactions that might happen in such lessons:

(99) Hand me *the* hammer.
Now hand me *the* nail.

(100) Hand me *a* hammer. [where there are two hammers on the table]
No, hand me *the* other hammer.

5.4.2.1.5. Categorization

In addition to Referential Identifiability uses, the use of articles to mark Non-referential NPs can also be taught in graduated fashion through the levels of English language learning. Categorization was the most commonly used Non-referential attribute in our model, and as such, it should be taught early on, perhaps at the very beginning along with or shortly after Initial vs. Subsequent Mentions in order highlight the difference between referring and non-referring expressions. A good place to introduce

Categorization in a curriculum would be where irrealis verbs such as *want* or negative expressions such as *is not* are being practiced, as in the following examples:

(101) I want *a new car*.

(102) He is not *a student*.

5.4.2.1.6. Idioms

Finally, as with Proper Names, learners should be instructed at some point that article use with Idioms also must be memorized, since the article use patterns do not change within idioms (as illustrated in items (54) and (55) on p. 48). Simply being aware of this fact may help learners learn to treat idioms as special cases and to avoid overgeneralizing the application of the model's Referentiality/Identifiability attributes to idioms.

5.4.2.2. Incorporating attributes of the Discourse Mode dimension

In order to facilitate acquisition of the article system, the use of two of the Discourse Modes in our model also should be included in English language curricula. Since the Normal discourse mode is by far the most commonly used mode, the attributes of the Referentiality/Identifiability dimension as they are expressed in the Normal discourse mode should be taught first (as just discussed in section 5.4.2.1). After learners have begun to acquire competence in using the Normal mode, the Reintroducing mode can be introduced in the curriculum. Since the basic function of the Reintroducing mode is simply to re-explain a previously-established Identifiable NPs (in order to clarify a real, perceived, potential, or posited misunderstanding or lack of understanding) and the basic

way to use the article system is simply to temporarily "reset" it to the indefinite pattern, explaining its operation in a way that learners can understand should be somewhat straightforward. Although speakers may not use the Reintroducing mode very often, when they wish to accomplish its "re-explaining" function, they will find its implementation very useful for communication, and therefore including it somewhere in the middle or late stages of the instructional curriculum of an English language program should serve the interests of learners well.

5.4.2.3. Incorporating Grammatical Number as a normal part of ESL vocabulary acquisition

In our model, Grammatical Number (which in our treatment includes countability) is prominently placed as one of the three dimensions determining article use, highlighting the idea that number/countability must be an input to the system that facilitates article choice. This kind of input information derives in part from the speaker-intended number of each noun she uses (e.g., whether she wishes to speak about a singular vs. a plural referent), but it is also influenced greatly by the lexical representation of those nouns, since some nouns are typically countable and some uncountable. The fact that number/countability information is cognitively represented has implications for teaching and learning English. Native English speakers have rich representations relating to number/countability for each lexeme in their vocabulary. For example, they have representations not only about whether a particular noun is normally countable or uncountable but also about how often and in which circumstances it can switch categories from one to the other. Native speaking listeners also use these representations to easily

notice any variations from typical usage and thereby to understand creative changes in the speaker's meaning. The frequency-based and context-relevant information in these representations enable native speakers to exploit the nuances of articles by forcing nouns into unusual number/countability patterns for purposes of subtle meaning manipulations (e.g., for playful communication and humor, creative expression, novel applications). When someone learns English as a second language, however, this kind of rich representation of grammatical number and countability is not normally part of their newly-acquired knowledge of the language. Even if they learn that a noun is normally countable or uncountable, they may not know how or how readily that noun can switch categories, and they may therefore be unable to produce or comprehend the subtleties changes of meaning that native-like speakers can achieve through the nuanced manipulation of article selection. Therefore, by prominently including number and countability as one of the three main dimensions in the model, this model highlights the importance of understanding the lexical semantics that influence the number/countability inputs to the model. This in turn has implications for teaching and learning English as a second language. In particular, it emphasizes the importance of the idea that learners of English will benefit from a relatively rich set of information about the countability of each English noun. One suggestion for facilitating such learning would be for English language instructional approaches to include the teaching of prototypical countability information (based on frequency of usage) for each new noun vocabulary item. This could be presented with each new noun in a way similar to how grammatical gender information is taught with each new noun in Romance language classes and textbooks.

5.5. Future directions

Although every attempt was made to create a model that was accurate and comprehensive, refinements in a number of areas could improve it further. First, as mentioned previously, because of lack of extant research regarding the non-referential functions of English articles, the attributes of the non-referential portion of the semantic map are currently a mix of functions and constructions. Finding a way to describe them all with a series of distinct but related cognitive functions would render the map – and its underlying conceptual space – more elegant and more useful for more widespread use in typological research. This would likely require an entire study devoted to understanding non-referential uses.

Second, the discourse mode "dimension" or semantic map currently includes four different discourse modes. These modes derive mostly from previous work by Du Bois (1980), but little attention has been given since then to how discourse modes affect article use. It might be the case that there are additional modes that are used as tactics to accomplish discourse goals that have yet to be notice or described. Future work into the nature and extent of discourse modes might reveal any additional modes and help better define the current modes. The "headline" discourse mode, in particular, is in need of research to discover precisely how various factors affect article use in headlines.

Third, the proper names semantic map has been proposed here based on work by Berezowski (2001), but it has not been tested. Testing and refining it is also a goal for future research.

Lastly, the nature of article use in secondary predicates and so-called "finite verb 'noun incorporation'" is only poorly understood at this stage (see Chapter IV). Future

work to understand what factors determine various article use patterns in these constructions is a necessary step toward making the model truly comprehensive.

Even in its current state, it is hoped that the model will offer a useful contribution to researchers interested in functions related to article use and will provide teachers and learners of English as a second language a comprehensive and intuitively accessible model of article use where none existed previously.

APPENDIX

SAMPLE OF ELICITED TEXT AND CODING

This appendix contains a sample of the transcribed conversations that were elicited as described in Chapter III. The first column, labeled "Line," contains a reference number that begins with the number of the conversation session followed by a decimal point followed by the line number for that conversation (for example, 2.013 refers to conversation 2, line 13). The text can be read from left to right, top to bottom, across the four columns labeled "Context," "Other Determiner/Genitive," "Article (hidden)," and "NP" (i.e., the columns in between the two sets of double vertical border lines). The articles that were actually spoken by the watchers of the video have been hidden in the column labeled "Article (hidden)," just as they were hidden when the data were coded. The next five columns – "Sequence of Mentions," "Number," "Referentiality/Identifiability," "2 Discourse Mode Model," and "3 Discourse Mode Model" – contain the coding fields that were described in Chapter III. To see what articles were actually spoken before the NP in each line, see the final column labeled "Articles (revealed)." The sample texts contain examples of all (or most) Grammatical Number, Referentiality/Identifiability Types, and Discourse Modes for both the 3 Discourse Mode Model and the 2 Discourse Mode Model, including all of the non-excluded Reintroducing mode items that were included in the analysis and results in Chapter III. It should be noted that only some (but not all) of the examples listed in the main text of the dissertation are reproduced here in the appendix.

In the column labeled "Context," the speakers in each conversation are labeled with a notation that begins with Q for Questioner or W for Watcher, followed by the conversation number, followed by F for female or M for male. For example, Q2M refers to Questioner 2 (Male), and W3F refers to Watcher 4 (female).

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
2.000	-	-		-		-	-	-	-	-
2.002	Q2M: uh, can you describe to me what's going on?								norm narr	
2.003	W2F: ok, so there's			guy	1	s	new	normal	norm narr	a
2.004	in			room	1	s	new	normal	norm narr	a
2.005	and he's getting dressed and	his		pants	1	p	new	normal	norm narr	
2.006	are too short									
2.007	Q2M: mm-hmm									
2.008	W2F: and that's it									
2.009	Q2M: OK, um can you describe to me what the place is like? The setting?								norm descr	
2.010	W2F: Uh, well it's just like			small room	2	s	current	reintroducing	norm descr	a
2.011	um, there's it's like cluttered and there's			lot	1	s	new	normal	norm descr	a
2.012	of			clothes	1	p	new	normal	norm descr	Ø
2.013	everywhere um, yeah,			walls	1	p	frame	normal	norm descr	the
2.014	are green.									

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
2.015	Q2M: OK, um, can you describe to me what the people in the scene are like.								norm descr	
2.016	W2F: Um, there's only, well there's			person	1	s	new	normal	norm descr	a
2.017	in			backgroun d	1	s	shared situation	normal	norm descr	the
2.018	but you don't really see him and			main, main person	2	s	current	normal	norm descr	the
2.019	is um wearing			white shirt	1	s	new	normal	norm descr	a
2.020	and			tie	1	s	new	normal	norm descr	a
2.021	and he's just putting			pants	2	p	current	reintroducing	norm descr	Ø
2.022	on									
2.023	Q2M: OK, so, um So there's a dude putting his pants on								deferred descr	
2.024	W2F: yeah [laughs]									
2.025	Q2M: in this apartment?								deferred descr	
2.026	W2F: I guess it could be	his		apartment,	1	s	new	normal	deferred descr	
2.027	yeah.									
2.028	Q2M: OK, and um and it's just a goofy dude by himself with one man in the background?								deferred descr	
2.029	W2F: mm-hmm, like			guy	2	s	current	reintroducing	deferred descr	a
2.030	walks by in			back	2	s	current	normal	deferred descr	the
2.031	and then goes out of			door	1	s	new	normal	deferred descr	a
2.032	and closes			door	2	s	current	normal	deferred descr	the
2.033	Q2M: OK.									
2.034	W2F: Yeah, that's about it.									
2.035	Q2M: Alright.									

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
2.036	Experimenter: K? Ready?									
2.037	Q2M: #1 Mm-hmm. #									
2.038	W2F: #2 Yeah. #									
2.039	Expimenter: OK.									
2.040	Q2M: Alright, so, um, can you tell me little bit									a
2.041	about what's going on so far?								norm narr	
2.042	W2F: OK, so he walked out of			room	2	s	current	normal	norm narr	the
2.043	where he was and it wasn't			apartment,	2	s	category	normal	norm narr	an
2.044	it's like			school	1	s	category	normal	norm narr	a
2.045	So he now he's in like			hallway	1	s	new	normal	norm narr	a
2.045	where like I guess			other teachers	1	p	new	normal	norm narr	Ø
2.046	are?									
2.047	Q2M: Mm-hmm.									
2.048	W2F: Um and he's,	his		pants	2	p	current	normal	norm narr	
2.049	are short, like they're too short for him, so he's looking at			other teachers	2	p	current	normal	norm narr	Ø
2.050	and like, all	their		pants	1	p	new	normal	norm narr	
2.051	are			normal length	1	u	new	normal	norm narr	Ø
2.052	Q2M: Mm-hmm.									
2.053	W2F: and so he's like, flustered about	his		pants.	2	p	current	normal	norm narr	
2.054	[laughs] Um. And he grabs	some one's		pants	1	p	new	normal	norm narr	
2.055	and like measured	his		pants	2	p	current	normal	norm narr	
2.056	against his.									
2.057	Q2M: Mm-hmm.									

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
2.058	W2F: Um. Yeah.									
2.059	Q2M: OK. Um, you kind of explained my second question, so it's like in a school?								norm descr	
2.060	W2F: Yeah. It's like			school hallway.	2	s	category	reintroducing	norm descr	a
2.061	Q2M: OK.									
2.062	W2F: There's like			fliers	1	p	new	normal	norm descr	Ø
2.063	on			wall	1	s	frame	normal	norm descr	the
2.064	and like			bulletin boards	1	p	new	normal	norm descr	Ø
2.065	and			stuff.	1	u	new	normal	norm descr	Ø
2.066	Q2M: And he was in a closet the first time you said? Or a classroom?								deferred descr	
2.067	W2F: It looked like			closet.	1	s	new	reintroducing	deferred descr	a
2.068	Q2M: OK.									
2.069	W2F: There was like			clothes	2	p	current	reintroducing	deferred descr	Ø
2.070	everywhere.									
2.189	Q2M: OK. So for the places you said, he-, the settings were um, he went, it was it was all in the bathroom?								deferred descr	
2.190	W2F: Yeah. So like it started out, I don't know	what		kind	UNSURE	s	unsure	reintroducing	deferred descr	
2.191	of			room	2	s	unsure	reintroducing	deferred descr	Ø

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
2.192	it was. There were like			clothes	2	p	current	reintroducing	deferred descr	Ø
2.193	in there.									
2.194	Q2M: Mm-hmm.									
2.195	W2F: And then he went in			hallway,	2	s	current	normal	norm narr	the
2.196	and then he was in			bathroom.	2	s	current	normal	norm narr	the
2.197	Q2M: Right. OK. And um, and who were the, what- who were the characters in this in this scene.								deferred descr	
2.198	W2F:			Mr. Bean	2	s	proper name	reintroducing	deferred descr	Ø
2.199	and	some		random guys.	2	p	current	reintroducing	deferred descr	
2.200	Q2M: Um, who were the other random guys?								deferred descr	
2.201	W2F: Um, well in in			room	2	s	current	normal	deferred descr	the
2.202	in			beginning,	1	s	shared situation	normal	deferred descr	the
2.203	um, it was just			Mr. Bean,	2	s	proper name	normal	deferred descr	Ø
2.204	and like			guy	2	s	current	reintroducing	deferred descr	a
2.205	walked through			room	2	s	current	normal	deferred descr	the
2.206	and walked out. And then in			hallway	2	s	current	normal	deferred descr	the
2.207	there was like maybe			three or four other teachers?	2	p	current		deferred descr	Ø
2.208	Q2M: Mm-hmm.									
2.209	W2F: And then in			bathroom,	2	s	current	normal	deferred descr	the
2.210	there was			guy	2	s	current	reintroducing	deferred descr	a
2.211	washing	his		hands,	1	p	frame	normal	deferred descr	

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
2.212	and then there was			guy	2	s	current	reintroducing	deferred descr	a
2.213	at			uniral, ur-	2	s	current	normal	deferred descr	the
2.214				urinal,	2	s	current	normal	deferred descr	
2.215	and then there was			guy	2	s	current	reintroducing	deferred descr	the
2.216	in			stall,	2	s	current	reintroducing	deferred descr	the
2.217	and then			security guard guy.	2	s	current	reintroducing	deferred descr	the
3.000										
3.002	Q3F: Can you uh, can you describe what's happening so far?								norm narr	
3.003	W3F: It looks like			Mr. Bean	1	s	proper name	normal	norm narr	Ø
3.004	is getting dressed in	some		sort	1	s	new	normal	norm narr	
3.005	of			locker room?	1	s	new	normal	norm narr	Ø
3.006	He's got	his		shirt	1	s	new	normal	norm narr	
3.007	and	his		tie	1	s	new	normal	norm narr	
3.008	on, and now he's pulling	his		pants	1	p	new	normal	norm narr	
3.009	on.									
3.010	Q3F: And um, what is, what does the setting look like, what's the environment like?								norm descr	
3.011	W3F: Well, it looks like									a,
3.012				locker room	2	s	category	reintroducing	norm descr	a
3.013	or			dressing room	1	s	category	reintroducing	norm descr	a
3.014	of	some		sort.	2	s	current	reintroducing	norm descr	

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
3.072	Q3F: OK. And then just for clarity could you just try to explain what happend again								norm narr	
3.073	W3F: OK.									
3.074	Q3F: 'cause I'm not quite...									
3.075	W3F: OK. So			Mr. Bean	2	s	proper name	reintroducing	norm narr	Ø
3.076	walked out of			dressing room	2	s	current	reintroducing	norm narr	the
3.077	he was getting dressed in and as he was walking out			door,	2	s	frame	normal	norm narr	the
3.078	he looked down and noticed that	his		pants	2	p	current	normal	norm narr	
3.079	were probably			4 or 5 inches	1	p	new	normal	norm narr	Ø
3.080	too short.									
3.081	Q3F: OK.									
3.082	W3F: And someone walked by with			slightly short pants	1	p	new	normal	norm narr	Ø
3.083	on and I think that was			woman.	2	s	current	reintroducing	deferred descr	a
3.084	She had			suit coat	2	s	current	reintroducing	deferred descr	a
3.085	as well on but he really examined			length	2	s	frame	normal	norm narr	the
3.086	of	her		pants.	2	p	current	normal	norm narr	
3.087	And then as he walked out into			hall,	2	s	current	normal	norm narr	the
3.088	he was looking at			length	2	s	frame	normal	norm narr	the
3.089	of	ever yone' s		pants.	2	p	current	normal	norm narr	

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
3.293	Q3F: OK. I think I get it. Could you just go over the, kind of the last part of what happened again?								norm narr	
3.294	W3F: Uh-huh.									
3.295	Q3F: Or like halfway through or something?									
3.296	W3F: Yeah. So while	this		man	2	s	current	normal	norm narr	
3.297	was looking up to see what			noise	2	s	current	normal	norm narr	the
3.298	and what was going on up there			Mr. Bean	2	s	proper name	normal	norm narr	ø
3.299	reached underneath			door	2	s	current	normal	norm narr	the
3.300	and pulled	his		pants	2	p	current	normal	norm narr	
3.301	Q3F: OK.									
3.302	W3F: off, but he couldn't get	his		pants	2	p	current	normal	norm narr	
3.303	all			way	1	s	unsure	normal	norm narr	the
3.304	off so he mananged to pull			man	2	s	current	normal	norm narr	the
3.305	off			toilet.	2	s	current	normal	norm narr	the
3.306	And			man	2	s	current	normal	norm narr	the
3.307		's		bare legs	2	p	current	normal	norm narr	
3.308	are sticking out of			door	2	s	current	normal	norm narr	the
3.309	and he was trying to get	his		shoes-	2	p	current	normal	norm narr	
3.310		his		pants	2	p	current	normal	norm narr	
3.311	off of him when			security guard	2	s	current	normal	norm narr	the
3.312	walked in									
3.313	Q3F: Walked in.									
3.314	W3F: So			Mr. Bean	2	s	proper name	reintroducing	norm narr	ø
3.315	pulled out			handkerchief	2	s	current	reintroducing	norm narr	a

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
3.316	and started polishing	this		guy	2	s	current	reintroducing	norm narr	
3.317		's		shoes	2	p	current	reintroducing	norm narr	
3.318	Q3F: Oh god.									
3.319	W3F: pretending that that's what he was doing									
3.320	Q3F: OK.									
4.000										
4.002	Q4F: What's happening?								norm narr	
4.003	W4F: Um. So essentially there's	this		guy	1	s	new	normal	norm narr	
4.004	in	this		dressing room.	1	s	new	normal	norm narr	
4.005	Q4F: Mm-hmm.									
4.006	W4F: And uh, he happens to be kind of uh extremely			comical type person	1	s	category	normal	norm narr	Ø
4.007	with	his		body.	1	s	frame	normal	norm narr	
4.008	And he's um putting he ha- doesn't have any	any		pants	1	p	VP negation	normal	norm narr	
4.009	on And he just put	his		shirt	1	s	frame	normal	norm narr	
4.010	on and			other guy	1	s	new	normal	norm narr	an=
4.010	comes up and grabs			pair,	1	s	new	normal	norm narr	a
4.011	there's			two pair	1	u	new	normal	norm narr	Ø
4.012	a'			pants	1	p	new	normal	norm narr	Ø
4.013	next to each other,									
4.014	Q4F: Uh-huh									
4.015	W4F: and so he comes up and grabs			one	unsure			normal	norm narr	Ø
4.016	of			pair	2	s	current	normal	norm narr	the
4.017	a'			pants,	2	p	frame	normal	norm narr	Ø

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
4.018	and that happens to be			other guy	2	s	current	normal	norm narr	the
4.019		's		pants	2	p	current	normal	norm narr	
4.020	but			other guy	2	s	current	normal	norm narr	the
4.021	doesn't know that so he just grabs			other pair	2	s	current	normal	norm narr	the
4.022	a' (=of)			pants	2	p	current	normal	norm narr	ø
4.023	and starts put them on and then realizes he looks down, and			pants	2	p	current	normal	norm narr	the
4.024	are too short.									
4.025	Q4F: OK, what, like what is the setting?								norm descr	
4.026	W4F: Um, I'd say it's uh/a, it reminds me of			older styled dressing room	1	s	category	normal	norm descr	an
4.027	or like			locker room	1	s	category	normal	norm descr	a
4.028	ya know from probably like			30s	1	p	shared lexis	normal	norm descr	the
4.029	or			40s	1	p	shared lexis	normal	norm descr	the
4.030		’		type style.	1	s	new	normal	norm descr	
4.031	Q4F: OK. And what's happening again in the dressing room?								norm narr	
4.032	W4F: Um, they're getting dressed,									
4.033	Q4F: OK									
4.034	W4F: and, yeah,	some		guy	1	s	new	reintroducing	norm narr	
4.035	stole			other guy	2	s	current	reintroducing	**norm narr**	an=
4.036		's		pants.	2	p	current	reintroducing	norm narr	
4.037	Q4F: And what are the people like in it?								norm descr	

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
4.038	W4F: Um, there's	this		one uh awkardly gangly really brunette looking guy	1	s	new	normal	norm descr	
4.039	Q4F: Mm-hmm.									
4.040	W4F: um, with kind of			weird look-face.	1	s	new	normal	norm descr	a
4.041	Q4F: Mm-hmm.					s				
4.042	W4F: And then there's			shorter guy.	1	s	new	normal	norm descr	a
4.043	Q4F: Mm-hmm.									
4.044	W4F: Um, he's definitely noticeably shorter									
4.045	Q4F: Mm-k.									
4.046	W4F: than			main guy.	2	s	current	normal	norm descr	the
4.047	Q4F: And where are they again?								deferred descr	
4.048	W4F: They're in			locker room.	2	s	current	reintroducing	deferred descr	a
4.049	Q4F: OK. OK.									
4.084	Q4F: Oh wow. And what do the people look like again? Like the guy who got his pants stolen and the woman.								deferred descr	
4.085	W4F: He is, well now he just looks really angry									
4.086	Q4F: OK.									
4.087	W4F: in			very bitter, funny type way.	1	s	new	normal	deferred descr	a
4.088	Q4F: OK.									

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
4.089	W4F: Um, h- h- he looks like he has um, what are tho-			Warheads	1	p	proper name	normal	deferred descr	the
4.090	in	his		mouth?	1	s	frame	normal	deferred descr	
4.091	#1 just a so- get(?)			little	1	s	unsure	normal	deferred descr	a
4.092	sour? #									
4.093	Q4F: #2 Yeah yeah yeah yeah. #									
4.094	W4F: And um, but yeah, I(?) still, awkward, gangly, tall. And then there was			woman	2	s	current	reintroducing	norm descr	a
4.095	with			blonde hair	1	u	new	normal	norm descr	Ø
4.224	Q4F: #2 Oh, god. #									
4.225	W4F: And then I think			security guard	1	s	new	normal	norm narr	a
4.226	walks in or something. I dunno, he's wearing			red sash.	1	s	new	normal	norm narr	a
4.227	I've never seen			security outfit	1	s	new	normal	norm narr	a
4.228	like that. But he's			secur-						a
4.229	he's got			official hat	1	s	new	normal	norm narr	an
4.230	on, so probably.									
4.231	Q4F: Wow. Does he look like a police officer, or what does he look like?									
4.232	W4F: He kinda looks like			police officer.	1	s	category	normal	norm narr	a
4.233	I don't-									
4.234	Q4F: But he has a sash?								deferred descr	
4.235	W4F: Yeah, it's			red sash.	2	s	current	reintroducing	deferred descr	a

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
4.236	He's got			police officer type hat,	1	s	new	reintroducing	deferred descr	the
4.237	but he's also got			wood club.	1	s	new	normal	deferred descr	a
4.238	W4F: #1 It's more like a wood stick, not a- even a club. #									
6.000										
6.130	Q6M: Cool, can you describe the place again?								deferred descr	
6.131	W6M: Uh, it's			bathroom.	2	s	current	reintroducing	deferred descr	a
6.132	It's got			three stalls,	1	p	new	reintroducing	deferred descr	
6.133	uh,			three urinals,	1	p	new	reintroducing	deferred descr	
6.134				red floor.	1	s	new	normal	deferred descr	
6.135	Um,			pretty old bathroom.	1	s	category	normal	deferred descr	
6.136	Probably smelly.									
6.137	Q6M: OK. [long silence] K, can you describe what's happening?								norm narr	
6.138	W6M: So, uh, he walks back into			same bathroom,	2	s	current	normal	norm narr	the
6.139	Q6M: Mm-hmm.									
6.140	W6M: because he, as he was walking out he saw that someone was in			stall	1	s	new	normal	norm narr	a
6.141	and uh he walks back in and uh he looks under			stall	2	s	current	normal	norm narr	the
6.142	and he sees that on			inside	1	s	frame	normal	norm narr	the

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
6.143	of			pants	1	p	frame	normal	norm narr	the
6.144	it says			"Bean"	2	s	proper name	normal	norm narr	
6.145	and in			parentheses	1	p	new	normal	norm narr	
6.146				"Mr."	2	s	proper name	normal	norm narr	
6.147	so they're	his		pants.	2	p	current	normal	norm narr	
6.148	Then he uh tries to distract him									
6.149	Q6M: Mm-hmm.									
6.150	W6M: Uh.									
6.151	Q6M: Can you see the person?								unsure	
6.152	W6M: Um, yeah. Well,									th-
6.153	from			different (camera angle[??])	1	s	new		unsure	
6.154	he tries to reach over			top	1	s	frame	normal	norm narr	the
6.155	to distract him and then while			person	2	s	current	normal	norm narr	the
6.156	inside			stall	2	s	current	normal	norm narr	the
6.157	is looking up he uh tries to go under and grab			pants	2	p	current	normal	norm narr	the
6.158	Q6M: Mm-hmm.									
6.159	W6M: and as he's grabbing			pants,	2	p	current	normal	norm narr	the
6.160	uh what seems to be like									a,
6.161				safety officer	1	s	new	normal	norm narr	a
6.162	#1 of			building	2	s	current	normal	norm narr	the
6.163	#									
6.164	Q6M: #2 Mm-hmm. #									
6.165	W6M: walks in with			stick	1	s	new	normal	norm narr	a
6.166	and uh			Mr. Bean	2	s	proper name	normal	norm narr	
6.167	pulls out									a,

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
6.168	pulls out			cloth	1	s	new	normal	norm narr	a
6.169	and pretends to be um polishing			guy	2	s	current	normal	norm narr	the
6.170		s		shoes	1	p	frame	normal	norm narr	
6.171	Q6M: Yeah.									
6.172	W6M: and uh so uh,			officer	2	s	current	normal	norm narr	
6.173	doesn't see anything wrong so he starts to walk out.									
6.174	Q6M: Uh, what does the person in the stall look like?								norm descr	
6.175	W6M: Um, it's			gentleman,	2	s	current	reintroducing	norm descr	a
6.176				grey hair,	1	u	new	normal	norm descr	
6.177	uh, don't r-, white, don't really get to he's pretty- sh- shorter than			Mr. Bean	2	s	proper name	normal	norm descr	
6.178	Q6M: Mm-hmm.									
8.000										
8.002	Q8M: Alright. Um. Can you tell me what was going on in the video?								norm narr	
8.003	W8M: OK, so in			video	2	S	shared situation	normal	norm narr	the
8.004	there's			man	1	s	new	normal	norm narr	a
8.005	who is preparing, he's getting dressed and he's getting dressed uh in			formal dress clothes	1	p	new	normal	norm narr	Ø
8.006	and he seems to be doing it rather awkwardly. He uh, it shows him first with	his		dress shirt	1	s	new	normal	norm narr	
8.007	on and			tie,	1	s	new	normal	norm narr	a

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
8.008	and he's about to put on	his		pants	1	p	new	normal	norm narr	
8.009	and he's doing it again very awkwardly, and he finally gets	his		pants	2	p	current	normal	norm narr	
8.010	up and he looks down and he realizes that he has forgotten	his		shoes,	1	p	new	normal	norm narr	
8.011	and	his		dress socks	1	p	new	normal	norm narr	
8.012	are all the way rolled down, and he has	no		shoes	2	p	current	normal	norm narr	
8.013	on. He's wearing			white dress shirt	1	s	new	normal	norm descr	a
8.014	with			red tie	1	s	new	normal	norm descr	a
8.015	and just			black slacks,	1	p	new	normal	norm descr	Ø
8.016	and he's in			changing room,	1	s	new	normal	norm descr	a
8.017	maybe at			work,	1	u	new	normal	norm descr	Ø
8.018	or at			gym,	1	s	new	normal	norm descr	a
8.019	I would say. And there seems to be			other person	1	s	new	normal	norm narr	an=
8.020	in there who is staring at him.									
8.021	Q8M: Right, and um, what did this place look like?								norm descr	
8.022	W8M: Uh, it just kinda looks like			dressing room,	2	s	category	normal	norm descr	a
8.023	like, there's			place	1	s	new	normal	norm descr	a
8.024	like, hang up			coats	1	p	new	normal	norm descr	Ø
8.025	and um well, like there're like			dress coats	1	p	new	normal	norm descr	Ø
8.026	and there're			dress clothes	2	p	new	normal	norm descr	Ø
8.027	on			sides.	1	p	frame	normal	norm descr	the
8.028	It's kind of			dull colored room.	1	s	new	normal	norm descr	a

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
8.029	Not very bright or anything. And there's			couple	1	s	new	normal	norm descr	a
8.030				little benches	1	p	new	normal	norm descr	Ø
8.031	that, for, to also put	**your**		belongings	1	p	new	normal	norm descr	
8.032	on.									
8.033	Q8M: And uh, what did the people look like?								norm descr	
8.034	W8M: Uh,			main character	2	s	current	normal	norm descr	the
8.035	has			medium length hair.	1	u	new	normal	norm descr	Ø
8.036	It's			man.	1	s	current	reintroducing	norm descr	a
8.037	It's			dark, dark hair.	1	u	new	normal	norm descr	uh(a?),
8.279	Q8M: Alright um, and what did the other people in the scene look like?								norm descr	
8.280	W8M: There's still			main character	2	s	current	normal	norm descr	the
8.281	with	his		black,	1			normal		
8.282	er,	his		brown slacks	2	p	current	normal	norm descr	
8.283	on that are too high and uh	his		rolled up socks,	2	p	current	normal	norm descr	
8.284	and			guard	2	s	current	normal	norm descr	the
8.285	comes in. He has			brown slacks	1	p	new	normal	norm descr	Ø
8.286	on. He's			older man,	1	s	new	normal	norm descr	an
8.287	uh, maybe around			50,	1	u	new	normal	norm descr	Ø
8.288	and he's, has			mustache,	1	s	new	normal	norm descr	a
8.289	and he comes in carrying			bigger stick,	1	s	new	normal	norm descr	a
8.290	and he has			hat	2	s	current	reintroducing	norm descr	a

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
8.291	on that resembles			police officer	1	s	category	reintroducing	norm descr	a
8.292		's		hat.	1	s	frame	reintroducing	norm descr	
8.293	He has			green sweater	1	s	new	normal	norm descr	a
8.294	on with			red sash	2	s	current	reintroducing	norm descr	a
8.295	that is covering, that's coming from	his		shoulder	2	s	current	reintroducing	norm descr	
8.296	to	his		hip	1	s	frame	reintroducing	norm descr	
8.297	that appears to um, show that he's of	some		kind	1	s	new	normal	norm descr	
8.298	of			authority.	1	u	new	normal	norm descr	Ø
8.299	And then there's			other older man	2	s	current	reintroducing	norm descr	an=
8.300	that's in			stall,	2	s	current	normal	norm descr	the
8.301	that's using			stall,	2	s	current	normal	norm descr	the
8.302	and he has			main character	2	s	current	normal	norm descr	the
8.303		's		pants	2	p	current	normal	norm descr	
8.304	on, and they are			other pair	2	s	current	reintroducing	norm descr	an=
8.305	of			slacks,	2	p	current	reintroducing	norm descr	Ø
8.306	and, yeah.									
9.000										
9.033	Q9F: Can you describe what's happening?								norm narr	
9.034	W9M: Yeah, so he left			changing room	2	s	current	normal	norm narr	the
9.035	and now he's like out in			hall,	1	s	frame	normal	norm narr	the
9.036	and he's walking around looking at all	these		other guys	1	p	new	normal	norm narr	
9.037		'		pants	1	p	new	normal	norm narr	

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
9.038	and comparing them to his own, and uh, he goes up to	this		one guy,	1		new	normal	norm narr	
9.039	and he starts like, he gets really close and starts comparing	his		pant length	1	u	frame	normal	norm narr	
9.040	to his, and he like pulls			back	1	s	frame	normal	norm narr	the
9.041	of	his		tag	1	s	frame	normal	norm narr	
9.042	up and starts looking at that.									
9.043	Q9F: Um, can you describe the place where it's happening?								norm descr	
9.044	W9M: Uh, it looks like just	some		hallway.	2	s	current	reintroducing	norm descr	
9.045	There's like			bulletin boards	1	p	new	normal	norm descr	Ø
9.046	on			wall,	1	s	frame	normal	norm descr	the
9.047	and			tile floor.	1	s	new	normal	norm descr	Ø
9.048	Q9F: What do the people in the scene look like?								norm descr	
9.049	W9M: Uh, they're all like			middle aged men	1	p	new	normal	norm descr	Ø
9.050	it looks like. There's			African American man,	1	s	new	normal	norm descr	an
9.051	that's			guy	2	s	current	normal	norm descr	the
9.052	who he goes up to and like pulls his like,			back	2	s	frame	normal	norm descr	the
9.053	of	his		pants	2	p	current	normal	norm descr	
9.054	and checks			tag.	2	s	current	normal	norm descr	the
9.055	And then there's like			two other guys	1	p	new	normal	norm descr	Ø
9.056	in			backgroun d,	1	s	shared situation	normal	norm descr	the
9.057	just kinda like looking at			bulletin boards	2	p	current	normal	norm descr	the
9.058	and			stuff.	1	u	new	normal	norm descr	Ø

Line	Context	Other Determiner/ Genitive	Article (hidden)	NP	Sequence of Mentions	Number	Referentiality/ Identifiability	2 Discourse Mode Model	3 Discourse Mode Model	Article (revealed)
9.059	Q9F: Can you describe the place again?								deferred descr	
9.060	W9M: Yeah, it's			hallway,	2	s	current	reintroducing	deferred descr	a
9.061	it's got			tile floor,	2	s	current	reintroducing	deferred descr	a
9.062				bulletin boards,	2	p	current	reintroducing	deferred descr	Ø
9.063	uh yeah,			yellow wall painting.	1	u	new	normal	deferred descr	Ø
9.064	Q9F: K.									
19.000										
19.110	Q19F: Where is he walking now? Where was he?								norm descr	
19.111	W19M: He's in, he's in			hallway,	2	s	current	normal	norm descr	a
19.112	um, I, [inaudible] Yeah,			hallway	2	s	current	reintroducing	norm descr	Ø
19.113	with, that comes to			T	1	s	new	normal	norm descr	a
19.114	at							normal		a,
19.115	at			reader board.	2	s	current	normal	norm descr	a

REFERENCES CITED

Abbott, B. (1993). A pragmatic account of the definiteness effect in existential sentences. *Journal of Pragmatics*, *19*(1), 39–55.

Abbott, B. (1997). Definiteness and existentials. *Language*, *73*(1), 103–108.

Abbott, B. (1999). Support for a unique theory of definiteness. *Proceedings from Semantics and Linguistic Theory IX* (Vol. 9, pp. 1–15). Ithaca, NY: CLC Publications.

Abbott, B. (2000). Presuppositions as nonassertions. *Journal of Pragmatics*, *32*(10), 1419–1437.

Abbott, B. (2001). Definiteness and identification in English. In N. T. Enikö (Ed.), *Pragmatics in 2000: Selected papers from the 7th International Pragmatics Conference* (Vol. 2, pp. 1–15). Antwerp: International Pragmatics Association.

Abbott, B. (2006). Definite and indefinite. In E. K. Brown, R. E. Asher, & J. M. Y. Simpson (Eds.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 392–399). Boston: Elsevier.

Abbott, B. (2007). Definiteness and indefiniteness. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 122–149). Oxford: Blackwell.

Abbott, B. (2008). Issues in the semantics and pragmatics of definite descriptions in English. In J. K. Gundel & N. Hedberg (Eds.), *Reference: interdisciplinary perspectives* (pp. 61–72). Oxford: Oxford University Press.

Abbott, B. (2011). Support for a unique theory of definiteness. *Proceedings of SALT* (Vol. 9, pp. 1–15).

Anderson, J. M. (2007). *The grammar of names*. Oxford: Oxford University Press.

Anderson, L. B. (1982). The "perfect" as a universal and as a language-particular category. In P. J. Hopper (Ed.), *Tense-aspect: between semantics & pragmatics*, Typological studies in language (Vol. 1, pp. 227–264). Presented at the Symposium on Tense and Aspect, May 1979, UCLA: John Benjamins Publishing Co.

Anderson, L. B. (1986). Evidentials, paths of change, and mental maps: typologically regular asymmetries. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The Linguistic Coding of Epistemology* (pp. 273–312). Norwood, NJ: Ablex Pub. Corp.

Azar, B. S. (1999). *Understanding and using English grammar* (3rd ed.). White Plains, NY: Pearson Education.

Barras, C. (2002). Transcriber (Version 1.5.1) [computer software]. Available from http: //trans.sourceforge.net/en/presentation.php

Barsalou, L. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization* (pp. 21–74). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Berezowski, L. (2001). *Articles and proper names*. Wydawn: Uniwersytetu Wrocławskiego.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Burge, T. (1973). Reference and proper names. *The Journal of Philosophy*, *70*(14), 425–439.

Chafe, W. (1996). Inferring identifiability and accessibility. In T. Fretheim & J. K. Gundel (Eds.), *Reference and referent accessibility* (pp. 37–46). Amsterdam: John Benjamins Publishing Company.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view in subject and topic. In W. Chafe & C. Li (Eds.), *Subject and Topic* (pp. 24–55). New: Academic Press.

Chafe, W. (1994). *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago: Univ. of Chicago Press.

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

Cormack, A., & Kempson, R. (1991). On specificity. In J. Garfield & M. Kitely (Eds.), *Meaning and Truth: Essential Readings in Modern Semantics* (pp. 546–581). New York: Paragon House.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.

De Haan, F. (2004). On representing semantic maps. *E-MELD Language documentation conference 2004: Workshop on linguistic databases and best practice*.

De Haan, F. (2010). Building a semantic map: top-down versus bottom-up approaches. *Linguistic Discovery*, *8*(1), 102–117.

Du Bois, 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* (pp. 203–274). Norwood, NJ: Ablex Publishing Corp.

Epstein, R. (1999). Roles, frames, and definiteness. In K. Van Hoek, A. A. Kibrik, & L. G. M. Noordman (Eds.), *Discourse studies in cognitive linguistics: selected papers from the fifth International Cognitive Linguistics Conference, Amsterdam, July 1997* (pp. 53–74). Presented at the Fifth International Cognitive Linguistics Conference, Amsterdam. Philadelphia: J. Benjamins.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1–88). New York: Holt, Rinehart, and Winston.

Fodor, J. D., & Sag, I. A. (1982). Referential and quantificational indefinites. *Linguistics and philosophy*, *5*(3), 355–398.

Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, *66*(2), 297–316.

Givón, T. (2001). *Syntax: an introduction* (Vols. 1-2, Vol. 1). Amersterdam: John Benjamins Publishing Company.

Givón, T. (2005). *Context as other minds: The pragmatics of sociality, cognition, and communication*. Amersterdam: John Benjamins Publishing Company.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.

Haspelmath, M. (1997). *Indefinite pronouns*. Oxford: Clarendon Press.

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language* (Vol. 2, pp. 211–242). Mahwah, NJ: Lawrence Erlbaum Associates.

Hawkins, J. (1978). *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom Helm.

Hawkins, J. A. (1991). On (in) definite articles: implicatures and (un) grammaticality prediction. *Journal of Linguistics*, *27*(2), 405–442.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases.* (Doctoral dissertation). University of Massachusetts, Amherst.

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In J. Gutiérrez-Rexach (Ed.), *Semantics.* Critical Concepts in Linguistics Series: Vol. 3 (pp. 108–135). London: Routledge.

Kadmon, N. (1990). Uniqueness. *Linguistics and Philosophy*, *13*(3), 273–324.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic inquiry*, *8*(1), 63–99.

Lambrecht, K. (1994). *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge; New York, NY, USA: Cambridge University Press.

Levinson, D. (2005). *Definiteness of Noun Phrases with Body Part Terms*. Manuscript.

Löbner, S. (1985). Definites. *Journal of Semantics*, *4*(4), 279–326.

Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics*. Cambridge; New York: Cambridge University Press.

Mårdh, I. (1980). *Headlinese: On the grammar of English front page headlines* (Vol. 58). Lund: Liberläromedel/Gleerup.

Matushansky, O. (2008). On the linguistic complexity of proper names. *Linguistics and Philosophy*, *31*(5), 573–627.

Murphy, R. (2007). *English grammar in use*. Cambridge: Cambridge University Press.

Palestinian prisoner ends 102-day hunger strike. (2012, July 23). *Al-Akhbar.com*. Retrieved from http://english.al-akhbar.com/content/palestinian-prisoner-ends-102-day-hunger-strike

Pederson, E. (1991). The Ecology of a Semantic Space. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (Vol. 17, pp. 457–468).

Pederson, E. (1991b). *Subtle semantics: Universals in the polysemy of reflexive and causative constructions*. (Doctoral dissertation). University Microfilms International, University of California at Berkeley.

Reuland, E. J., & ter Meulen, A. G. B. (1987). *The representation of (in)definiteness*. Cambridge, Massachusetts: The MIT Press.

Rice, S., & Prideaux, G. (1991). Event-packing: the case of object incorporation in English. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Grammar of Event Structure* (Vol. 17, pp. 283–298).

Riley, J. A. (2007). Mental representations: Reference and definiteness. *Journal of Pragmatics*, *39*(5), 831–871.

Russell, B. (1919). *Introduction to mathematical philosophy*. Mineola, NY: Dover publications.

Russell, B. (1905). On denoting. *Mind*, *14*, 479–493.

Sapir, E. (1929). Central and North American languages. In *Encyclopedia Britannica*. Chicago: Encyclopedia Britannica Inc.

Vonnegut, K. (1975). *The Eden Express*. New York: Bantam.