Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kimberly Edmunds April 16, 2013

Chinese ESL Learners' Overuse of the Definite Article: A Corpus Study

by

Kimberly Edmunds

Marjorie Pak, Ph.D.
Adviser

Program in Linguistics

Marjorie Pak, Ph.D.
Adviser

Yu Li, Ph.D.
Committee Member

Lynna Williams
Committee Member
2013

Chinese ESL Learners' Overuse of the Definite Article: A Corpus Study

By

Kimberly Edmunds

Marjorie Pak, Ph.D.
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2013

Abstract
Chinese ESL Learners' Overuse of the Definite Article: A Corpus Study
By Kimberly Edmunds

Even at an advanced level of English proficiency, non-native speakers will sometimes produce errors of unidiomaticity: language that is clear in its meaning, but sounds "strange" to native speakers. Chinese ESL learners sometimes unnecessarily insert the definite article in their academic writing, producing just such unidiomatic patterns. Much of the previous research on this phenomenon has been limited to prompted tasks or narrative contexts, leaving a gap in the understanding of *the*-overuse that is particularly relevant to college-level ESL students. By compiling and analyzing data from the CEECUS corpus of Chinese ESL learners' English essays, this project demonstrates how language competence can be more fully measured without prompting subjects. Furthermore, it places the writing in an argumentative discourse style, removing it from a narrative context. Every common noun in the corpus was coded according to specific semantic, morphological, and discourse-status features, including the presence or absence of an article. Results indicate that generic and non-referential indefinite noun phrases are more likely than others to induce errors of *the*-overuse. Within these groups, additional features such as ambiguity and sentence position of the noun demonstrate effects on the rate of error. A confluence of factors may have contributed to the patterns observed: interference from the L1, previous classroom experience, the distribution of semantic and morphological features, and the argumentative context. Bringing these perspectives to pedagogy is essential to a Chinese ESL learner's accurate understanding of the definite article, and to their success in the argument-driven American educational system.

Chinese ESL Learners' Overuse of the Definite Article: A Corpus Study

By

Kimberly Edmunds

Marjorie Pak, Ph.D.
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2013

Acknowledgements

I would like to thank my committee members, Dr. Yu Li and Professor Lynna Williams, for their feedback, patience, and enthusiasm throughout the writing process. I would also like to express my gratitude to my adviser, Dr. Marjorie Pak, for her encouragement and support every step of the way.

Table of Contents

"So often the patient language-learner is told by the native speaker that a particular sentence is perfectly good English...but that native speakers would never use it. How are we to explain such a state of affairs?" –David J. Allerton, 1984

CHINESE ESL LEARNERS' OVERUSE OF THE DEFINITE ARTICLE: A CORPUS STUDY

KIMBERLY EDMUNDS
Emory University
Program in Linguistics
Undergraduate Honors Thesis

ABSTRACT

Even at an advanced level of English proficiency, non-native speakers will produce errors of unidiomaticity: language that is clear in its meaning, but sounds "strange" to native speakers. Chinese ESL learners sometimes unnecessarily insert the definite article in their academic writing, producing just such unidiomatic patterns. Much of the previous research on this phenomenon has been limited to prompted tasks or narrative contexts, leaving a gap in the understanding of *the*-overuse that is particularly relevant to college-level ESL students. By compiling and analyzing data from the CEECUS corpus of Chinese ESL learners' English essays, this project demonstrates how language competence can be more fully measured without prompting subjects. Furthermore, it places the writing in an argumentative discourse style, removing it from a narrative context. Every common noun in the corpus was coded according to specific semantic, morphological, and discourse-status features, including the presence or absence of an article. Results indicate that generic and non-referential indefinite noun phrases are more likely than others to induce errors of *the*-overuse. Within these groups, additional features such as ambiguity and sentence position of the noun demonstrate effects on the rate of error. A confluence of factors may have contributed to the patterns observed: interference from the L1, previous classroom experience, the distribution of semantic and morphological features, and the argumentative context. Bringing these perspectives to pedagogy is essential to a Chinese ESL learner's accurate understanding of the definite article, and to their success in the argument-driven American educational system.

1.0 Introduction

Idiomaticity exists; otherwise, second language learners would be able to take the rules they have learned through classroom study and observation, and eventually be able to produce L2 (second language) speech that, apart from accent, would be indistinct from that of native speakers. However, this does not occur; even very advanced L2 speakers are often told that something "gives them away;" they lack the idiomaticity of native speakers. Even when one is communicating perfectly clearly--which is the professed goal of many modern foreign language

classrooms--one's speech can still be called unidiomatic. In certain cases, it is easy to identify the reason why: a misplaced preposition, or the use of an outdated word. In others, however, it is difficult to explain why a native speaker just would not say something that way in normal conversation.

As an ESL writing tutor for my university's international student community, I have had the urge many times to dismiss a student's error as something that is simply "the way it is." The international students at my university are mainly from China, and adapting to a system of learning based on written, rhetorical argumentation is not always easy for them, though most are relatively proficient English speakers. I began to wonder: what, exactly, contributes to unidiomaticity in their ESL writing?

Of course, it is not one thing: it is quite a few, and in order to make any headway, I would have to pick just one angle, one pattern, and examine it systematically. I made this choice based on my experience with my Chinese tutees, after reading a good deal of their writing and taking note of the types of errors they tended to make. Many of these errors did nothing in the way of obscuring meaning, and perhaps were not even ungrammatical if interpreted in a certain way. However, their occurrence contributed to the unidiomaticity of their writing language and style, and if there is indeed a pattern, we may be able to call the students' attention to these errors in the earlier stages of their learning process. In this thesis, I chose to analyze and study the specific phenomenon of Chinese ESL students' overuse errors with the English definite article, *the.* In this study, I define total *the* overuse as the occurrence of definite articles in excess of the number that is actually required in a given discourse. In other words, I consider how much more often learners used *the* than was necessary overall.

To understand the inappropriate use of the definite article, it is first essential to define when such an article is used *appropriately* (which will be addressed shortly). Even without such a briefing, however, the following examples, drawn from the CEECUS corpus of advanced Chinese ESL writers, would still have sounded strange to a native speaker's ear (note that none of these nouns has been previously mentioned in the discourse):

1. When we come to *the college*, almost every new student will be into *the situation.* (PTJ 10)
2. Maybe study *the knowledge* from *the books* well is important. (PTJ 24)
3. *The good score* is the precondition for a good job. (PTJ 27)
4. They can put what they have experience and get to know *the society*.[1] (PTJ 35)

Why should these writers overuse the definite article in such a way? **Overuse** of an article is not what would be intuitively expected for this particular group of speakers. Mandarin Chinese itself does not have an article system commensurate to that of English; one might expect that they would be more likely to **omit** articles (Robertson 2000), as these speakers are not accustomed to using them. Furthermore, there is the question of whether or not the errors appear at random, as if the writers lack any knowledge of article use whatsoever and thus overuse *the* in all contexts, or if there are predictable patterns to the errors. To be clear, I am not claiming that overuse of the definite article necessarily occurs more often than underuse; rather, I simply hope to investigate overuse specifically.

Other researchers have noticed the phenomenon of *the*-overuse, and have proposed various hypotheses for its occurrence. However, much of the work has been limited to specific, prompted

[1] These examples are drawn from the corpus used in the present study, the Corpus of English Essays by Chinese University Students, or CEECUS. "PTJ10" etc. reflects the number of the essay in which the sentence was found.

tasks or to narrative contexts, and tends to emphasize only particular semantic features of nouns. This previous research has left a gap in the understanding of *the*-overuse, a gap that is particularly relevant to college-level ESL students: how does their knowledge of English manifest itself in a non-narrative discourse style, like the persuasive essays so prevalent in American higher education? In addition, how do context and additional factors such as classroom experience play a role in this type of error?

By analyzing the spontaneous English writing of Chinese ESL learners, this thesis will demonstrate how aspects of language competence, such as the understanding of the definite article, can be more fully measured without (or certainly in addition to) instructive priming. Furthermore, it will place the writing in an argumentative discourse style by removing it from narrative discourse styles, in which certain types of noun phrases are disproportionately likely to occur. Also, rather than emphasizing only certain semantic features of noun phrases, I will argue that an understanding of how generic nouns are distributed not only semantically, but also morphologically[2], is essential to a Chinese ESL learner's accurate grasp of the definite article. Finally, I will consider how influences from these learners' first language and from their previous English-learning experience may also contribute to patterns of *the*-overuse. A thorough understanding of this wide array of factors is especially crucial to the development of ESL pedagogy and instructor training, and to the success of Chinese ESL students in the American educational system.

The paper will be organized as follows: in the first section, I will differentiate between narrative discourse styles and argumentative discourse styles in how I define them and how they

[2]"Semantics" refers to the way that native speakers understand the meaning of a word—the information carried in that word, and what it means in any given context. "Morphology" characterizes how words change in terms of sound or structure in order to add new semantic information. Morphemes, or meaningful units, attach to or are associated with words and change the way we interpret them: for example, the –s on *dogs* is a plural morpheme, so that we interpret it as "many dogs," rather than one.

are relevant to this project. I will also discuss the use and distribution of the English articles, and the differences between overuse and omission. Section 2 will focus on previous research on this topic, how the relevant studies relate to my points here in terms of the discourse styles tested and their findings. Section 3 will detail the methodology of the present study, followed by section 4, which will report the results of the study. In section 5, I will elaborate further on some of the more concrete findings, including potential interference factors; I will also consider the implications of this study for the results of the previous research. Finally, in section 6, I will consider some possible pedagogical applications of the findings of this study, and make concluding remarks, including speculation about future research.

1.1 Narrative versus Argumentative Styles

It is important for the reader to understand the difference between **narrative discourse styles** and **argumentative discourse styles** as they are defined in this paper, as the different purposes and structures of these styles, I believe, play a role in how Chinese ESL learners use and/or overuse the definite article.

Narrative styles, as Liu & Gleason describe, consist of "telling stories or describing events," which in turn "require a speaker to rely heavily on the accurate textual use of the definite article to communicate clearly." (Liu & Gleason 15) This *textual* use refers to the role that the definite article plays in referring to entities that have already been introduced to a discourse, throughout the discourse. See the following example:

> **5. Narrative discourse style:**
> A woman came home one day to find that her house had been robbed. Glass from a window was everywhere; *the window* had been smashed. A neighbor of ours said he hadn't seen or heard anything, but *the house* had been trashed. *The woman* also realized that a new laptop had been stolen she had been planning to sell *the laptop* to an online buyer. We found some muddy footprints on the floor, but *the footprints* didn't lead anywhere.

Argumentative discourse styles, on the other hand, generally require less grappling with *textual* use of articles. The act of persuasion, at least in an English-speaking context, is instead likely to require the writer to use the language of generalization: to make broad statements, refer to groups of entities, acknowledge commonly understood states of the world, and perhaps even refer to generally accepted "rules". The use of the definite article with referential nouns is less frequent; I have observed this during my years as a university ESL tutor, and this is the style of writing found in CEECUS. The following is an example of this style as I define it:

> **6. Argumentative Discourse Style:**
> Preventing home *burglaries* requires preparation in advance. We recommend having *a security system* installed, certainly. In addition, it is prudent to lock all *doors* and participate in *neighborhood watch systems.* Sometimes, *burglars* will break in through *windows,* so *homeowners* should take special precautions and place *valuables* in secure *locations.*

What these examples serve to show is that the distribution of the different English articles is going to differ between argumentative styles and narrative styles. In narrative styles, there tend to be more referential nouns like *the window, the woman,* and *the footprints*; in argumentative styles, by contrast, there are fewer referential nouns and more generic nouns like *burglars, homeowners,* and *doors.* As a result, I believe that ESL learners' language performance and use of *the* will vary according to the discourse style in which they are producing language. A deeper discussion of the use and distribution of the English articles, including their semantic and morphological features, will be undertaken in the next section.

My argument here is that not only do factors like specific reference (referentiality) and hearer knowledge influence an ESL learner's use of *the,* but also does the discourse context in which they are producing language. This is why experimental methods that actively elicit articles

from subjects using narrative-style sentences may not accurately demonstrate the full extent of a learner's knowledge, nor help us determine what precisely causes them to overuse *the.* In the present study, I have chosen to take a different approach to this question: by analyzing spontaneously written, argumentative material, I hope to be able to understand how discourse style in addition to a combination of factors may contribute to *the*-overuse unidiomaticity in the writing of this group of Chinese ESL learners. More and more students from China are coming to the United States to attend college, and understanding American rhetorical strategy and argumentative style is essential to their success.

1.2 The Articles

Cross-linguistically, articles serve a variety of purposes, indicating different aspects of the nouns they accompany. Not all languages have what can be described as an article system. However, in the case of English, we are dealing with three: the indefinite article (*a/an),* the definite article (*the*), and the zero (sometimes called the null) article. The zero article is not pronounced or written, so it is indistinguishable from cases of omission or non-use of an article (the zero article is notated with ∅) (Master 1997).

Dilin Liu and Johanna Gleason discuss Bickerton's (1981) theory of articles, which remains one of the most influential works on the topic. He postulated two discourse characteristics of nouns: **Specific Reference** (SR) and **Hearer Knowledge** (HK). If a noun is **referential,** it has specific reference [+SR]; it "picks out" one particular entity in the world and specifically refers to it; nouns with no specific reference [-SR] do not refer to a specific entity; they are **non-referential**. If a noun is hearer known [+HK], it means that the person on the receiving end of the utterance (spoken or written) has a preexisting belief or assumption that the referent of the noun is exists in the world or the relevant context (Liu & Gleason 2).

7. I have some glasses I wear for driving. *The glasses* are fairly ugly. [+SR +HK]
8. Around the corner, you'll find *a friendly dog*. [+SR –HK]
9. I really need *a job interview*. */ Job interviews* are difficult to get. [-SR +HK]
10. She doesn't have *a bicycle* to ride to work. [-SR –HK]

In example (7), the glasses were previously mentioned in the discourse, so the noun is referential, and the hearer assumes that the glasses exist. In example (8), this is the first time the dog in question is introduced—so, while the noun is referential, the hearer does not have previous assumptions of that dog's existence. Both noun phrases in example (9) are not referential, because they "pick out" categories of referents (to be discussed in further detail later). In such cases, the hearer *does* assume that such a thing as a job interview exists. Finally, in example (10), the negation (she *doesn't)* presupposes the non-existence of the bicycle, so that the noun cannot logically be referential, and the hearer could not have had previous assumptions of its existence.

In English, native speakers use the three different articles to denote how various combinations of these factors apply to certain nouns.

The Indefinite Article[3]. To begin, we will consider two uses of the indefinite article *a/an.* These uses of the indefinite article are illustrated in examples (11) and (12):

11. *(Non-referential indefinite noun phrase)* You must bring *a cat* to the annual meeting of the Cat Fanciers' Association. [-SR +HK]

[3] Note that the terminology in this section was adopted from Ionin & Wexler (2002). See their article for a fuller discussion on the topic, and their interpretation of how it relates to ESL learners' article use, pp. 151-153.

The noun phrase *a cat* in example (9) is what is called **non-referential [-SR].** It does not refer to a particular cat—any cat will do. As such, it does *not* have specific reference. Non-referential noun phrases refer to one non-specified instance of a category (above, you must bring one [currently unspecified] cat to the meeting). Therefore, non-referentials occur with singular count nouns only (a cat, a job, a plan). However, as the hearer of such a sentence would have assumed that cats in general exist, such a noun phrase has hearer knowledge. An additional example would be this sentence from the example of argumentative discourse style (example 6):

12. "We recommend having *a security system* installed, certainly." [-SR +HK]

The hearer of this sentence would have assumed that security systems exist, but the speaker of the sentence is not referring to a specific system. Now consider example (13):

13. *(Referential indefinite noun phrase)* My friend has *a cat* that eats from my vegetable garden. [+SR -HK]

The version of *a cat* in example (13) is **referential [+SR]:** it picks out a certain cat and means to refer to that one in particular; thus, it has specific reference. In these cases, a single instance of a category has been identified, but the hearer does not have any preexisting assumptions that there exists a cat that eats from the speaker's garden. However, whether or not an indefinite noun phrase is non-referential or referential is not always clear within a single sentence.

14. *(Ambiguous definite noun phrase) A cat* sits outside my porch door every morning.

Example (14) is ambiguous. *A cat* could refer to the same cat, every morning, or to a different cat, every morning. We need more information to make a decision.

15. *(Non-referential indefinite noun phrase) A cat* sits outside my porch door every morning. Every day there's a different one. [-SR +HK]

16. *(Referential indefinite noun phrase) A cat* sits outside my porch door every morning. It has markings like spectacles around its eyes [+SR –HK].

Example (15) is non-referential; example (16) is referential. This distinction is an important one to grasp. While both types of noun phrase use the indefinite article, non-referentials refer to an entire category of entities, while referentials refer to one instance of a category of entities.

Indefinite articles, then, occur in three combinations: [+SR –HK] as in examples (8, 13, 16), [-SR +HK] as in examples (9, 11, 12, 15), and [-SR –HK] as in example (10).

The Definite Article. To discuss the definite article, it makes sense to return to example (14). We might have continued the first sentence, "A cat sits outside my porch door every morning," in the following way:

17. A cat sits outside my porch door every morning. *The cat* has markings like spectacles around its eyes [+SR +HK].

In example (17), the definite article is used the *second* time the cat is mentioned. This is because it has already been referred to once, and by including *the,* we have confirmed some new information: a) that it is one cat and one cat only, and b) it is a specific cat, unique in whatever conversation we are having, story we are writing, etc. Since this is the second time the cat has been mentioned, it now has specific reference and hearer known [+SR +HK], whereas in the original sentence, we could not be sure whether *a cat* was specific. *The* will continue to crop up every time we mention this cat in the discourse. Here is a further example of the definite article used in this way, drawn from the sample of narrative discourse (example 5):

18. Glass from a window was everywhere; *the window* had been smashed.

In addition to the well-known "second mention" use of *the,* English uses the definite article in several other contexts. For example, it is used to describe something that is, by nature, one of a kind:

19. *(Unique by nature)* In *the future,* we will most likely be ruled by cats [+SR +HK].
20. *(Unique by nature)* He might be the only fan of cow-tipping on *the planet* [+SR +HK].

Because there can be only one *future* and one *planet,* they are inherently referential and the hearer is always presumed to understand what is being referred to. In other instances, *the* is used in what has been termed "cultural" contexts. These uses are "determined, to a large extent, by conventional practice" (Liu & Gleason 7), and thus tend to be arbitrary and unpredictable. As a result, learners typically have to memorize them.

21. *(Cultural context)* Eli cared for his girlfriend, Lilith, when she got *the flu* last month [-SR +HK].
22. *(Cultural context)* I play *the piccolo* very well. [-SR +HK].

Finally, *the* is sometimes used to describe a race of people, a species, or the people of a nation.

23. *(Species) The Giant Panda* is very near extinction. [-SR +HK]

This use of *the* is special, because it is not referential. As a result, instances such as those in (23) above also commonly called **generics**, which refer to most or all members of a particular group of things (a category)**.** As such, they are automatically considered hearer-known—the

hearer can be assumed to have previously acknowledged the existence of categories of things. Using *the* in this way is relatively rare outside of a scientific context (Liu & Gleason 6); for instance, native speakers would not say anything like:

24. *The book is an enjoyable way to pass time.

Because they are so specialized, this type of generic is not discussed in the context of this project.

Much of this information seems self-evident to native English speakers, but the information that is coded in our articles can be very difficult to grasp for non-native speakers, particularly for those who speak a first language with no articles whatsoever. Most importantly for the current discussion, *the* is used with nouns that are both referential and hearer known, such as nouns that are mentioned for the second (third, etc.) time in a discourse; the exception is the uncommon generic *the* for species, races, or peoples.

The Zero Article. For some time, nouns occurring with a zero article were simply called bare nouns (Master 1997), and sometimes still are. Generally speaking, the zero article is used most often with nouns that are either mass nouns (oatmeal, sand, water) or are pluralized count nouns (boxes, jaguars, bathtubs). A key point to remember is that with the zero article, a noun can be either generic or referential, whether it is mass (examples 25 and 26) or pluralized count (examples 27 and 28). Notice that the generics, like those that sometimes occur with *the,* are not referential but are hearer-known. The referential nouns, having just been introduced to the discourse, *are* specific but are not yet hearer-known.

25. *(Zero article with generic mass noun).* ∅ *Sand* can be fine and soft or coarse and gritty. [-SR +HK]

26. *(Zero article with referential mass noun).* After my cat ran on the beach, he was covered in ∅ *sand.* [+SR –HK]

27. (*Zero article with generic plural count noun).* Sometimes, burglars will break in through *windows*. [-SR +HK]

28. *(Zero article with referential plural count noun)* I recently had new *windows* installed in my home. [+SR -HK]

Notice that the zero article can sometimes occur with a referential noun, just like a referential indefinite noun phrase. To make this clearer, insert *some* before the relevant nouns in examples (26) and (28)—the meaning of the original sentences is preserved. This is because *some* can only be used with nouns that are referential but are not hearer known, like referential indefinite noun phrases. Once the noun does become hearer-known (it is mentioned a second time), having *some* sounds odd. Imagine continuing (28) with, "Some windows were look great."

By contrast, inserting *the* into (25) or (27) sounds awkward to native speakers because generic nouns are not referential (except in those rare, specialized cases).

Article	**A/An**	**The**	**Zero (Ø)**
Is used to…	-Introduce new referential nouns to a discourse [+SR –HK] **"*A woman* came home one day to find that her house had been robbed."** -Refer to a single, nonspecific member of a category [-SR +HK] **"You must bring *a cat* to the meeting."**	-Refer to old referents in a discourse [+SR +HK] **"Glass from a window was everywhere; *the window* had been smashed."** -Refer to unique things and/or things that definitely exist [+SR +HK] **"He seemed to fall off *the planet."*** -Refer to culturally unique entities [-SR +HK] **"I have *the flu."*** -Refer to species, races, groups of people (rare) [-SR +HK] **"We must protect *the orca* from extinction."**	-Refer generically to all members of a category [-SR +HK] **"Preventing home *burglaries* requires preparation in advance."** **"*Salt water* is not good habitat for salamanders."** -Introduce referential plural count or mass nouns to a discourse [+SR –HK] **"We found *some muddy footprints* on the floor."** **"He was covered in *flour."***
With…	-Singular count nouns	-Singular or plural count or mass nouns	-Plural count nouns -Mass nouns

Table 1. Semantic and Morphological Distribution of the English Articles

Table 1 above summarizes the features of the articles as they have been discussed here.

Note that while the definite article can sometimes be used generically, most often that use is granted to the zero article. However, we cannot forget the **non-referential indefinite noun phrase,** as in "You must bring *a cat* to the annual meeting of the Cat Fanciers' Association." Semantically, both non-referentials and zero article generics are not referential but *are* hearer-known ([-SR +HK]). The difference is morphological, in the type of nouns that occur in each phrase: non-referentials are always singular countable nouns with indefinite *a/an*, referring to one non-specified instance of a category. Generics that are marked with the zero article are either mass nouns or pluralized countable nouns. Therefore, this paper will make a distinction between **generic nouns** like *sand* and *windows* in (25) and (27), and **non-referentials** like *a cat* in (11). Table 2 summarizes some of the key terms introduced in this section, for later reference.[4]

Term	Definition
Indefinite	Not (or not yet) having uniqueness and/or existence within a given discourse context
Definite	Having uniqueness and/or existence within a given discourse context
Referential	Referring to a specific entity
Non-referential	Not referring to a specific entity; referring to one nonspecific instance of a category. A singular count noun phrase, e.g. *a security system* (10)
Generic	Referring to a category of entities; not referring to a specific entity. A mass or plural count noun phrase e.g. *windows* (25*), sand* (26)

Table 2. Key Terms for Articles and Noun Phrases

[4] Further overview of the English articles can be found in the handout on articles produced by the Writing Center at the University of North Carolina at Chapel Hill. http://writingcenter.unc.edu/handouts/articles/

1.3 Errors of Omission vs. Errors of Overuse

An **error of omission** is the scenario opposite to overuse or insertion; an ESL writer will fail to provide an article where one is required. Examples (29) and (30) illustrate this (the sentences contain several errors; focus on the italicized noun phrases).

29. Part-time job also can help the student whose family have not much money to support *a student* to go to college.

30. It will make us college students more confident and mature for *future.*

The noun phrases emphasized in the sentences above required a definite article. In example (29), *a student* should have instead been *the student,* because the noun in question had already been mentioned once. In example (30), because "the future" is naturally unique and referential, it always requires a definite article, which the writer here omitted.

Now, recall the examples of definite article overuse given at the end of the first section (focus only on the errors involving the definite article):

31. *When we come to *the college*, almost every new student will be into *the situation.*

32. *Maybe study *the knowledge* from *the books* well is important.

33. **The good score* is the precondition for a good job.

34. *They can put what they have experience and get to know *the society*.

In certain of the above cases, an indefinite article was necessary instead, or a zero article (or nothing at all). These sentences sound unidiomatic because the unnecessary definite articles send informational signals to the brains of native speakers, signals that suggest things about the context that are not true. As discussed earlier, definite articles tell us that a noun phrase is

referential and unique, either by nature or because it has already been mentioned in the discourse. These characteristics do not apply to the nouns emphasized in examples (31-34).

In example (31), *college* is used to refer to college in general, not to a specific college as the definite article suggests. In the text from which example (31) is drawn, this sentence is the first. Therefore, it is also incorrect to use *the* with the noun *situation,* as the writer has yet to specify which situation they mean. In Example (32), both *knowledge* and *books* are meant to be generics—note that *knowledge* is a mass noun and *books* is a pluralized count noun, both types of nouns which should occur with the zero article. In example (33), *score* is a singular count noun; thus, if the writer wanted to refer to the general category of good scores, she needed to use an indefinite article and thus create a non-referential indefinite noun phrase. Finally, in example (34), *society* is an example of a singular count noun that can occur without any article and still have a non-referential meaning. Therefore, including the definite article converts it into a referential noun, as would including an indefinite article.

Accounting for errors of *the*-overuse has been both a question for researchers and challenge for teachers of English, who must find a way to explain seemingly idiosyncratic distribution of the definite article to students from diverse language backgrounds. An explanation that applies to one group of learners or one linguistic context may not apply in alternate situations. Essentially, most studies of this phenomenon attempt to describe the knowledge, or linguistic competence, that underlies what these ESL learners actually produce—their performance.

This performance may be influenced by many factors: the environment in which the speech/writing is produced, the type of writing, the timing, the intended audience, the formality of the writing, the style in which the students were taught, etc. One of the most important factors is the type of communicative task being attempted by the writer: not only in English, but in other

languages, the language (in terms of words and structure) we use will vary depending on whether or not we are narrating a story or arguing a point. This is a crucial issue, as sometimes a speaker's first language has expectations for what kind of language to use in a certain situation that do not match up with the expectations of their second language. Trying to establish a theory that accounts for linguistic performance regardless of context is strikingly difficult, and one must be careful not to over-apply findings. Many of the previous studies that have been done on the topic of ESL article use base their findings on narrative uses of *the,* while the present study is primarily concerned with the question of the argumentative rhetoric that challenges so many Chinese ESL college students. Additionally, testing students in only one type of discourse context will necessarily produce results that reflect the input, because some types of nouns are more likely than others to occur in certain discourse contexts; an additional aspect of their understanding may be lost because certain nouns are not represented. Furthermore, priming students with specific instructions making them alert to the definite article will not produce an entirely accurate reflection of their competence, because they will specifically attend to that grammatical issue. That being said, considering the implications of these results is important to understanding how the present study can complement what work has already been conducted in this field of research.

The null hypothesis of the present paper is that Chinese ESL speakers' overuse of the definite article is random and generalized, not adhering to any predictable pattern. This hypothesis implies that these speakers do not have any knowledge of appropriate article use. The eventual aim of this study is to refute this hypothesis and establish finer distinctions (and thus specific causes) within *the* overuse. However, previous work on this topic has already begun to

extricate trends from a variety of kinds of data, with a particular emphasis on the semantic framework of [±SR ±HK] described by Bickerton.

2.0 Previous Research on ESL Speakers' Article Use

2.1 Master

Peter Master, in 1987, conducted a study of 20 non-native English speakers coming from two different language backgrounds: +article and –article. This means that some of the participants natively spoke languages with article systems (e.g. Spanish), and some natively spoke languages without article systems (e.g. Chinese). He grouped these speakers into different levels of English proficiency and quantified their *spoken* use of the three English articles. For the definite article, he explains that "once [-ART] learners have realized that Ø is not always appropriate, their first hypothesis appears to be that all nouns require *the. The* is thus overused at first […]*"* (Master 4). This phenomenon has been termed "*the*-flooding" (Wong & Quek 230); it describes how ESL speakers of –article languages will, apparently, insert *the* into numerous, varied contexts, including those in which no article is necessary at all. This occurs in post-low-proficiency levels, flaring particularly at the intermediate level.

Master theorized that, once learners begin to realize that *the* is not appropriate in all contexts, they begin to pull back from flooding. Even so, however, overuse persists among higher levels. Master postulates that ESL learners associate *the* with the hearer-known aspect of nouns; thus, when they believe that a noun should be understood by the hearer in a discourse, they will insert *the* (Lu 3). This might occur, for instance, with generic mass or pluralized count nouns that take the zero article. However, this idea has come under some criticism from researchers who point out that Master's findings are, in fact, "short of support" (Lu 5) from generic noun phrases, relying more on other +HK contexts.

Furthermore, there is a well-attested divergence between the second-language skills of speaking and writing. Each is acquired differently and at a different pace; each tends to reflect different sets of knowledge that together form an ESL speaker's general English competence. In other words, performance will vary according to not only the skill under study, but also according to the format in which it is being studied, as "different tasks have been shown to cause variation in L2 accuracy" (Liu & Gleason 15).

The notion of an omnipresent "flood" of definite articles merits additional study; Master himself argues that ESL learners insert unnecessary *the* when the noun can be understood to be hearer-known, although the basis for this has been called into question. There are additional variables that could shed light on the situation; for instance, other researchers have suggested that referentiality (a.k.a. specific reference) is in fact what gives learners trouble.

2.2 Ionin & Wexler

Some researchers have suggested that *the*-overuse is a result of learner confusion about the distinctions of referentiality outlined in section 1 of this paper. Recall **referential** indefinite noun phrases, as in example (13), repeated here:

35. *(Referential indefinite noun phrase)* My friend has *a cat* that eats from my vegetable garden. [+SR –HK]

This type of indefinite noun phrase refers to a particular individual entity that has just been introduced to the discourse, so that it is referential but is not yet hearer-known. Meanwhile, **non-referential indefinite noun phrases,** as in (36), refer to one unspecified member of a category by using a singular count noun, and thus are not referential but *are* hearer-known:

36. *(Non-referential indefinite noun phrase)* You must bring *a cat* to the annual meeting of the Cat Fanciers' Association. [-SR +HK]

In a 2003 study, Tania Ionin and Ken Wexler posited that referentiality is one aspect of a noun's meaning that could be encoded in the English definite article[5] (Ionin & Wexler 150), and that it represents an alternative to definiteness that might be an option for an ESL speaker who is trying to determine when and where to use *the.* Therefore, they predicted that ESL speakers of a language with no articles (their subjects spoke Russian) would "associate *the* with referentiality" (Ionin & Wexler 153). They gave their subjects two types of tasks: translation, in which they translated a sentence in their native language into English, and elicitation, in which the speakers read a series of dialogues and inserted the appropriate article into the final sentence. In their results, they did find that learners tended to overuse *the* more with referential indefinite noun phrases, as in:

37. She wanted to see *a certain painting (*Ionin & Wexler 154). [+SR –HK]

The subjects were more likely to render (37) as *She wanted to see the certain painting* than they were to insert *the* in a context like that below:

38. I want to have *a dog.* [-SR +HK]

This trend was particularly strong when the noun phrase was modified with a relative clause, as in *She wanted to see a certain painting, which featured a smiling maiden* (Fodor & Sag, Ionin & Wexler).

[5] Ionin and Wexler propose that referentiality and definiteness are two mutually exclusive options for the definite article's function cross-linguistically. They suggest a universal grammar hypothesis in which speakers with no articles in their first language must decide which of these two features the L2 definite article codes for.

But does this pattern hold up with native Chinese speakers? Marta Tryzna conducted a similar study with L1-Chinese speakers, and found that they also tended to significantly overuse *the* when the indefinite noun phrase was referential. However, she found that the trend only persisted with singular NPs; in the plural context, *the* was overused with generics, instead (Tryzna 78).

Since these referential nouns are not hearer-known, there is a conflict between this theory and the theory of Master's camp. It may be that neither one of these factors—specific reference and hearer knowledge—can neatly account for the entire phenomenon. As I have explained, learner performance depends on many variables. In both Master's and Ionin and Wexler's cases, the format of the study will have had an influence on their results. Master's study observed spoken speech. Ionin and Wexler used translation and elicitation tasks. Noting their contrasting research findings allows us to focus more sharply on the variation we find and better pinpoint the exact circumstances that lead to *the*-overuse.

2.3 Liu and Gleason

A study conducted by Dilin Liu and Johanna L. Gleason begins to examine the way that non-native speakers of English (not solely native Chinese speakers) use the definite article in various specific contexts. They presented 128 non-native, student speakers of English with a 91-sentence survey and asked the participants to insert *the* "wherever they deemed it necessary" (Liu & Gleason 10). The survey, or instrument, was composed of sentences either with an obligatory *the* deleted, or control sentences (in which no *the* was necessary) (Liu & Gleason 9).

The researchers then quantified the number and type of *the*-omissions and *the*-overuses for each proficiency level. It is interesting to note that they considered the overuse of *the* to be

"unexpected" (Liu & Gleason 11)—they primarily predicted that their participants would underuse *the,* rather than overuse it.

What they found was that those strange, rather arbitrary cultural uses constituted the greatest portion of overuse errors. Students tended to most frequently commit errors such as "she plays *the* basketball" or "Sally Ride was the first woman in *the* space" (Liu & Gleason 25). After these errors, overuse of *the* with generics [-SR +HK] were most frequent: "Someday, I think that *the* computers will replace people everywhere" and "*The* rocket ships are launched from Cape Canaveral in Florida" (Liu & Gleason 25).

This test was given to speakers from a variety of L1 (first language) backgrounds, although when the researchers compared the performance of speakers of Indo-European languages and non-Indo-European languages (which include Chinese), they found no significant differences.

Bee Eng Wong and Soh Theng Quek performed (for all intents and purposes) the same study, but with a group of participants limited to Chinese speakers and Malay speakers. Commensurate with Master's *the*-flooding, they noted that "there was a substantial increase in the unnecessary use of *the* by the intermediate proficiency level learners […] The levels of overuse of this definite article then dropped for advanced proficiency level learners" (Wong & Quek 228). However, despite this eventual increase in accuracy, the researchers claim that "the results of [their study] seem to suggest that overuse of the definite article *the* remains a problem for the advanced proficiency group" (Wong & Quek 230).

In contrast to Liu and Gleason's finding, Wong and Quek's results indicate that their speakers *underused* the definite article in cultural contexts. ESL learners do not, according to these results, tend to overuse *or* tend to underuse cultural nouns; it is quite possible that their insertion or lack thereof of *the* is the result of random guessing or a failure to memorize the

arbitrary formulas. This may be a weakness of this type of experiment design, as it obligated the subjects to deal with cultural nouns, which they are likely to avoid when producing their own, original writing. This is particularly salient, given the fact that "such use of *the* [with cultural nouns] is often not framed by situation, but is determined […] by conventional practice" (Liu & Gleason 7). If we remove fixed expressions and other "just because" formulas, we are left with a distribution pattern that may give us a better picture of the state of these speakers' grammatical knowledge. In other words, forcing ESL speakers to grapple with certain kinds of nouns will necessarily produce results that reflect what they were given.

The more intriguing aspect of their findings is that generic nouns appeared to give the subjects trouble. As generic nouns are [-SR +HK], these findings align with Master's theory of *the*-overuse. The notion that hearer-known nouns tend to induce ESL learners to insert *the* remains is a potential explanation. However, in both of these experiments, the students participating were aware that their knowledge of *the* was being tested. The instructions were to insert *the* "wherever they deemed it necessary"(Liu & Gleason 10); having been thus primed, they would have been hunting suspicious, *the*-less phrases. Such transparent directions, under such controlled conditions, could have easily made these participants overly-likely to plug in definite articles because they knew that *some* were missing. Preparing subjects in such a clear and specific way has its experimental benefits, particularly in a study that is focused on acquisition. However, the priming may have affected their performance, which further clouds our ability to judge their competence: their underlying knowledge of how the definite article works. Analyzing spontaneously produced language would add an additional perspective to what has been found with this type of research, producing a more thorough, accurate, complex model of the problem.

3.0 The Present Study: Methodology

In order to construct a picture of *the* overuse based in spontaneous, unprimed language production, the present study relied on a corpus of short essays written in English by university students in China. The corpus used was the Corpus of English Essays Written by Chinese University Students, or CEECUS. It comprises a sub-section of a larger corpus called the Corpus of English Essays Written by Asian University Students[6]; CEEAUS also includes sub-corpora of essays written by Japanese, Korean, and Taiwanese students, as well as native English speakers.

CEECUS consists of 96 short essays, one half written in response to one topic, and the other half written in response to different topic. The topics were: "Is it important for college students to have a part-time job?" and "Should smoking be banned in all the restaurants in China?" The students had a set amount of time to respond to the prompt, and were not allowed to use dictionaries or other outside sources for help.

A research question focusing on article use required that all nouns in the corpus be recorded and coded according to a number of relevant characteristics. Only common nouns were included--proper nouns such as "China" and "Singapore" were not recorded; nor were pronouns. The coding was done in Microsoft Excel to allow for later analysis.

What follows will be a brief description of each of the characteristics for which all nouns were coded in the spreadsheet. All of the coding was done manually.

Text Number. Indicated the essay number from which the noun was extracted, to simplify the process of re-examining source material.

---

[6] Ishikawa, S. (2009). Vocabulary in interlanguage: A study on Corpus of English Essays Written by Asian University Students (CEEAUS). In K. Yagi & T. Kanzaki (Eds.) Phraseology, corpus linguistics and lexicography: Papers from Phraseology 2009 in Japan (pp. 87-100). Nishinomiya, Japan: Kwansei Gakuin University Press. http://language.sakura.ne.jp/s/ceeause.html. Licensed under Creative Commons.

Noun. The noun itself, in whatever form (erroneous or not) the student wrote it.

Sentence/Clause. The actual sentence (or clause, if the sentence was very long) in which the noun occurred.

Singular/Plural. Indicated the singularity or plurality of the noun as it occurred in the corpus.

Ambiguous/Unambiguous. If the noun was always either count or mass (meaning it never occurred in more than one form), it was unambiguous; if a noun could be either count or mass (as in: *freedom, string, thought*), it was ambiguous.

Abstract/Concrete. If a noun referred to something tangible in the real world, it was coded as concrete (*students, restaurants, money*). If the noun referred to intangible things or ideas, it was abstract (*opinion, views, attention*).

Sentence Position. Indicated the position of the noun relative to the verb: **subject**, **object**, or **object of a preposition**.

Discourse Status. Indicated whether a noun was **new** to the discourse, **old** to the discourse, **generic,** or **non-referential indefinite.** Two additional categories were recorded, although they were not considered on par with the other four discourse statuses. Rather, they were sub-categories that merited special attention: ambiguous nouns **in-between** generic and non-referential, and **bare singular generics** (certain special nouns like *college* that can occur as singular nouns without articles). This determination of discourse status aligned with an English perspective of what is considered topic-comment and subject-predicate; the English generic, for instance, is considered the equivalent of the Chinese topic-comment sentence because they share similar contexts.

Article/Plural. Indicated whether or not the writer made an error either with an article or a plural. Both indefinite and definite articles were recorded, with codes either for **insertion** or

omission; the insertion or omission of plurals was also recorded. Errors with zero articles were indirectly recorded: if a different article was inserted, a zero article was therefore omitted, and if a different article was omitted, a zero article was therefore inserted.

Article Type. Indicated the type of article used, whether correctly or incorrectly.

Every noun in the corpus was coded according to each of the above variables, in order to track *the*-overuse and determine whether or not any of these aspects of the noun or its context influenced the likelihood of an ESL writer error.

To briefly demonstrate how the coding worked, we can look at an example. Take, for instance, the noun *job,* which appeared relatively frequently, given one of the essay prompts was on the topic of part-time jobs. This *job* occurred in essay PTJ01, in the sentence, "In my opinion, it's important for students to have a part-time job." Both of these pieces of information were recorded. It was also noted that it was singular and unambiguously count. Additionally, the noun was recorded as abstract and its sentence position is as the object of a verb. The discourse status of this noun is non-referential indefinite noun phrase: it is a singular, indefinite count noun being used to refer to a category of things (namely, part-time jobs). It does not refer to a specific job. Finally, no "error" code was included because there was no error, but the presence of a correctly used indefinite article *a* was indicated.

Once all nouns had been coded for in this way, the following totals were calculated over the entire corpus:

a. *The*-overuse
b. *The*-omission
c. *A/an*-overuse
d. *A/an*-omission

Then, each of the variables was totaled; that is, the number of tokens (nouns) that fit each parameter were calculated, and taken as raw frequencies from the total number of nouns in the corpus. Note that for each variable, the sub-totals of the different expressions of the variable added to 100%. For instance, the total number of nouns of each type of discourse status (new, old, generic, non-referential, ambiguous, bare singular) added to 100%. These totals were taken not to analyze the breakdown of the corpus overall, but instead to calculate the proportion of each variable that occurred with a *the*-overuse error. In other words, 60% of generic nouns may have had a *the* overuse error, while perhaps only 30% of new nouns had a *the* overuse error. Frequencies like these were calculated for each expression of each variable in the corpus.

Chi-square tests were then performed for each set of expressions within a variable, in order to determine statistical significance. Chi-square tests are necessary to account for the weight of each variable in the overall corpus. For example, even if it turns more generic nouns had an error than did new nouns, that might only be a result of the fact that there are simply more generic nouns in the corpus overall. Calculating frequencies as explained above and performing chi-square tests for each variable allows us to ascertain whether or not our results truly demonstrate a significant trend.

4.0 Results

In the entire corpus, the rate of overall *the*-overuse was 9.2%. This means that of all 3,370 nouns in the corpus, 311 occurred with an unnecessary *the.* If this number does not seem substantial, consider it as a frequency compared to the total number of required definite articles in the corpus. By totaling the number of omitted necessary definite articles and the number of correctly used articles, I found that the number of required contexts for *the* in the corpus was 476. The number of actual uses of *the*—the sum of those used correctly and those incorrectly

inserted—was 787. This means that the ESL learners used *the* approximately 165% of the time; 65% more than it was necessary. Compare this to the rate of overuse of the indefinite article, *a/an,* which came to only approximately 109% of the time, or 9% more often than necessary. From these findings alone, we can establish that overuse of *the* is not a problem isolated to narrative or experimental contexts. But does the patterning differ in an argumentative context? And if so, how?

We now turn to our variables. Discourse status, with its implications of referentiality and hearer knowledge, should be considered first, as previous research could not agree on which, if either, factor had more influence on speakers' overuse. Table 3 displays the rates of error for each of the four types of discourse status recorded in the corpus. Below the table are examples of each type of noun occurring with an unnecessary indefinite article.

Error	**Non-referential Indefinite**	**Generic**	**New (Referential Indefinite)**	**Old (Second Mention**
***The*-overuse**	29 (6.1%)	158 (7.8%)	19 (3.7%)	0 (0%)
No *the*-overuse	462	1864	491	126
Total	491	2022	510	126

Table 3. Rates of *the*-overuse for discourse status χ^2=20.932, p=0.00010876

39. **Non-referential indefinite:** no matter how the meeting is important, they are always late. [PTJ 09]

40. **Generic:** In a word, a part-time job is important for *the students*. [PTJ 01]

41. **New:** I even can't feel my sister's change after she did *the part-time job*. [PTJ 07]

Generic nouns had the highest rate of *the*-overuse at 7.8%, followed closely by non-referential indefinite nouns at 6.1%. The rates for new, or referential, and old nouns are lower,

at 3.7% and 0.0%, respectively. Note also how the overall numbers of new and old nouns, which require writers to keep track of information in the discourse, are relatively low when compared to generics; this lines up with the notion that such types of nouns occur less often in topical-argumentative discourse styles.

Remember that both non-referential indefinite noun phrases and generic noun phrases are not referential but are hearer-known [-SR +HK]. A chi-square test just between these two values yields results that are not statistically significant, with a p-value of 0.19426.

It should be further noted that there was a fifth category of nouns: nouns that were ambiguous between non-referential indefinites and generics. Nouns were marked as in-between if their status as singular or plural could not be determined from context due to the lack of a conjugated verb or other indicator (as this is the defining morphological difference between non-referentials and generics). For instance, if it was evident that a required plural marker was omitted from a noun or if a verb did not agree with its subject noun, the noun was considered ambiguous in status. Because definite articles were often used with ambiguous nouns, they served to further obscure these nouns' singular/plural status (definite articles can occur with both). Out of 79 ambiguous nouns in the corpus, 49 included an unnecessary *the,* for a rate of 61%. These nouns will be considered separately in discussion. I will deal primarily with nouns that are unambiguously non-referential or generic.

Because the non-referential indefinites and the generics differ from one another primarily morphologically, the generic nouns themselves were further broken down into the two different types of nouns that can have this label: mass nouns, and pluralized count nouns. Each total consists of a sub-total within generic nouns overall. The figures are displayed in table 4, followed by illustrative examples from the corpus.

Error	Mass Generic	Plural Count Generic	Bare Singular Generic
***The*-overuse**	59 (5.2%)	93 (11.1%)	56 (41.5%)
No *the*-overuse	1083	747	79
Total	1143	840	135

Table 4. Distribution of *the*-overuse by morphology of generic nouns, χ^2=182.3 p=0

42. **Mass Generic: […]**can't be learned from the courses we take in the school, but only can learn from *the practice*. [PTJ 12]

43. **Plural Count Generic:** Yes, we have *the no-smoking restaurants*. [SMK 09]

44. **Bare Singular Generic:** With the development of *the society*, Smoking has became a world problem. [SMK 44]

Bare singular nouns are those strange nouns that are semantically countable, like colleges and schools and classes, but in certain contexts can appear with neither an article nor pluralization, a freedom normally forbidden to count nouns. Hence, we get expressions like *I'm in class* or *I can hardly afford to attend college.* These nouns have a [-SK +HK] connotation, and thus fall semantically under the generic heading, but morphologically, because they are "bare," it is pertinent to see how they measure up when compared to the other manifestations of generic nouns, those we see more typically: mass and pluralized count. As it happens, the funny bare singulars have a much higher rate of *the*-overuse than do the other categories, at 41.5%. The next highest rate belongs to pluralized count nouns, at 11.1%. Finally, generic mass nouns have a rate comparable to those of the new nouns, at 5.2%. The statistically significant differences here can tell us something about what precisely about generic nouns may give students so much trouble. It pays, then to report further characteristics of generic nouns here.

Error	Unambig. Generic	Ambiguous Generic	Unambig. Non-Ref	Ambig. Non-Ref
***The*-overuse**	116 (8.7%)	42 (6.1%)	24 (6.4%)	6 (7.4%)
No *the*-overuse	1212	651	360	102
Total	1328	693	383	106

Table 5. Unambiguous vs. Ambiguous Generics and Non-Referentials
Generics χ^2=4.519 p=0.03352
NR χ^2=0.031, p=0.86024

45. **Unambiguous Generic: t**his will also lighten the financial pressure from *the parents*. [PTJ 45]

46. **Ambiguous Generic:** a part-time job will cost *the physical power*. [PTJ 27]

47. **Unambiguous Non-Referential:** no matter how *the meeting* is important, they are always late. [PTJ 09]

48. **Ambiguous Non-Referential:** spending all the time on study in *the room* is not a wise choice. [PTJ 36]

Table 5 illustrates the rates of *the*-overuse for unambiguous nouns and ambiguous nouns, within both generics and non-referentials. Of generic nouns, unambiguous nouns (those solely count or solely mass) had the higher rate, at 8.7%, not quite significantly higher than the rate of error for ambiguous generics. For non-referentials, ambiguous nouns had the higher rate, and there was no significant difference between the rates for unambiguous and ambiguous.

Error	Concrete Generic	Abstract Generic	Concrete Non-Ref	Abstract Non-Ref
***The*-overuse**	89 (15.3%)	69 (4.7%)	6 (6.3%)	24 (5.8%)
No *the*-overuse	494	1414	74	388
Total	583	1483	79	412

Table 6. Concrete vs. Abstract Generics and Non-Referentials
Generics χ^2=66.3741, p=0
NR χ^2=0.03, p=0.86249

49. **Concrete Generic:** take *the airports* and the cinema for example. [SMK 13]

50. **Abstract Generic:** About *the working experience* or *the society experience*, I think these are ought to be realized after graduation. [PTJ 14]

51. **Concrete Non-Referential:** the smoke when *the cigarette* burning contends nearly all the poisons that the cigarette has. [SMK 03]

52. **Abstract Non-Referential:** I think *the part-time job* is necessary for college students to have more chance to know the situation of themselves. [PTJ 08]

Table 6 displays the rates of error by concreteness or abstractness of nouns within generics and non-referential indefinites. Concrete nouns refer to entities that exist as tangible things within the world, while abstract nouns refer to things that do not. Again, there was only a statistically significant effect within generic nouns, with concrete nouns showing more than triple the rate of *the*-overuse of abstract nouns. Within non-referentials, there was no significant difference between the two rates of error.

Finally, sentence position was taken as a factor in determining the patterns of *the*-overuse for generic and non-referential indefinite nouns. Table 7 displays the data for the generics:

Error	**Subject Pos. Generic**	**Object Pos. Generic**	**Obj. of Prep Pos Generic**
***The*-overuse**	27 (5.1%)	35 (5.0%)	96 (12.2%)
No *the*-overuse	498	672	693
Total	525	707	789

Table 7. Error Rates of Generic Nouns by Sentence Position, $\chi^2=33.99$, $p=4e^{-8}$

53. **Subject Pos. Generic:** *the companies* also want the college students to have more ability. [PTJ 09]

54. **Object Pos. Generic:** we are students, we have *the severe competition* […] [PTJ 40]

55. **Obj. of Prep Pos. Generic:** there is more knowledge which can not be learned from *the textbooks*. [PTJ 24]

The error rates for nouns in subject and object position were comparable, 5.1% and 5.0% respectively. The rate for nouns as objects of prepositions, however, was significantly higher, at 12.1%. Compare this pattern of findings to those of non-referential indefinites, in table 8 below:

Error	**Subject Pos. Non-Referential**	**Object Pos. Non-Referential**	**Obj. of Prep Pos. Non-Referential**
***The*-overuse**	11 (15.1%)	9 (3.3%)	10 (7.0%)
No *the*-overuse	63	267	133
Total	73	276	143

Table 8. Error rates of Non-Referential indefinites by Sentence Position, χ^2=14.04, P=0.000894

56. **Subject Pos. Non-Referential:** For example, *the student* who is major in computer takes a part-time job in a computer company. [PTJ 15]

57. **Object Pos. Non-Referential:** no people like *the man* whose mouth is bad smell [SMK 17]

58. **Object of Prep Pos. Non-Referential:** In fact, I think it is very rude to smoke in *the restaurant*. [SMK 38]

The rates by sentence position for non-referentials were more evenly distributed than those found for generics, with clearer "low," "mid," and "high" groups. However, in this case, the highest rate was found in subject position, with 15.1% of these nouns occurring with an unnecessary *the.* Sentence position is the only factor for which non-referentials demonstrate a statistically significant distribution, and the fact that the patterning differs from that of generics is something to be emphasized. This—and all these findings—will be explored in the discussion.

To summarize, the results indicate that there is an overall 9.2% rate of *the*-overuse within the corpus, as students used *the* 65% more than it was necessary. When broken down by discourse status of the noun, it was found that generic nouns and non-referential indefinite noun phrases had the highest rates of error: 7.8% and 6.1%, respectively. As there was no statistically significant difference between these two values, both groups were analyzed in terms of

unambiguity/ambiguity, concreteness/abstractness, and sentence position. For generics, those nouns that were unambiguous had the higher error rate at 8.7%, as did concrete nouns, at 15.3%. For non-referentials in these categories, no statistically significant differences were found. In sentence position, however, while generics in objects of preposition position had the highest error rate, non-referentials had the highest error rate in subject position.

5.0 Discussion

The data collected from CEECUS have yielded results revealing an aspect of definite article overuse not previously explored in the field. By studying the argumentative, academic output of Chinese ESL writers, we can see how the context of a task influences student performance; we cannot construct the full picture of learner knowledge without exploring its manifestations in different contexts.

Despite the fact that these are college-level learners, the roughly 10% overall rate of error suggests, as Wong and Quek asserted, that the definite article remains a source of difficulty for learners even at this stage (Wong & Quek 228). This certainly requires those of us who want to teach the English language to understand that the rules of article use are anything but hard and fast. Those learners who want to achieve a higher degree of idiomaticity will need to grasp not only the basic rules of using *the,* but also the relative frequencies of different *the* constructions (like the fact that generic *the* is rather rare), the special cases arbitrarily determined by culture, and the precise circumstances of discourse that are more likely to require nouns with definite articles.

The fact that generic nouns demonstrated the highest rate of error (and that they were the most frequent type of noun) aligns with my hypothesis that a written argumentative discourse would influence the type of nouns with which definite articles were overused. This weakness in

learners' understanding of definite article use would not have been apparent in studies focusing on narrative tasks, because generic nouns do not occur as frequently in such contexts.

Writing essays is a core component of the American higher education system. In my own experience with second language classrooms, however, learning how to structure and compose essays with rhetoric appropriate to that discourse style only became a focus in specialized courses at the higher level; most of my years of study consisted of conversation practice, description, and the juggling of tenses and anaphora required in explaining stories or situations to others. I am not suggesting that this is a flawed approach, but rather that it was aimed at achieving what is called **communicative competence**: essentially, allowing me to function in the day-to-day life of the culture of the native speakers.

It is possible that the curricula of Chinese English classrooms, too, focus on communicative competence and de-emphasize written proficiency. As it happens,

> "many people [in China] hold the view that 'communicative ability refers to spoken ability' and this has lead to the phenomenon whereby oral ability in teaching practice has changed from being ignored to being emphasized too much. The fact is, however, that most Chinese learners use English more in written forms." (Xu 754)

Xioayan Xu, reviewing Dingfang Shu's book *FLT in China: Problems and Suggested Solutions,* also explains how studies have found that as many as 70% of English teachers in China "are still using the traditional grammar-translation method" in their classes (Xu 755), which involves learning grammatical rules and creating sentence structures based on those rules

(Yu 195). Grammar-translation, and the more recently instituted communicative language teaching (CLT) method (Yu 195), may not be not easily applied to the notion of higher-order, stylistic contextualization and argumentation.

Certainly not all Chinese students study with the eventual goal of attending American universities; if learning to compose written arguments in the appropriate register of English is not emphasized in the classroom, then learners may fail to understand that the distribution of nouns in an essay—and the way articles interact with them—differs from how nouns are used in speech and narratives. I suspect, then, that pure unfamiliarity with generic nouns could be one reason why Chinese learners overuse *the* with them.

There are, though, further aspects to consider. Remember, for instance, that non-referential indefinite noun phrases demonstrated an error rate that was not significantly different from that of generic nouns. Both types of nouns have the same semantics in terms of Bickerton's framework: [-SR +HK]. While one could argue that these findings support Master's hypothesis that learners associate *the* with hearer knowledge, they could also be interpreted as suggesting that learners associate *the* with a lack of referentiality. This lack of referentiality is what makes both types of nouns semantically generic; the difference between them is morphological, because non-referential indefinite noun phrases are always singular count nouns, while generics are mass nouns or plural count nouns.

Again, I believe that discourse context can may explain the relatively high error rate of non-referentials. Arguments require writers to comment on topics and make broad statements, which is why non-referential nouns tend to prevail—they are generalizable, lending themselves well to the acts of persuasion and thesis-writing. The Chinese language itself maintains a distinction between topical-argumentative styles and narrative styles, although it is reflected not

only at the discourse level, but also at the sentence level. Subject-predicate sentences, which are "used for narrative purposes" (Po-Ching & Rimmington 109), permit only definite nouns in the subject position. By contrast, topic-comment sentences are "designed for descriptive, explanatory or argumentative purposes" (Po-Ching & Rimmington 111). Topic-comment sentences consist of a topic (some given information) and then a comment about the topic. The topic is presumed to be hearer-known, so that any noun in topic position can be "interpreted as definite or generic" (Wong & Quek 213). Here, they also include non-referential indefinite nouns in their definition of "generic." Definite nouns are [+SR +HK], while generic nouns are [-SR +HK]. Below are three examples of topic-comment sentences, taken from Po-Ching and Rimmington's grammar of Chinese (111):

59. Gōngjù yīnggāi fàng zài zhèr.
Tool should put at here
The tools should be placed here. [±SR +HK]

60. *Zìdiǎn hěn yǒuyòng.*
Dictionary very useful.
Dictionaries are useful. [±SR +HK]

61. *Yī gè rén bù néng bù jiǎng lǐ.*
One person not able not talk reason
A person must be reasonable. [-SR –HK]

The topic in (59) (*gōngjù*, "tools") is definite [+SR +HK]; however, in a different context, it could also be interpreted as generic (simply "tools"). Similarly, while (60) is an example of a generic topic, it could also be interpreted as definite in a different context. Finally, the topic in (61) (*yī gè rén,* "a person") is indefinite—specifically, a non-referential indefinite noun phrase [-SR +HK]. Therefore, in an argumentative or explanatory context, topics in Chinese can be definite, generic, *or* non-referential indefinites.

Syntactically, these topic-comments in Chinese are similar to the relatively uncommon English construction that allows native speakers to say things like, "My cottage I love, but your cottage I despise." However, I am less concerned with the syntactic structure of topic-fronting in both languages, and more interested in the fact that in Chinese, topic-comment sentences are more likely to occur in argumentative or explanatory contexts than are subject-predicate sentences. Because topic-comment sentences allow for elaboration on an idea (Shi 388), they may be part of the Chinese language of generalization. In English, the language of generalization is typically encoded with generic and non-referential indefinite nouns—two of the three types of noun phrases most often used in argumentative discourse contexts in Chinese.

Therefore, because Chinese permits topics to be both definite (be referential) and generic/non-referential (not be referential), I believe that Chinese speakers may be more likely to confound the two types when arguing about or explaining a topic in English. Imagine being a Chinese speaker, coming from a language background that allows a choice between two semantic types of nouns as a topic. In other words, the first language predisposes the speaker to associate definiteness with argumentative constructions, as well as genericity. When trying to make argumentative statements in English, the speaker may be more inclined to use definite nouns, and hence overuse the definite article, than would a native English speaker.

Furthermore, since both generic nouns (like the plural count noun in (60)) *and* non-referential indefinite nouns (like *a person* in (61)) are both permitted alongside definite nouns in topic position, this explanation can account for the fact that both types exhibit relatively high rates of *the-overuse*.

Morphological distinctions, however they may be blurred in the above situation, do remain relevant to the patterning of errors in the corpus. In particular, they appear to play a

crucial role in the breakdown of *the*-overuse within generic nouns (see Table 4). Non-referential indefinites lack morphological variation: they are always singular count nouns with an indefinite article. However, their sister-category, generics, does permit several manifestations: mass nouns, pluralized count nouns, and special, bare singular count nouns.

In the case of the bare singular count nouns like *college, society, class, campus,* etc., these learners produced the strikingly high error rate of 41.5%. There were many sentences like these:

62. First of all, it is a good chance for student raise their ability of living in *the society*. [PTJ 03]

63. [...] students in *the school* take classes in the campus everyday […] [PTJ 13]

64. After graduating from *the college*, we are to step into *the society*. [PTJ19]

In addition, learners demonstrated a relatively high rate of *the*-overuse with pluralized count nouns (see Table 4). Since bare singular count nouns and pluralized count nouns are related by the shared presence of count nouns, I suspect that difficulty determining a noun's countability, or dealing with it in some way, may underlie these learners' *the*-overuse. Since most nouns in Chinese take the same surface form regardless of whether they are count or mass, it is possible that unfamiliarity with this distinction in English could cause learners to insert *the* unnecessarily. However, such speculation about the role of Chinese morphology in this pattern of findings is beyond the scope of this research.

What the above discussion serves to show, however, is that the differing systems of English in Chinese may overlap and interlock in the mind of an ESL speaker. In addition, it may not be the case that any single aspect of a given noun (like specific reference or hearer knowledge) can account for the learner's overuse of *the.* A combination of factors (language

interference, imperfect grasp of L2 grammar, morphological patterning, count/mass distinctions) should also be considered in error analysis.

Let us briefly consider the nouns that were ambiguously **in-between** non-referentials and generics. All of the in-between nouns were superficially singular, though there was evidence of an omitted required plural, as in example (61):

65. […] for some student who just left school to find a good job since *the company* do not admit they ability by a piece of paper [PTJ 03]

Cases such as these, while it is not clear whether they should be grouped with generic nouns or non-referential indefinite nouns, may provide further support for my idea that isolating morphemes like articles, particularly *the,* could be more salient to native Chinese speakers than suffixes likes English plurals. If, perhaps, a student is unsure of whether or not a plural marker would be appropriate, the definite article may present itself as an acceptable option for modification. Another possibility is that the students *did* intend to use a singular noun, and perhaps only erred in the conjugation of the verb. Therefore, while these in-between nouns should be explored further, experimenting with their effects on the more clear-cut categories is beyond the scope of the current project. I will leave these cases to be investigated in my own future research.

I turn now to the further aspects of generic and non-referential nouns that were reported from the corpus. As we have just seen, dealing with the countability of nouns presents a special challenge to Chinese ESL learners. One of the categories by which the generic and non-referential nouns were broken down was unambiguous versus ambiguous (see table 5), which refers to whether or not a noun is always mass or always count (unambiguous: *dog, book, dust*), or can be either, depending on context (ambiguous: *freedom, room, string*). As count nouns, in both singular and plural form, gave the learners trouble, the fact that unambiguous nouns showed

a higher rate of *the-overuse* could be accounted for by considering the proportion of these unambiguous nouns that were count. In fact, there were more than twice as many unambiguous count nouns as unambiguous mass nouns, whereas the ambiguous nouns were approximately evenly divided between count forms and mass forms. Therefore, because an unambiguous noun was more likely to be count, it is possible that it was, in turn, more likely to be used with an unnecessary *the* due the one or more of the effects of countability discussed above.

Finally, the question of sentence position remains (tables 7 and 8). This is the only category in which non-referential indefinites differed from generics to a statistically significant extent. While generics tended to have more *the*-overuse as objects of prepositions (12.1%), non-referential indefinites had more *the*-overuse as subjects (15.1%). Of these two patterns, only the latter is predicted by the effects of interference from Chinese. This is because the topic position in a Chinese sentence is pre-verbal (it precedes the verb), which is, syntactically at least, equivalent to the English subject position. Since both non-referential indefinites and definite nouns can appear in the pre-verbal topic position, it follows that subject position would be the site of most confusion. However, generics can also appear in topic position, and yet there is a higher rate for the object of preposition position in English.

One possible explanation for this distribution is that there is an effect of formulaicity. Formulas are "chunks" of language, composed of words that frequently co-occur, such as "down the stairs" or "take a walk." Sometimes, they are called **prefabrications** or **formulaic language.** In English, prepositions like "in" and "on" serve two purposes: to indicate the position of things relative to a viewer and to surrounding objects in the physical world, and to suggest relationships between more abstract concepts (as in *in society* or *on drugs).* When used for the former purpose, the viewer typically describes the location of an object relative to things immediately nearby,

which definitely exist and are usually unique—the characteristics of a definite noun phrase. Therefore, when learning how to use prepositions, ESL students might internalize phrases like "in *the* _____" or "on *the*_____" and extend this formulaic expression to abstract expressions, leading to unidiomatic utterances like *in the society.*

There has been a not insignificant amount of research on the subject of formulaic language in L2 pedagogical technique (Ellis et al., Girard & Sionis, Jiang & Nekrasova). Jiang and Nekrasova, who studied the holistic processing of formulas by native and non-native English speakers, speculated that instruction "during which formulaic sequences may be introduced to students as unanalyzed phrases" could play a role in how the phrases are represented in students' vocabulary: as rigidly bound lexical items (Jiang & Nekrasova 442). In fact, Ellis et al. (citing Lewis 1993, Nattinger and DeCarrico 1992) describe how the *lexical approach,* which was developed to mirror first language acquisition of phrases and formulas, "focuses instruction on relatively fixed expressions that occur frequently in spoken language" (Ellis et al. 378). All this serves to show that not only might L2 students be predisposed to learn and memorize formulas, but also that there is even a pedagogical theory advocating the use of formulas as "the pedagogically applicable unit" (Ellis et al. 378). Therefore, it is possible that Chinese ESL learners have been exposed to such formulas or such teaching techniques, and as a result may overextend "preposition + *the* + noun" patterns to necessarily abstract concepts, like those represented by generic nouns. The potential effects of formulaicity, and the implications of this research for pedagogical methods that employ it, will be explored further in section 6.

Thus far, I have attributed much of the definite article overuse seen in the corpus to a combination of interference from Chinese and the issue of argumentative contexts in English, as well as to effects of learners' classroom experiences. Much of the previous research, however,

centered instead on two competing theories of Bickerton's [±SR ±HK] classification system. Some studies (Master 1987, 1997) suggested that hearer knowledge induced *the*-overuse, while others (Ionin & Wexler) argued that referentiality, instead, played that role. The findings of the present study currently support the notion that hearer knowledge is the more relevant factor. Both types of nouns that had the highest error rates in the corpus, non-referentials and generics, were [-SR +HK]. By comparison, those nouns in the corpus that were [+SR –HK], (the "new" nouns), had a relatively low rate of *the*-overuse. This rate of error conflicts with Marta Tryzna's (2009) finding that Chinese speakers tended to overuse *the* in just such situations of referentiality; it also argues against Ionin and Wexler's overall argument that a noun's referentiality may induce *the*-overuse.

In addition, my findings align with Liu and Gleason's finding that *the*-overuse was most common with generic nouns, second only to the arbitrary cultural cases like *the flu* vs. *the cold.* This would indicate that argumentative discourse styles, which are more likely to have a greater prevalence of such nouns (especially as confirmed by this corpus), may present additional challenges to Chinese ESL learners, who must use and respond to generic nouns in such contexts. Furthermore, it is possible that Liu and Gleason's arbitrary cultural uses of *the* did make an appearance in this corpus: with bare singular count nouns. Since they are exceptions to the usual rule, we could consider bare singular count nouns like *college* and *society* as phenomena accessible only to native English speakers with cultural familiarity. Whether or not the occurrence of certain count nouns as bare singulars is truly arbitrary, however, must be investigated in future research.

What is certainly apparent from the findings of the present study is that learners do not seem to randomly overuse *the* in a wide variety of positions; nor are the rates of *the*-overuse

relatively equally distributed among all types of nouns and in all sentence positions. This would be the case if learners were simply "flooding" their language production with the definite article; however, the results of this study indicate that there is significant variation in how and where *the* is unnecessarily inserted, with discourse context playing one of the most important rules—particularly at this level of proficiency. This is because the types of nouns that learners are required to manipulate will necessarily influence the patterns of error they produce.

However, I believe that the association of *the* with hearer knowledge is further compounded by issues of morphological variation between L1 and L2, by the confluence of competing grammatical rules, by the preparation received in the classroom, and by the context in which language production takes place. If the notion of hearer knowledge were purely responsible for *the*-overuse, there would be little variation according to within-group factors such as concreteness, ambiguity, or sentence position—yet these all demonstrate effects.

6.0 Implications and Conclusions

6.1 Pedagogy

What we have discovered is that explanations of ESL learner competence should incorporate the effects of variables beyond simply the semantic; and not only this, but these explanations should also be tailored to the unique qualities of different learner groups. Chinese speakers will have challenges different from those met by French or Persian or Tagalog speakers. As convenient as it would be to standardize ESL curricula so that they are equally accessible to learners of all backgrounds, this is not the ideal situation.

For Chinese learners in particular, this study highlights the importance of introducing America- (or Europe) bound students to the concept of written argumentation in English. A path of study suited to such a goal would expose students to ample examples of this type of writing, in

order to balance out any previous experience with narrative rules and noun distribution patterns. It would make students aware of how generic nouns pattern with articles, both in their language and in English—since, in my experience, formal, explicit knowledge of one's native language can be extremely helpful in learning a second.

To accomplish this, I believe that a side-by-side analysis of the two discourse styles would be an excellent place to start. Simply introducing students to the concept of English argumentative writing without comparing it to the knowledge of narratives they already have would, in my opinion, fail to make use of the gift of scaffolding. The students already have a foundational understanding of articles in at least one context—in introducing another context, a teacher might ask them about what article rules they already know, and have them produce several sentences or correct errors in a narrative discourse. This bolsters their confidence while it explicitly re-familiarizes them with their knowledge.

To capture their active attention and engage them in the lesson, beginning with something surprising—that very clearly violates the rules they already have—is, I have found, an effective hook. It is the most obvious way to demonstrate that there *is* variation and it *can* dramatically affect the meaning of or the response to a learner's utterance. For example, one might illustrate some of the arbitrary cultural uses of *the* as an introduction to idiomatic definite article use:

66. "Did you know that English speakers sometime say they play *the piano,* even though they aren't talking about one piano in particular? They say it that way even if nobody has said anything about a piano yet in conversation."

This type of discussion serves to show students two things. First, it conveys that the definite article is not limited to use with second-mention nouns in a discourse. Second, it functions as a segue into the concept of generic nouns in English. The next step would be to communicate very clearly that the type of noun phrase above is actually relatively rare and

specialized, but that there *are* such things as generic nouns in English, and they ordinarily occur as pluralized count nouns or mass nouns. Furthermore, generic nouns can be mentioned more than once in a discourse and still not take the definite article. I believe the best way to demonstrate this would be to move from a sentence level (using nouns of each morphological type and in each sentence position) to a discourse level, directly indicating how generic nouns are repeated.

67. "Orca *whales* are not actually endangered."

68. "Some doctors think that not all patients should take *medication*."

69. "Some people seem to spend all their time in *shopping malls."*

Following this discussion, the teacher might ask the students to write their own sentences with generic nouns. Then, most likely, it would be wise to begin looking at discourses by focusing on one generic noun and how it is carried through, ignoring other examples of generics for the time being.

> 70. "*Orca whales* are large, carnivorous mammals that live in cold, arctic water. They primarily hunt seals and sea lions, usually in packs. Not only are *orca whales* extremely stealthy hunters, but they are also very intelligent. They can be trained to perform tricks and can live long lives in captivity. Some agencies, which believe *orca whales* should live only in the wild, protest against places like SeaWorld, which keep and train *orcas."*

Once a text such as this has been analyzed in terms of how the key generic word patterns, the teacher might ask students to read it over again and point out other generic words that appear: *mammals, water, seals, sea lions, packs, hunters, tricks, captivity, agencies, places.* Repetition of information is important, so the teacher may ask about what features these nouns have in common. Ideal answers: they are pluralized count nouns, or mass nouns. Emphasize the fact that none occurs with a definite article.

This discussion could be followed by additional comparisons of narrative versus argumentative discourses, similar to the two examples (5 and 6) given at the beginning of this paper. The instructor could, for instance, present students with a narrative discourse describing a woman who spends all of her time at a particular mall, including the things she buys and what she does with them. Then, the instructor could ask the students to write a short *argumentative* paragraph on the topic of malls, shopping, and purchasing, perhaps requiring that they use some of the nouns from the narrative in generic form. Such an exercise could reinforce the differences in noun and definite article distribution between the two types of discourse styles. Eventually, the teacher could introduce longer, more essay-like argumentative examples. For instance, samples from a corpus like CEECUS would be an excellent way to teach students to recognize and self-correct errors.

This sort of lesson could potentially be taught over the course of one class period, or broken up into stages of learning and acquisition over a longer period of time, with a ten- to fifteen-minute session each day. The earlier the exposure to this argumentative register begins, the less trouble learners are likely to have once they enter college, since it has been shown that overuse of *the* does tend to drop with increasing English proficiency (Master; Wong & Quek).

Apart from such structured lessons of definite article use, the present study suggests that the distinction between count and mass nouns in English remains a possible source of confusion for Chinese ESL learners. Since the ramifications of this may carry over into other issues of English grammar—such as the use of the definite article—the results of the project indicate that additional support in this topic may be required for advanced level students. In particular, students should be introduced to those nouns in English that are permitted to occur as bare singular count nouns. In fact, the data in this study have been coded for additional features of the

nouns that may aid in the design of appropriate curricula for clarifying this topic: the presence or absence of modifying phrases, singular/plural status, and the number and type of errors in assigning count/mass status. The wide array of noun phrases and sentences captured in the course of this study is easily searchable by these various factors, and could be used (as could many learner corpora) as a useful source of error examples to show students and have them actively correct.

These approaches to teaching often-idiomatic aspects of English are a partial departure from communicative language teaching (CLT), with its emphasis on "common communicative expressions" (Yu 195). I intended for them to emphasize recognition of L2 variation, especially in the written domain; although, because they closely examine rules and exceptions in grammatical patterns, they are also somewhat analytical in nature. My main hope is that a focus on grammar and idiomaticity, as well as allowing students to work at the discourse level, can represent a middle-ground between the traditional, highly analytical grammar-translation approach and the "participatory, experiential ways of learning" (Stern 1983, qtd. in Yu) embraced by CLT. In other words, the use and overuse of the definite article, like most other aspects of English, should be considered in the context of the task being performed and the intended meaning.

The reason, it would seem, that unidiomaticity persists beyond pure ungrammaticality in ESL language is the sheer number and complexity of the variables involved. Errors that blatantly violate English grammar are easily explained and corrected—the rules are fairly black and white--, while, as we have found, errors that render something only strange-sounding require a great deal of detailed explanation, much of which is inaccessible to English teachers not trained in formal linguistics (and even within linguistics, the relevant knowledge is quite specialized). As a

result, errors of unidiomaticity are explained away by telling students to just memorize chunks or that they will "eventually get it." As native speakers, however, I feel we have a responsibility to investigate and understand the subtleties of our language, if only so that those students who are especially motivated have the opportunity to pursue a higher level of fluency. In my future with TESOL, I hope to find--and foster—the exchange of theory and technique between linguists of English and teachers of English. Even though most English language classrooms in the United States are quite diverse, I believe that simply acknowledging the combined influences of L1 interference, morphological rules, and discourse style variation is an important step toward improved pedagogy.

6.1 Future Research

I realize that the current study is limited in its scope and applicability, and I want to address opportunities for further research. First of all, this project is meant to complement studies of narrative discourse styles—I do not claim that it presents a full picture of learner competence, but rather an additional angle from which to observe the puzzle. In addition, a larger corpus of essays, written in response to a wider variety of prompts, could further strengthen (or perhaps change) the results found here. Companion studies of corpora from different language backgrounds would also provide an interesting foil to these data, either corroborating my claim that L1 interference plays a significant role in *the*-overuse, or perhaps neutralizing them. A study of a comparable English corpus would also confirm the overall distribution of noun types (generic, new, etc.).

Finally, future studies may look with greater precision at the effects of the variables discussed here (concreteness, ambiguity, subject position, and the in-between nouns). In particular, the effect of concreteness if nouns found in this study could be further explored, as

speculating about the cause of this result is beyond the scope of this project. Such pursuits would only add to the knowledge base from which pertinent curricula could be developed.

References

Allerton, David J. "Three (or four) levels of word co-occurrence restriction. *Lingua* 63: 17-40.

Articles. Handout by The Writing Center, University of North Carolina at Chapel Hill. Web. http://writingcenter.unc.edu/handouts/articles/

Bickerton, D. *Roots of language*. Ann Arbor: Karoma, 1981.

Corpus of English Essays Written by Chinese University Students, see: Ishikawa, S. (2009). "Vocabulary in interlanguage: A study on Corpus of English Essays Written by Asian University Students (CEEAUS)." In K. Yagi & T. Kanzaki (Eds.) Phraseology, corpus linguistics and lexicography: Papers from Phraseology 2009 in Japan (pp. 87-100). Nishinomiya, Japan: Kwansei Gakuin University Press.

Ellis, Nick C., Rita Simpson-Blach and Carson Maynard. "Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and Tesol." *TESOL Quarterly* 42.3 (2008): 375-396. Web. 4 June 2012.

Fodor, Janet Dean and Ivan A. Sag. "Referential and Quantificational Indefinites." *Linguistics and Philosophy* 5.3 (1982): 355-398. Web. 26 December 2012.

Girard, Marie and Claude Sionis. "Formulaic Speech in the L2 Classroom: An Attempt at Identification and Classificiation." *Pragmatics* 13.2 (2003): 231-251. Web. 6 June 2012.

Ionin, Tania and Ken Wexler. "The Certain Uses of *the* in L2-English." Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002), ed. Juana M. Liceras et al., 150-160. Somerville, MA: Cascadilla Proceedings Project.

Jiang, Nan and Tatiana M Nekrasova. "The Processing of Formulaic Sequences by Second Language Speakers." *The Modern Language Journal* 91.2 (2007): 4330445. Web. 4 June 2012.

Lewis, M. *The lexical approach: The State of ELT and the way forward.* Hove, England: Language Teaching, 1993.

Liu, Dilin and Johanna L. Gleason. "Acquisition of the Article *The* by Nonnative Speakers of English." *SSLA* 24: 1-26. Web. 16 December 2012.

Lu, Crystal Fen-Chuan. "The acquisition of English articles by Chinese learners." *Second Language Studies* 20.1 (2001): 43-78.

Master, Peter. *A cross-linguistic interlanguage analysis of the acquisition of the English Article system.* Unpublished doctoral dissertation, UCLA (qtd. In Master 1997)

Master, Peter. "The English Article System: Acquisition, Function, and Pedagogy." *System* 25.2 (1997): 215-232. Web. 21 January 2013.

Nattinger, J.R. & J. DeCarrico. *Lexical phrases and language teaching.* Oxford: Oxford University Press, 1992.

Po-Ching, Yip and Don Rimmington. *Chinese: An Essential Grammar.* New York: Routledge, 1997. Print.

Prince, Ellen F. "The ZPG letter: Subjects, definiteness, and information-status." *Discourse description: diverse analyses of a fund raising text* (1992): 295-325. Web. 25 January 2013.

Robertson, Daniel. "Variability in the use of the English article system by Chinese learners of English." *Second Language Research* 16.2 (2000): 135-172. Web. 11 November 2012.

Shi, Dingxu. "Topic and Topic-Comment Constructions in Mandarin Chinese." *Language* 76.2 (2000): 383-408. Web. 23 March 2013.

Stern, H.H. *Fundamental concepts of language teaching.* Oxford: Oxford University Press, 1983.

Tarone, Elaine and Betsy Parrish. "Task-related variation in interlanguage: The case of articles." *Language Learning* 38: 21-44. Print.

Tryzna, Marta. "Questioning the Validity of the Article Choice Parameter and Fluctuation Hypothesis: Evidence from L2 English article use by L1 Polish and L1 Mandarin Chinese Speakers." *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications.* Eds. María del Pilar García Mayo and Roger Hawkins. Philadelphia: John Benjamins Publishing Company, 2009. 67-86. Print.

Wong, Bee Eng and Soh Theng Quek. "Acquisition of the English Definite Article by Chinese and Malay ESL Learners." *Electronic Journal of Foreign Language Teaching* 4.2 (2007): 210-234. Web. 16 December 2012.

Xu, Xiaoyan. Rev. of *FLT In China: Problems and Suggested Solutions. Applied Linguistics* 27.4 (2006): 754-766. Web. 8 June 2012.

Young, Richard. "Form-Function Relations in Articles in English Interlanguage." *Second Language Acquisition and Linguistic Variation.* Eds. Robert Bayley and Dennis R. Preston. Philadelphia: John Benjamins Publishing Company, 1996. 135-175. Print.

Yu, Liming. "Communicative Language Teaching in China: Progress and Resistance." *TESOL Quarterly* 35.1 (2001): 194-198. Web. 23 March 2013.

Zhang, Hong. "Numeral Classifiers in Mandarin Chinese." *Journal of East Asian Linguistics* 16 (2007): 43-59. Web. 21 July 2012.