台灣學生英文寫作冠詞錯誤分析

English article errors in Taiwanese college students' EFL writing

Neil Edward Barrett
國立成功大學外國語文學系（所）
Department of Foreign Languages & Literature
National Cheng Kung University
baoneer@hotmail.com

陳麗美 Li-mei Chen
國立成功大學外國語文學系（所）
Department of Foreign Languages & Literature
National Cheng Kung University
leemay@mail.ncku.edu.tw

Abstract

The English articles, *the, indefinite a/an*, and *zero* can be troublesome for English language learners. Thomas [1] demonstrated that English second language (L2) learners from first languages (L1) that do not have the equivalent of an article system encounter problems using articles. Ionin and Wexler [2] found that such learners fluctuate between definiteness and specificity. This study examined English L2 article use with Taiwanese English learners to determine the potential factors influencing English article substitution and error patterns in their academic writing. The corpus-based analysis used natural data collected for the Academic Writing Textual Analysis (AWTA) corpus [3]. A detailed online corpus tagging system was developed to examine article use, covering semantic (specific and hearer knowledge) as well as the other features of the English article. The results indicated that learners overused both the definite and indefinite articles but underused the zero article. The definite article was substituted for the indefinite article in specific environments. Although no significant difference existed between specific and non-specific semantic environments in zero article errors, a significant difference emerged between plural and mass/non-count nouns. These results suggest that, in regard to writing, learners need to focus on the semantic/pragmatic relationships of specificity and hearer (or reader) knowledge.

Keywords: definite article, indefinite article, zero article, hearer knowledge.

1. Introduction

The use of cohesive devices in writing is a well-researched topic in second language acquisition research, taking on a greater significance in recent years as increasingly more students are being asked to present their work in English, thus pointing to the need for greater accuracy and cohesion in students' writing. Errors within the article system (i.e., *a*, *an*, *the*,

Proceedings of the 22nd Conference on Computational Linguistics and Speech Processing (ROCLING 2010), Pages 266-280, Puli, Nantou,Taiwan, September 2010.

and the zero article) have been noted in studies examining L2 learners' writing, and such errors can be found in advanced learners' texts as well [4]. To put this issue into perspective, a corpus study of 668 TOEFL essays from Chinese, Japanese, and Russian students found that 13% of sentences—or 1 in every 8 noun phrases—had article errors [5].

Research into article errors has revealed that English article errors by English learners may be due to an inability to require the semantic feature of specificity [6] [2] resulting in the overuse of the definite article in specific environments or it may be a pragmatic deficit [7], when learners overuse the definite article due to extra-linguistic features such as world knowledge. Other studies [8] [9] [10] have investigated noun countability and its influence on article errors.

Although previous research has examined Mandarin English L2 article use in spoken discourse [11] [12] or article use in a cloze test [4] [13], only [7] have investigated texts from Mandarin L1 English learners. They concluded that Mandarin L1 English learners had both a grammatical and pragmatic deficit. However, in that study, only the semantic features of the articles were examined and essays were completed with time restrictions.

The aim of this study is to identify the features that influence students' article use or misuse. We first noticed the frequency of article errors in undergraduate writing while tagging cohesion errors for the Academic Writing Textual Analysis (AWTA) [3] corpus, an online corpus of Taiwanese undergraduate writing. Although the article errors did not seriously impair communication, they nevertheless interrupted the flow of the writing. Consequently, it was felt that the reasons for these errors deserved further attention.

To investigate the factors that influence article errors, this study will ask the following questions:

1. What is the influence of specificity and definiteness on the English article substitution and error patterns in the academic writing of Taiwanese college students?
2. What other potential factors influence English article substitution and error patterns?

2. Literature review

English has three articles, the definite, indefinite, and zero, which have a wide range of semantic and syntactic functions in discourse [11]. A widely used theory related to English article use is [14]. According to Bickerton, English noun phrases (NPs) can be classified according to two features: specific reference [+/-SR] and hearer knowledge [+/-HK]. Table 1 illustrates the four NPs. Many studies have shown that the failure to recognize [HK] has led to article errors in article production tasks [12] [4] and cloze tests [8] [15]. A more recent development in article system research was proposed by [2]. Based on their studies of Russian and Korean—two languages that do not have an article system—these authors proposed that articles are governed by semantic parameters. Their theory is known as the article choice parameter (p. 10).

Evidence for this comes from languages such as Samoan, which has different articles to indicate if a NP is specific or non-specific. English does not have the [+/-specific] setting, but instead has the definiteness setting [+/- definiteness]. Samoan uses the article *le* with specific noun phrases and *se* with non-specific, but does not mark definiteness [2]. The Samoan data analyzed by Ionin and Wexler demonstrate that definiteness may be irrelevant in languages like Samoan. Thus, the authors proposed the article choice parameter, which states that, "A language that has two articles distinguishes them as follows: The Definiteness Setting: Articles are distinguished on the basis of definiteness; The Specificity Setting: Articles are distinguished on the basis of specificity" [2, p. 12].

Table 1. Bickerton's noun phrase environments [8] p. 478.

Noun phrase environment	Example
[−SR, +HK], (*the, a, zero*): Generics.	*A* cat likes mice. *The* whale is a mammal. *(zero)* Language is a great invention of human kind.
[+SR, +HK], (*the*): Unique, previously mentioned, or physically present referents.	When I found a red box in front of my house, it was too late. *The* box blew up with a terrific explosion. This book did not sell well even though *the* author was a famous writer.
[+SR, −HK], (*a, zero*): First-mention NPs or NPs following existential "has/have" or "there is/are."	There is *a* new version of the I-phone. Did you see it? I keep sending *(zero)*messages to him.
[−SR, −HK], (*a, zero*): Equative NPs or NPs in negation, question, or irrealis mode.	He used to be *a* lawyer. *(zero)* Foreigners would come up with a better solution to this problem.

For [–] article languages, [3] proposed the fluctuation hypotheses, which states that learners fluctuate between the two parameter settings until they have enough input and the settings stabilize. Moreover, L2 learners may adopt parameter settings not found in their L1 or their L2 because if an L2 learner lacks articles in his/her L1, no language transfer should occur as there should be no parameter preference [2]. Thus, if languages such as Mandarin Chinese are seen as having neither articles for definiteness nor specificity, learners should fluctuate between the two settings for definite and specific. Based on this, [2] made specific predictions for [–] article L2 learners (see Table 2).

Table 2. Definite and Indefinite Fluctuation Hypothesis Predictions[13, p. 32]

Semantic type	+ definite	-definite
+ specific	Correct use of *the*	Overuse of *the*
-specific	overuse of *a*	Correct use of *a*

Although studies indicate that the fluctuation hypothesis correctly predicts L2 output [13], it has been criticized for several reasons. First, the fluctuation hypothesis does not take the zero article into account. For many first mention mass and plural nouns, specificity—as in first mention singular nouns—can be a semantic feature of zero article NPs, so the fluctuation hypothesis should also be able to predict error types.

2.1 Definite article in English and definiteness in Mandarin Chinese

Hawkins [16] initially based his location theory on previous article studies and subsequently revised his theory [17]. Hawkins identified eight different types of definite articles. By using *the,* a writer or speaker asks the reader/listener to locate the referent using knowledge that is available in the text (anaphoric and associative anaphoric use), can be sensed in the vicinity (visible and immediate situation use), or is available from local or general knowledge (immediate and local situation use). The other types of use—what Hawkins [16] called 'structural information' which refers to prepositional phrases, relative clauses or adjectives—help locate the referent.

In 1991, Hawkins revised his location theory based on both Gricean pragmatics and work by Levinson. According to Hawkins, the referents are located in pragmatic sets (p-sets) that are available to the speaker/hearer via discourse sets that contain information about a certain situation or event. These p-sets are associated knowledge shared by the discourse participants and can be accessed from present or prior discourse, the local environment, shared knowledge, or general knowledge. The main point of the p-sets is that they allow the hearer or reader to accept information as definite.

In Mandarin Chinese, only NPs that are referential (used to refer to an entity) can be definite or indefinite. A definite NP refers to a noun that a speaker/writer thinks is known to the hearer/reader whereas an indefinite NP is used for nouns the speaker/writer believes the hearer/reader does not know. If a NP has or can be used with a classifier, it is seen as referential; if this classifier is a demonstrative, the NP is definite [18].

Definiteness is also marked in Mandarin by word order, as Mandarin is a topic-prominent language. This means the topic appears sentence initial and shows either known information or generic uses, such as referring to an entire class of objects. The second part of the sentence is the comment, which contains new information [11].

2.3 English articles in second language acquisition studies

Numerous studies in second language acquisition (SLA) research have examined English articles, starting with Brown [19]. Research indicated that both young L1 children and L2 learners tend to associate the definite article with specific contexts rather than hearer/discourse knowledge. This became known as *the* flooding [19, p. 369], whereby a beginning learner overuses the definite article in all article contexts. Chaudron and Parker [21] found evidence that English learners misused articles in specific, discourse-first locations.

Moore [11] found that intermediate/advanced Mandarin Chinese subjects tended to overuse the indefinite article in both a cloze test and an oral narration task. Most of the indefinite errors occurred in *a for the* errors during the cloze test, but *zero/the* accuracy was almost the same during the narration task. Lee [4] found that for definite article error types, *the for zero* errors were more common in front of unique common nouns and in specific contexts.

More recently, [7] carried out a corpus study into article acquisition in Spanish and Chinese English L2 learners. The definite article was overused in specific contexts, but the zero article was also overused demonstrating issues related to noun countability. They suggested that the overuse of the definite article was a pragmatic problem as the writers did not consider the readers knowledge, while noun countability was seen as a grammatical problem.

2.4 Noun countability and English articles in SLA studies

Noun countability has been an issue in article acquisition, especially for learners from languages that do not use an article system [22] [8] [9]. Using a cloze test, Yoon [9] found that Japanese learners had problems with *indefinite for zero* errors, especially with mass/non-count nouns. Goto-Butler [8] found that noun countability was also a source of errors with Japanese participants and mass/non-count nouns could influence definite article errors. Furthermore, Snape [10] found that Japanese L1 English L2 learners made more errors with the definite article within plural and mass contexts compared to singular contexts. For Mandarin L1s, Hua and Li [22] found that participants were able to distinguish between countable and uncountable nouns in English L2 but were more accurate with abstract nouns.

3. Methodology

A total of 30 students participated in this study. The students were juniors who had attended writing class with the same instructor for four semesters. These participants were chosen for several reasons. Participants who had taken writing class with the same instructor were needed in order to avoid the effect of different writing instruction. In addition, all participants had received the same length of writing instruction. Although an earlier pilot study had shown no longitudinal effect, some of the essays may have been too short to provide an adequate amount of tokens; thus, it is possible that longitudinal changes could affect article

accuracy. In order to control for this, the participants had to be students who had a similar amount of exposure to writing instruction.

The corpus consisted of 30 argumentation essays, with a total of 28,020 words. Only 30 essays were coded due to time limitations. Unlike automatic parsing, the article types and error types had to be coded manually. The pilot study revealed that shorter essays did not contain enough articles and article errors. Therefore, argumentation essays were coded, because their lengths ranged from 789 to 1,449 words, resulting in a mean of 980 words per essay. The original drafts of the essays were coded because they had not been corrected by the student, instructor, or peers.

To explore article use and error patterns in Taiwanese students' EFL academic writing, a coding scheme was developed to annotate the data with linguistic information. The coding scheme is based on a modified version used by Moore [11].After investigating other schemes, Moore's [11] was found to be the most comprehensive system as it was based on article research conducted by Hawkins [16], and Robertson[12]. This coding scheme has many advantages over the other schemes used in SLA article research because it combines the semantic environments, the definite article types identified by Hawkins [16], and the language transfer features described by Robertson [12]. The coding scheme was needed for a larger investigation into English article use. Although this scheme follows the procedure described by Moore [11], it was sometimes necessary to make some adaptations or collapse some of Moore's categories.

In terms of coding, Figure 1 shows a brief diagram of the actual data as they would appear in the window of the corpus. The tagging system and AWTA corpus are described in detail in [23]. The first pair of brackets indicates the meta-linguistic tag used in the corpus and the annotation shows either the article type or the error type after the equals sign. The original text is in the arrowed brackets, followed by the meta-linguistic information to make the tags clear in the reviewing process. The tagging system works as follows: Inside the bracket is the name of the article (e.g., the semantic or article type); information regarding whether it is used correctly is indicated by the letters Y or N, which represent correct and incorrect use. This is followed by a number indicating the general error type. For example, in<tag D PN N annotation="2 ">, the D is a definite article, PN stands for plural noun, which is the error type, N indicates an article error, and 2 is the code number for *definite for zero specific* errors. In this way the article error can first be identified and meta-linguistic information can be added. Figure 1 is an extract taken from the AWTA corpus. The tag ＜tag D IA N annotation="5"＞ indicates a definite for indefinite article substitution. The D is a definite article, IA stands for indefinite article, which is the error type, and 5 is the code number for *the for specific indefinite a/an* errors.

Many studies have showed that it would be better for the hearing disabled to have ＜tag D IA N annotation="5"＞the＜/tag＞ cochlear implant at an early age. Also, if implanted the cochlear implant at the age one to two, their language learning could come out of great improvement. However, the situation now seems that the elder people who are more than 55 years old, are not suitable to have the cochlear implant. They are usually told only to use ＜tag D PN N annotation="2 " ＞the＜/tag＞ hearing aids for that most people think it would be too late for them to have the implantation.

Figure 1. Annotating meta-linguistic information.

To deal with the repetition of a NP, which is often necessary when writing as it has a cohesive function [23]—although overuse or repetition can be interpreted as an immature writing style,—a types/token distinction was used. Here, token counts refer to the frequency of a particular word or phrase whereas type refers to the occurrence of a distinct word or phrase in a text. In terms of errors, token counts would record the same error throughout the text, whereas type frequency would only record a mistake once. Therefore, if tokens were classified as errors, it would present an inflated picture. This paper only coded the types to avoid inflating the number of errors.

Once the coding procedures were decided, the data were coded for errors as article errors are often discourse dependent, making it necessary to read the essays first without the distraction of tagging every English article. All the errors were highlighted and subsequently coded according to their error type. Next, the essay was coded for article use, starting with the definite article, followed by the indefinite, and then the zero article. This was done to collect information for related research into L2 English article use. The annotation system consisted of two main parts: the semantic and pragmatic relations of each article and a description of the common error patterns.

Article error types can tell a researcher a lot about what kind of articles the participants were using in their writing [24]. The most important contribution is that they can indicate if any patterns of underuse or overuse exist or if the errors are purely random. All together, 37 possible error types were identified. Article errors in the text that could not be tagged according to the error system were labeled “unclassified”; these included definite and indefinite articles that were erroneously used outside the NP, meaning they were general errors, not errors within the article system. Furthermore, it was presumed that these were writing mistakes as there was no pattern to the errors.

Cohen’s Kappa analysis was used to measure inter-rater reliability. In the inter-rater procedure, the second coder was trained to use the corpus over three essays. If agreement was not reached, the two coders discussed the coding problems, and extra training was provided as necessary. In this study, 20% of the data was randomly selected from the argumentation

essays and coded by the two raters. The Kappa statistic was calculated at 0.332, which indicates a fair level of agreement between the two raters.

4. Results

This section explains the rationale and formulas for reporting accuracy and presents the accuracy of the three articles. Following this is a description of the distribution patterns of each article, including the semantic and structural functions. After the essays had been tagged, the data was checked for inter-rater reliability, and then the raw frequency counts for each error type were computed.

In order to report the frequency of the article errors, the data had to be normalized to allow data from different texts to be accurately compared. As the lengths of the essays differed between participants, reporting the raw frequency counts would not present an accurate account of the errors. In a longer text, there are more opportunities for errors to occur, so 'normalization' is a formula that adjusts the raw frequency counts so that texts of different lengths can be compared [26]. In normalization, the raw frequency counts are divided by the number of words in the text and then multiplied by the mean essay lengths for the 30 essays, which is 980 words per essay. The following example illustrates the normalization formula:

definite for zero errors 26 /1020 x 980 = 24.98 *definite for zero* errors per 980 words.

In this formula, there are 26 *definite for zero* errors in one essay. This is divided by the total number of words in the essay, and then multiplied by the mean essay length, giving a total of 24.98 errors per 980 words. If raw frequency counts are used, not only is it difficult to demonstrate a direct comparison, it can also present an inflated or deflated count.

Table 3 reports the distribution of the article types and article errors throughout the corpus. It is presented as a matrix table and it is read from left to right. The article type *the* on the horizontal axis shows the definite article, and reading the column from left to right indicates where the definite article is substituted for another article. If the table is read from left to write, starting with the definite article, it indicates where the definite article is being substituted in place of another. For example, reading the matrix from left to right indicates that 9.16% *the for a* substitution errors occurred. The highest frequency is *definite for zero* errors at 34.47%.Countability errors occurred when the indefinite article was substituted for the zero article or vice versa. The results illustrate that 6.75% *zero for a/an* errors occurred, and 2.25% *a for zero* errors occurred. The number of *zero for the* and *a for the* errors are low at 5.33% and 0.79% respectively, indicating that the frequency of definite article underuse is low. Table 3 clearly indicates that overuse of the indefinite and zero article is low, more errors are made with the definite article, while countability errors are relatively lower. In other

words there are far more semantic or pragmatic errors than grammatical errors. Grammatical errors are due to noun countability errors where the writer must assign the indefinite article to singular nouns and the zero article to plural or mass/non-count nouns.

Table 3. Article error distribution

Article	The		Indefinite a/an		Zero	
	freq	%	freq	%	freq	%
The	922.71	**93.87**	**48.41**	**9.16**	**159.24**	**34.47**
Indefinite a/an	**7.81**	**0.79**	443.95	**84.07**	**10.4**	**2.27**
Zero	**52.43**	**5.33**	**35.65**	**6.77**	292.21	**63.26**
Total	982.95	100	528.01	100	461.85	100

The next section analyzes the influence of semantic NP environments, and countability in order to determine their effects on article errors as the effects of specificity and countability have been well documented as factors that influence L2 learners' article errors.

Table 4 illustrates the frequency of the main article errors according to NP environment and countability. The highest frequency of errors can be found in *definite for zero specific plural* errors followed by *the for specific indefinite a/an* errors. These descriptive results suggest that specificity influences the frequency of the *for indefinite a/an* errors as more errors occur in specific NPs. The frequency of *zero for a* errors is low at 10.58% of total errors, but suggests that some participants have trouble using the correct article with singular and plural nouns. The number of *a for zero* and *zero for the* errors was not reported as their frequencies were very low indicating that this is not a problem for the participants. The frequency of *definite for zero* errors in both specific and non-specific environments indicates that specificity may not be the only influence on *definite for zero* errors.

Further statistical analysis was needed to investigate the influences on error types. It has been predicted that for English L2 learners with no article system in their L1, more errors are found in specific indefinite noun phrase environments [2].To determine the effect of specificity on definite for indefinite errors, a paired sample *t*-test was conducted. As there are only two independent variables, a *t*-test could show if the difference between specific and non-specific *the for indefinite a/an* errors is significant. It revealed a significant difference between the two groups (t (29) = 6.94, $p < .001$). The mean of the specific indefinite errors was significantly higher (m = 1.36, sd = 1.03) than the mean of the non-specific errors (m = 0.25, sd = 0.46), indicating that specificity influences definite article errors in indefinite specific environments. In other words, the definite article is being substituted for the indefinite article in specific environments as predicted by Ionin & Wexler's [2] fluctuation hypotheses. The implications of this are discussed in section 5.

Table 4. Error types across the corpus per 980 words

Error type	freq. of errors	% of total errors
Zero for A	26.38	10.58
Definite for zero		
Non-count specific	23.98	9.62
Plural specific	**45.12**	**18.10**
Plural non-specific	**35.12**	**14.09**
Definite for zero non-count non-specific	16.14	6.47
The for A		
Specific	**41.09**	16.48

Note. N = 30 (N = shows the size of the data pool which is 30 subjects.)

Table 5 presents the descriptive statistics for definite article for zero article errors where the four independent variables are *definite for zero specific plural* errors, *definite for zero non-specific plural* errors, *definite for zero non-count/mass specific* errors, and *definite for zero non-count/mass non-specific* errors. Some researchers [8] and [9] believe that in addition to semantic environments, the difference between count and mass/non-count nouns may have an influence on article errors. Due to this, more errors are expected with mass/non-count nouns than plural nouns. Also, due to the hypothesis [2] that specificity influences article errors, more errors are expected in specific NPs. It was suggested that a repeated measure ANOVA would be able to show any significant differences between NPs environments, and any differences between plural and mass noun errors.

Table 5. Descriptive statistics for definite article for zero article errors

Substitution type	M	SD
Definite for zero specific plural errors	1.50	1.67
Definite for zero specific non-count/mass errors	0.79	0.99
Definite for zero non-specific plural errors	1.19	1.51
Definite for zero non-specific non-count/mass errors	0.53	0.73

Note: N = 30

Table 6 shows the repeated measure ANOVA results for the definite article for zero article errors. A significant effect was found (F (3, 87) = 5.66, $p < .005$). Follow-up protected t-tests revealed a significant difference between *definite for zero plural* (m = 2.70, sd = 2.67) and *definite for zero non-count/mass* substitution errors (m = 1.33, sd = 1.54), showing an effect with noun countability on *definite for zero* errors. In other words, more *definite for zero* substitution errors are found with plural nouns indicating that for these participants, mass/non-count nouns do not have a significant influence on definite article errors. The

follow-up protected *t*-tests between *specific definite for zero* (m = 2.37, sd = 2.31) and *non-specific definite for zero* errors (m = 1.73, sd = 2.09) revealed no significant difference between specific and non-specific zero indicating that specificity is not a significant influence in *definite for zero* article errors. The implications of this are discussed in section 5.

Table 6. ANOVA results for definite article for zero article errors

	df	F	η^2	*p*
Between subjects				
Definite for zero subs	3	5.66	.003	.001**
Within-group error	87	(0.96)		

Note: N = 30; **p<. 001

5. Discussion

The results indicated that Taiwanese college students majoring in English in this study had problems using the English article in terms of distinguishing between a definite and indefinite noun phrase. Correct article use in terms of noun countability was not a major problem for these writers. This section discusses the influence of specificity on article error patterns. First, the indefinite article is discussed, followed by the zero article.

Errors with specificity may stem from some participants' identification of a specific noun clause as definite, as predicted by the fluctuation hypothesis [2]. The fluctuation hypothesis predicts that writers whose L1 has no articles will overuse the definite article in specific environments until the correct article parameters are reset by exposure to the L2. The results of this study support the view that the definite article is overused in specific noun phrases with indefinite a/an as the results of the *t*-test show a significant difference between *the for indefinite a/an* errors, with more errors occurring in specific NPs. However, the fluctuation hypothesis also predicts overuse of the indefinite article with definite non-specific nouns (i.e., *a for the generic* errors). No such errors were evident in the results of this study, although only 54 generic indefinite noun types were counted in the data.

Zero articles not taking a generic, proper noun or idiomatic reading can be specific or non-specific in accordance with [24]'s specifications. The repeated measure ANOVA and follow up protected *t*-tests revealed no effect of specificity on *definite for zero* errors as no significant difference was found between specific and non-specific errors. Thus, unlike the indefinite article, specificity was not the only influence on the overuse of the definite article with zero articles. A misrepresentation of the pragmatic functions of the definite article is a possible reason for these errors and this will be discussed below.

The results demonstrate that the learners in this corpus lacked accuracy with regard to the zero article regardless of semantic type. As a result, the participants often compensated for this by using the definite article. The indefinite article cannot be used for plural nouns or

mass or non-count nouns due to countability rules. Thus, a writer has two article options: the zero or the definite. Although the fluctuation hypothesis may explain the errors in specific environments, it cannot explain definite article overuse in non-specific environments; thus, the effects of other influences need to be considered—particularly mass/non-count nouns or the hearer knowledge [HK] feature of definite articles.

The results clearly reveal that participants made significantly more errors with plural nouns than with mass/non-count nouns. Errors involving the definite article with mass/non-count nouns have been found in other studies [8] [13], although the results in this study reveal plural errors have a greater effect on error patterns. A *t*-test indicated a significant difference between mass/non-count nouns and plural nouns, indicating that definite article errors with mass/non-count nouns are less frequent than errors with plural nouns. This differs from what [8] found with their Japanese participants, who made more errors with mass/non-count nouns. In other words, for the participants in this study, the influence of mass/non-count nouns is not a factor in English definite article errors.

As previously pointed out, Mandarin Chinese has different ways in which to indicate definiteness (e.g., demonstratives, discourse context, or word order). In English, the context—namely, the speaker's and hearer's knowledge of the context—determines whether a NP can be located by both participants. If the writer believes that the hearer is aware of the noun, the definite article is used. In other words, as [7] pointed out, a writer often takes the readers' knowledge into account when using the definite article.

According to Hawkins [17], using the definite article enables the hearer to access the NP in the p-set (a set of knowledge known by the hearer/reader as being definite). The speaker/writer should use the definite article when he/she is confident that the other party knows that the NP is definite. A communication breakdown will occur if the speaker/writer uses the definite article erroneously or mistakenly believes that the hearer has such knowledge. The writers in this corpus may not have been falsely assuming that the reader had definite knowledge, but the writers may not be fully aware of how the definite article signals this knowledge and the resulting communication problems. Thus, errors with the definite article could be classed as errors regarding the pragmatic functions of the English definite article.

Although noun countability has been seen as a problem for English L2 learners, especially learners whose L1 does not have an article system [8] [9], for the participants in this study, the number of errors in *zero for a* and *a for zero* contexts was relatively low (9.51% and 2.87% of the total errors, respectively). The majority of these types of errors (18) occurred with count nouns, indicating that the writer incorrectly analyzed the noun as a non-count/mass noun. The following examples illustrate this error type.

> Example 1. Every citizen is suitable by the law. No one is exception if he or she committed *crime.*

Example 2. For some losers may bankrupt and then rob *bank* in order to win back.

These examples indicate that zero articles were substituted for the indefinite article. One reason for this was that the noun was incorrectly determined to be a non-count noun due to L1 influence. In Mandarin Chinese, all nouns are realized as mass/non-count nouns; as such, the countability feature of English nouns has to be acquired by these language learners.

When it comes to L2 writing, article use is often overlooked as it is often considered too difficult to teach all the features of the English article system [7]. The results of this study demonstrate that only certain aspects of the English article system need to be emphasized in college writing: the specific/definite distinction and the use of the definite article with zero nouns. For this it is suggested that reader awareness needs to be taught along side some of the uses of the definite article as this pragmatic feature is a major influence on English article errors.

This study helped with our understanding of the influence of specific knowledge, hearer knowledge and noun countability on English article errors in writing. Future studies need to investigate this across a larger corpus and different proficiency levels. It is also hoped that an English article teaching method can be devised based on the results of this study.

References

[1] M. Thomas, The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics, 10*, pp. 335-355, 1989.

[2] T. Ionin, and K. Wexler, Article Semantics in L2 Acquisition: The role of specificity. *Language Acquisition*, *12*, pp. 3-70, 2004.

[3] National Cheng Kung University. (n.d.). *Academic Writing Textual Analysis (AWTA) corpus*, [online], Available http://awta.csie.ncku.edu.tw/.

[4] E. H. Lee, *English article usage in online graduate forums by non-native EFL teachers,* Doctoral dissertation, Indiana University, Available from ProQuest Dissertation and Theses database, (UMI No. 3274275), 2007.

[5] N.R Han, M. Chodorow, and C. Leacock, Detecting errors in English article use by non-native speakers. *Natural Language Engineering*, *12*(2), pp. 115-129, 2006.

[6] N. Snape, L2 acquisition of definiteness and specificity in English by advanced Japanese and Spanish learners, In A. Belletti, A. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari (Eds.), *Language Acquisition and Development, Proceedings of the Generative Approaches to Language Acquisition Conference 2005,* pp. 591-596, Cambridge, UK: Cambridge Scholars Press/CSP, 2006.

[7] M. B. Diez-Bedmar, and S. Papp The use of the English article system by Chinese and Spanish learners, In G. Gilquin, M. B. Diez-Bedmar, & S. Papp (Eds.), *Linking up*

contrastive and learner corpus research pp.147-175, New York: Cambridge University Press, 2008.

[8] Y. Goto-Butler, Seond language learners' theories on the use of English article: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system, *Studies in Second Language Acquisition*, *24*(3), pp. 451-480, 2002.

[9] K. K. Yoon, Challenging prototype descriptions: Perception of noun countability and indefinite vs. zero article use. *International Review of Applied Linguistics, 31*(4), pp. 269-289, 1993.

[10] N, Snape, Resetting the Nominal Mapping Parameter in L2 English: Definite article use and the count-mass distinction, *Bilingualism: Language and Cognition*, *11*, 63-79, 2008.

[11] J. M. Moore, *Articles and proper names in L2 English,* Doctoral dissertation, Northwestern University, (UMI No. 3132569), 2004.

[12] D. Robertson, Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, *16*(2), pp.135-172, 2000.

[13] N. Snape, N, Exploring Mandarin Chinese speakers' article use. In N. Snape, Y. K. I. Leung, & M. Sharwood Smith (Eds.), *Representational Deficits in SLA: Studies in Honor of Roger Hawkins*, pp. 27-51. Amsterdam: John Benjamins, 2009.

[14] D. Bickerton, *Roots of language*, Ann Arbor, MI: Karoma Press, 1981.

[15] D. Trenkic, The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition, 11*(1), pp. 1-18, 2008.

[16] J.A. Hawkins, *Definiteness and indefiniteness*, London: Croom Helm Ltd, 1978.

[17] J. A. Hawkins, On (in) definite articles: Implicatures and (un)grammaticality predictions. *Journal of Linguistics*, *27*(2), pp. 405-442.

[18] C. N. Li, and S.A. Thompson, *Mandarin Chinese. A functional reference grammar*, Los Angeles: University of California Press, 1981.

[19] R. Brown, *A first language: The early stages*, Cambridge, MA: Harvard University Press, 1973.

[20] C. Chaudron, and K. Parker, Discourse markedness and structural markedness: The acquisition of English noun phrases, *Studies in Second Language Acquisition*, *12*(1), pp. 43-64, 1990.

[21] D. Hua, and H. Lee, Chinese ESL Learners' Understanding of the English Count-Mass Distinction. In D. Laurant, R. A. Sprouse, & A. Liljestrand (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)* pp. 138-149, 2005.

[22] T. Kao, and L. M. Chen, Coherence in Chinese students' English writing: An initiative

to a learner's corpus, In Y. Leung, & H. Chang (Eds.), *Selected Papers from the Seventeenth International Symposium on English Teaching* pp. 167-175, Taipei: Crane Publishing, 2008.

[23] J. E. Trademan, *The acquisition of the English article system by native speakers of Spanish and Japanese: A cross-linguistic comparison,* Doctoral dissertation, University of New Mexico, Available from ProQuest Dissertation and Theses database. (UMI No. 3058290), 2002.

[24] F. C. Lu, The Acquisition of English Articles by Chinese Learners, *Working Papers in Second Language Studies*, *20*, pp. 1-36, 2001.

[25] D. Biber, S. Conrad, and R. Reppen, *Corpus linguistics: Investigating language structure and use*, Cambridge: Cambridge University Press, 1998.

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