A CORPUS-BASED MULTIFACTORIAL ANALYSIS OF JAPANESE AND CHINESE SPEAKERS' ENGLISH ARTICLE USE: QUANTIFYING THE DEVIATION USING MUPDAR

By

Tatsuya Aoyama

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Teaching English to Speakers of Other Languages—Master of Arts

2020

ABSTRACT

A CORPUS-BASED MULTIFACTORIAL ANALYSIS OF JAPANESE AND CHINESE SPEAKERS' ENGLISH ARTICLE USE:
QUANTIFYING THE DEVIATION USING MUPDAR

By

Tatsuya Aoyama

The English article system poses a unique challenge to learners of English, especially for those with article-less first language backgrounds. This multifactorial corpus-based study investigates Chinese and Japanese speakers' use of definite, indefinite, and zero articles, based on 2,461 noun phrases annotated for relevant syntactic, morphological, and semantic factors. A multinomial extension of Multifactorial Prediction and Deviation Analysis with Regressions (MuPDAR; Gries & Deshors 2014) provides insights into how such factors affect the nativelikeness of the non-native speakers' article use, and how such effects differ for the three article types and for the first language backgrounds. The results show that noun countability and pluralization, among other independent variables, had significant effects on the accuracy of Chinese and Japanese speakers' use of English articles, and such effects are significantly different for the three types of English articles: definite, indefinite, and zero articles. Limitations of this study will be discussed at the end.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude for my thesis advisor, Dr. Sandra Claire Deshors, who has always been patient and supportive throughout. She was also the first person who introduced me to the field of corpus linguistics. I was immediately enthralled by the level of granularity a corpus-based study can afford, and this entire work would not have been possible if it were not for her.

I would also like to extend my gratitude to my second reader, Dr. Kristen Johnson. As a professor in Computer Science and Engineering (CSE) department, she has warmly welcomed me in her Natural Language Processing (NLP) class even though I was neither a CSE student nor proficient enough in programming languages. Most of the automatic data preprocessing I used in this study would have had to be done manually without her.

I am also indebted to amazing professors outside my university. Dr. Stefan Gries, a professor at University of California Santa Barbara, has provided me with keen insights into the methodological and statistical problems I had. Dr. Akira Murakami, a research fellow at Birmingham University in the U.K., patiently helped me decipher all kinds of problems I had. His kind responses to my sudden, random questions undoubtedly kept me moving forward.

Needless to say, as much as I am grateful for all the academic and personal support from these amazing professors, I would like to acknowledge that all the mistakes and incompleteness of this study are my own.

Last but not least, I would like to thank my parents, Tatsushi and Setsuko Aoyama, who have warmly welcomed me back at home, in the midst of the unprecedented chaos due to the

COVID-19 pandemic. Without their selfless support and love (and BIG breakfast, lunch and dinner), I would have given up halfway. *Arigato*!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

CFD	Communicative Function of Definiteness
CLC	Cambridge Learner Corpus
DA	Definite Article
EF	Education First
EFCAMDAT	The EF-Cambridge Open Language Database
EFL	English as a Foreign Language
ESL	English as a Second Language
IA	Indefinite Article
ICLE	The International Corpus of Learner Corpus
ICNALE	The International Corpus Network of Asian Learners of English
IL	Interlanguage
LOCNESS	The Louvain Corpus of Native English Essays
MuPDAR	Multifactorial Prediction and Deviation Analysis with Regressions
NL	Native Language
NNS	Non-Native Speaker
NP	Noun Phrase
NS	Native Speaker
SLA	Second Language Acquisition
ZA	Zero Article

1 INTRODUCTION

Studies in corpus linguistics have documented a number of linguistic phenomena, such as regional variabilities (e.g., Collins, 2007), gender differences (e.g., Fuchs, 2017), and lexical bundles (e.g., Hyland, 2012). The development of learner corpora and the advent of learner corpus research have extended this line of research to learner language, enabling corpus-based second language acquisition (SLA) research. Topics of such studies include L1 effect on dative alternation (e.g., Song & Sung, 2017), genitive alternation (e.g., Gries & Wulff, 2013), complementizer (e.g., Durham, 2011), relativizer (e.g., Lester, 2019), and articles (e.g., Diez-Bedmar & Papp, 2008), to name a few. With specific regards to articles, even though several studies probed into the acquisition of the English article system, few studies have approached this topic from a multifactorial perspective by simultaneously investigating various contextual factors that affect the use of English articles. Furthermore, the focus of L1 effect seems to be placed on the dichotomy between the first languages that have article system and the ones that do not, and few studies compared L1 effects within article-less languages (e.g., Crosthwaite, 2016).

In this context, the present study aims to investigate (i) how syntactic, morphological, and semantic factors affect the nativelikeness of the use of English articles; (ii) how those factors differentially affect the use of the three types of English articles, namely, definite (DA), indefinite (IA), and zero articles (ZA); and (iii) how the effects in (ii) differ within article-less languages.

2 LITERATURE REVIEW

In this literature review, several aspects crucial to understanding the English article system in L2 are discussed. First, I will focus on the uniqueness of the system with a particular emphasis on why it poses a particular challenge to L2 learners (Section 2.1). Secondly, an important factor contributing to the difficulty of the English article system, namely, L1 transfer, will be discussed (Section 2.2). However, studies on L1 transfer in the acquisition of English articles remain descriptive, and a multifactorial approach that incorporates various contextual factors is necessary. In this context, factors affecting the use of English articles will be then reviewed (Section 2.3). Lastly, research questions will be formulated given what has been discussed in Sections 2.1 – 2.3 (Section 2.4).

2.1 Uniqueness of the English Article System

The English article system poses a unique challenge to learners because of cross-linguistic differences of the article system (e.g., Larsen-Freeman, Celce-Murcia, & Williams, 1999; Section 2.1.1) and the difficulty in teaching and learning the use of the intricate English article system (e.g., Dulay, Burt, & Krashen, 1982; Master, 1994; Section 2.1.2). Morpheme order studies have quantitatively informed us of the relative difficulty of the article acquisition, in comparison with other morphemes (e.g., Goldschneider & DeKeyser, 2001; Murakami & Alexopoulou, 2016; Section 1.1.3). In what follows, I discuss each of these aspects.

2.1.1 Cross-linguistic Heterogeneity

Among a variety of grammatical morphemes available in English, the article system is particularly unique. Specifically, the presence of article system itself is not common across other languages, and the presence and absence of an article system in a particular language is often

expressed as [± ARTICLE]. For example, most Asian and African languages do not have an article system (Larsen-Freeman, Celce-Murcia, & Williams, 1999), and are thus described as [- ARTICLE]. Such languages often have a different way of expressing what the English article system expresses. Japanese, for example, is a [- ARTICLE] language, and it employs grammatical particles (or sometimes referred to as postpositions) *wa*, a topic marker, and *ga*, a subject marker, to distinguish an already-introduced topic from a newly-introduced subject. Furthermore, Japanese demonstratives are often used as determiners to limit the interpretation of noun phrases, and it can sometimes function as the English article system (Butler, 2002). An article system is not only unique in the sense that many languages do not even have such systems, but also in the sense that even though some languages, such as Spanish or German, have an article system (i.e., [+ ARTICLE]), the articles or article-like morphemes in such languages often behave differently from the English article system. For example, the Spanish article system is very similar to the English article system in that it has both definite articles *el* and *la*, and indefinite articles *un* and *una*, but it differs from English in many ways. One of such differences is that Spanish requires the pluralization of articles, changing definite articles to *los* and *las*, indefinite articles to *unos* and *unas*. Another difference is that mass nouns have to follow definite articles in Spanish; a restriction that is absent in English (Snape, García-Mayo, & Gürel, 2013). The different article systems across different languages require that learners have to learn a new, separate article system as part of their SLA process (Snape, 2008). However, whether or not the English article system can be taught or learned has been hotly debated, as I discuss below.

2.1.2 Learnability and Teachability of English Articles

Another aspect of the article system that sets itself apart from other morphemes is its learnability and teachability. Even though it is clear that the English article system is extremely complex and difficult to acquire (e.g., Larsen-Freeman et al., 1999), its learnability and teachability is under debate. For example, Dulay, Burt, and Krashen (1982), on the one hand, argued that the English article system is unlearnable and thus unteachable, and that abundant exposure is the only solution to its acquisition. Master (1994), on the other hand, conducted a quasi-experimental study to determine the teachability of the article system and concluded that it is indeed teachable. In his study, he performed a nine-week, focused and well-structured article instruction to the treatment group but not to the control group, and compared the gain score of those two groups. With this pretest-posttest design, he showed that the treatment group exhibited a statistically significant improvement in the accuracy score between pre- and posttest, whereas no such difference was found for the control group. However, given that the treatment group's gain score was minimal (from 26.79 to 29.08 out of 36), and that the control group also showed a small improvement (from 26.61 to 27.24), the effect of instruction is, although statistically significant, rather small. Also, only the immediate posttest was conducted in this study, and the long-term effect of article instruction remains unclear. Considering these factors, the notion of learnability and teachability of the English article system remains inconclusive, and warrants more research. More specifically, why is it that the learnability or teachability of other grammatical morphemes are hardly ever debated, while that of English articles is so controversial? I now turn to morpheme order studies, which will help us explain what aspects of the English article system account for this unique difficulty of its learning and teaching.

2.1.3 Morpheme Order Studies

The idea that various grammatical morphemes are acquired in a particular order was first proposed by Brown (1973), and was subsequently applied to L2 studies. Various studies have probed into this notion of acquisition order of morphemes in SLA (e.g., Dulay, Burt, & Krashen, 1982; Pienemann, 1998), and a well-cited, seminal study by Goldschneider and DeKeyser (2001) meta-analyzed those studies. In their meta-analysis, Goldschneider and DeKeyser (2001) identified six aspects of morphemes that determine their acquisition order, which was operationalized by the accuracy score. Such aspects included perceptual salience, semantic complexity, morphophonological regularity, syntactic category, frequency, and L1. However, the sixth variable, L1, was eventually excluded from their meta-analysis due to methodological constraints; that is to say, the studies reviewed in the meta-analysis by Goldschneider and DeKeyser (2001) did not group learners based on their L1s, making it impossible to gauge the effect of L1 transfer. The resultant multiple-regression model without L1 showed that 71% of the total variation in the accuracy scores can be accounted for by the combination of the abovementioned five aspects of morphemes ($R = .84$, $R^2 = .71$, $p < .001$) (p. 34).

Interestingly, a closer look at each of the articles' scores of the abovementioned five morpheme aspects reveals that articles' accuracy scores are higher than other morphemes in most of the five aspects, indicating that they are relatively easy to learn (p. 47). This discrepancy between the purported difficulty of the English article system and the highly reliable ($R^2 = .71$) multiple-regression model seems to be accounted for by the sixth variable excluded from their study: L1 effect. Despite this exclusion of L1 effect, they acknowledge the importance of this variable: L1 transfer "clearly is a factor that must be taken into consideration as one of the factors that could interact with morpheme acquisition and accuracy orders" (Anderson, 1978, p.

267, quoted in Goldschneider & DeKeyser, 2001, p. 31). Therefore, L1 effect could potentially account for the difficulty of the English article system, which the aforementioned five factors did not predict.

A more recent study by Murakami and Alexopoulou (2016) showed the effect of L1 on the acquisition order of various morphemes. Murakami and Alexopoulou (2016) conducted a large scale corpus study to address this problem, arguing that specific morphemes' "differential sensitivity to L1 influence" has seldom been discussed despite the well-documented L1 effects on morpheme acquisition order (pp. 394-395). Utilizing Cambridge Learner Corpus (CLC), they scrutinized approximately 10,000 written texts produced by English learners from seven different L1s (Japanese, Korean, Spanish, Russian, Turkish, German, and French) across five proficiency levels A2 – C2 in Common European Framework of References (CEFR). Particularly noteworthy in this study is the finding that the presence and absence of a specific morpheme in L1 exerts a strong influence on the morpheme acquisition order in English, and that the strength of the influence depends on the type of the morpheme. Specifically, Murakami and Alexopoulou found that L1 affected the acquisition of morphemes in the following order: articles (most susceptible to L1 effect), progressive *–ing*, plural *–s*, possessive *'s*, and third person *–s* (least susceptible to L1 effect) (p. 393). Therefore, Goldschneider and DeKeyser's (2001) exclusion of L1 effect could have led to the underestimation of the difficulty of articles, a morpheme that is most susceptible to L1 effect. In what follows I turn to the studies that have investigated the L1 effect in article acquisition.

2.2 Studies of L1 Effects on Article Development

Amongst the studies of L1 effects on article development, several aspects need to be taken into consideration such as how these studies have dichotomously compared [+ARTICLE]

and [-ARTICLE] languages (Section 2.2.1). This will be followed by the introduction of some of the rare studies that conducted comparisons of such L1 effects within [-ARTICLE] languages (Section 2.2.2). Once the differences of L1 effects within [-ARTICLE] languages are established, the next logical question is: why do those differences arise? Section 2.2 will conclude by proposing how to answer this question.

2.2.1 Comparisons between [+ARTICLE] [-ARTICLE] languages

As L1 effect has become more and more widely acknowledged, many studies have been conducted to investigate L1 effect on the acquisition of the English article system (e.g., Master, 1988; Snape et al., 2013). As is already discussed in the previous section, articles are found to be the most susceptible morpheme to L1 effect, and many studies have compared the article acquisition of learners whose L1s are [+ ARTICLE] and [- ARTICLE] languages. For example, Snape et al. (2013) compared how Spanish, Turkish, and Japanese learners of English use articles in the context of generic references. In their study, Spanish is treated as [+ ARTICLE] language as it has both definite and indefinite articles, and Japanese is treated as [- ARTICLE] language as it does not have any equivalent morpheme. Turkish is treated as an intermediate language in terms of the presence of article, as it only has a morpheme *bir*, which can represent a numeral "one" or an indefinite article "a/ an," depending on the way it is stressed. Snape et al. (2013) found a strong L1 effect in their production of articles in generic contexts. Whereas Spanish speakers enjoyed the overall higher accuracy than other L1 groups, Turkish and Japanese speakers performed relatively poorly on definite and indefinite articles, regardless of their proficiency levels. Both Turkish and Japanese speakers performed better on zero articles, and Snape et al. (2013) speculate that this is because they were more likely to drop articles as a result of the L1 transfer.

In contrast, what remains to be studied in the field of acquisition is the comparison *within* [- ARTICLE] languages, which is what the current study is set up to explore. An often-cited, classic study of article acquisition, which has incorporated multiple [+ ARTICLE] and [- ARTICLE] languages is Master (1988). In his study, he adopted a pseudo-longitudinal design by looking at learners from four proficiency levels in order to study their differential developmental trajectories based on L1s. Speakers of [+ ARTICLE] languages (i.e., Spanish and German) and [- ARTICLE] languages (Chinese, Japanese, and Russian) participated in this study. Major findings of this study include (a) similarities within article groups (i.e., [± ARTICLE]), (b) differences between article groups. Differences within article groups received relatively little attention in this study, and this aspect warrants further investigation. Even though this pivotal study has been influential, it is not without methodological problems. Here, key problems relevant to the present study are identified.

First, Master (1988) is too small in scale for its findings to be generalized. It consisted of 20 groups (four proficiency levels and five L1s), and each group was represented by only one participant. Master (1988) acknowledges this problem as his study's limitation: "The greatest need for further research, as has been mentioned throughout this study, is the reduplication of the present study with a much larger number of subjects" (p. 38).

Second, the operationalization of the English proficiency level makes the comparison with other more recent studies difficult. Master (1988) uses "negation criteria" to categorize the participants into four proficiency levels. Developed by Canino, Rosansky, and Schumann (1978), these negation criteria label learners as "basilang," "mesolang," and "acrolang," based on the characteristics observed from the learners' production of negation. However, negation criteria are not a holistic proficiency measure; rather, it is an emergence criterion that measures the mastery

of a certain linguistic structure (Pallotti, 2007). For this reason, findings in Master (1988) are difficult to compare with other studies that operationalize learner proficiencies by a different measure, such as Common European Framework of Reference (CEFR).

Third, the operationalization of the article "usage" and "accuracy" does not allow us to capture the precise picture of the learners' interlanguage development. Master (1988) uses "Suppliance in Obligatory Context" (SOC), developed by Brown (1973), and his own measure "Used in Obligatory Context" (UOC). These two measures are expressed as:

$$SOC = \frac{Number\ of\ correct\ usages}{Number\ of\ obligatory\ contexts};\ UOC = \frac{Total\ Number\ of\ usages}{Total\ Number\ of\ obligatory\ contexts}$$

The presence and absence of "total" indicates whether one-to-one correspondence of the usage and context is considered or not. In other words, SOC identifies a certain number of obligatory contexts of a particular morpheme, and for each context, (in)correct usages are identified. The biggest concern of SOC is that it is not capable of accounting for overgeneralization of a particular morpheme. For example, in the case of articles, if a learner uses a definite article *the* for all noun phrases, the learner will receive the SOC score of 100% for *the* because SOC only looks at the obligatory context, where the use of *the* is required. UOC, on the other hand, does not have a one-to-one correspondence. It is a simple division of the total number of times a particular morpheme was used by the total number of times the morpheme was required. To account for this drawback in the usage and accuracy measure, Pica (1984) devised the Target Like Use (TLU) measure. TLU is a revised version of SOC and it is capable of accounting for overgeneralization by adding the number of incorrect uses of a particular morpheme to the denominator:

$$TLU = \frac{Number\ of\ correct\ uses}{Number\ of\ obligatory\ contexts + Number\ of\ incorrect\ uses}$$

These problems can be overcome with the use of a corpus-based approach. The first problem, namely the number of participants, can be taken care of with the use of large-scale learner corpora, such as *The International Corpus of Learner English* (ICLE; Granger, Dagneaux, Meunier, & Paquot, 2002), *The International Corpus Network of Asian Learners of English* (ICNALE; Ishikawa, 2011, 2013), the *Cambridge Learner Corpus* (CLC), and The EF-Cambridge Open Language Database (EFCAMDAT; Huang, Murakami, Alexopoulou, & Korhonen, 2018; Geertzen, Alexopoulou, & Korhonen, 2013). For example, EFCAMDAT contains more than 83,000,000 word tokens in the essays written by more than 170,000 learners of English (Huang, Geertzen, Baker, Korhonen, & Alexopoulou, 2017).

The second problem, namely the operationalization of learner proficiency levels, can also be taken care of, with the rich learner variables already included in such learner corpora. For example, even though many corpora only include an institutional proficiency variable (e.g., third year undergraduate), which has been shown to be unreliable (Carlsen, 2012), some (e.g., CLC, EFCAMDAT) provide the researchers with each learner's proficiency level based on Common European Framework of Reference (CEFR), a more reliable measure of proficiency. I now turn to corpus-based studies that addressed the abovementioned problems.

2.2.2 Corpus-based comparisons within [- ARTICLE] languages

Few studies have taken a corpus-based approach to compare L1 effects within article-less languages, and Crosthwaite (2016) is one of such rare studies. He adopted ICNALE to compare the developmental patterns of article usage of Chinese, Korean, and Thai learners of English. Particularly noteworthy was that this study was able to overcome the abovementioned problems: i.e., a large number of written texts were investigated to enhance the generalizability of the findings (the first problem); CEFR-based proficiency categorization was included as a learner

variable in ICNALE (the second problem); and TLU score, rather than SOC or UOC, was used to accurately quantify the learners' article use (the third problem). Furthermore, to prevent the variation of essay prompts from potentially confounding the analysis, Crosthwaite only included two essay prompts into his study, and conducted separate analyses based on the essay prompt, as well as a combined analysis to analyze the overall tendency.

Among the major findings of Crosthwaite (2016), particularly relevant to the present study is that Chinese speakers outperformed other L1 speakers in definite, indefinite, and zero article use: "The TLU performance of the Mandarin L2 English group was consistently higher than that of other L2 groups for definite, indefinite and zero article production" (p. 94). This study is informative in that it clearly showed the difference *within* [- ARTICLE] languages: an aspect of article acquisition research that has long been understudied. Crosthwaite (2016) attributes this Chinese speakers' outperformance to L1 transfer: "Given the claims regarding definiteness marking in Mandarin alongside the higher use of demonstratives in definite referential contexts in the small sample of L1 data provided in the present study, the higher TLUs of the Mandarin L2 English data may be explained as a significant effect of positive L1 transfer into L2 English article production" (p. 94). This explanation warrants more investigation; for example, do other [- ARTICLE] languages with rich demonstratives (e.g., Japanese) outperform [- ARTICLE] languages without demonstratives, like the Chinese speakers did in Crosthwaite (2016)?

A clue to answer this question is found in Han, Chodorow, and Leacock (2006). They devised a machine training (MT) model to automatically detect and annotate (in)correct use of articles, and tested it with Chinese, Japanese, and Russian learners of English. They used a total of 664 TOEFL essays to measure the accuracy of automatic error tagging of articles.

Unfortunately, no cross-linguistic comparisons were provided, which is reasonable considering that the aim of their study was to develop a reliable MT model for automatic detection of article errors.

However, some data provided in Han et al. (2006) are relevant to the present study. Han et al. (2006) compared the proportion of the text units (i.e., essay, sentence, and noun phrase) that included one or more article errors. The results showed that Japanese learners produced more texts with at least one article error than Chinese learners (98% and 95%, respectively), more of such sentences (34% and 30%, respectively), and more of such noun phrases (NPs) (15% and 12%, respectively). Whether these differences are statistically significant or not remains unclear, but considering the large number of TOEFL essays produced by Japanese and Chinese learners (234 and 225 essays, respectively), it is worth investigating why these differences arose, despite the presence of similarly rich demonstratives in the two languages.

Despite numerous advantages abovementioned corpus-based studies brought about, the common limitation that characterize these studies is that the level of explanation they can afford is limited to a descriptive level. In other words, although they provide us with insights into *how* learners with various L1s differ from each other based on the accuracy of their use of English articles, *why* they do so is yet to be deciphered. This is an extremely challenging, if not impossible, problem to address, given the large number of rules and exceptions that govern the use of English articles. However, one way to tackle this problem is through Multifactorial Prediction and Deviation Analysis with Regressions (MuPDAR; Gries & Deshors, 2014). Methodological and statistical details of this approach will be explained in detail in Section 3. Conceptually, though, MuPDAR allows us to quantify how much the use of a particular linguistic structure by NNS deviates from that of NS, through building regression models on

richly annotated corpus data. Hence, I will now turn to various semantic and morphosyntactic factors that have been shown to affect the use of English articles by NNS (e.g., Liu & Lu, 2020; Master, 1994).

2.3 Contextual Factors Affecting the Use of English Articles

This section reviews studies that have investigated various factors affecting the use of English articles. Specifically, semantic and morphological factors relevant to nouns themselves, such as noun countability, will be first introduced (Section 2.3.1). Secondly, a discourse level factor that plays a central role in the article use will be reviewed (Section 2.3.2). Lastly, other relevant factors, such as syntactic modifications of nouns, that are less frequently studied in the existing literature will be introduced (Section 2.3.3).

2.3.1 Noun Countability, Noun Animacy, Number, and Noun Type

Among the most challenging factors that govern the use of English articles is noun countability (Master, 1988, 1994; Butler, 2002). As Master (1994) points out the importance of the distinction between countable and uncountable nouns, it poses a particular challenge to English learners. In a more recent study, Liu and Lu (2020) concluded, through a series of grammaticality judgment test, forced elicitation test, and the subsequent stimulated recall, that the most important factor contributing to the misuse of English articles by Chinese EFL learners was their misconception of noun countability.

Noun countability is not a straightforward notion, as it varies not only from noun to noun (e.g., **car** is countable whereas **water** is not), but also from context to context (e.g., **life** as a particular person's life is countable, whereas **life** as a general state of existence is not). Yet, the former (i.e., noun to noun variation) is more straightforwardly interpretable, and Doetjes (2017) maintains that "both agents and cohesive objects are normally denoted by count nouns, as in two

dogs or three pens" and that "agentivity is associated with animacy" (p. 201). Therefore, Noun Animacy, as a semantic factor underlying the property of countability, also appears to be an important factor affecting the use of English articles.

It follows naturally then, that pluralized nouns will be relatively easy in terms of the selection of the article, because the presence of the plural marker *–s* signifies that the noun is countable. This is indeed the case; Leroux and Kendall (2018), in their corpus-based analysis of Chinese EFL learners' acquisition of the English article system, conducted a regression analysis and found that article use with pluralized nouns was the most accurate. In the regression analysis, they also found that, among other significant predictors, the effect of pluralization had a significant positive effect on the accurate use of articles ($b = 1.44$, $se = 0.22$, $p < .001$). Hence, in addition to noun countability, it is also worth incorporating into analysis whether or not a noun is morphologically marked as plural.

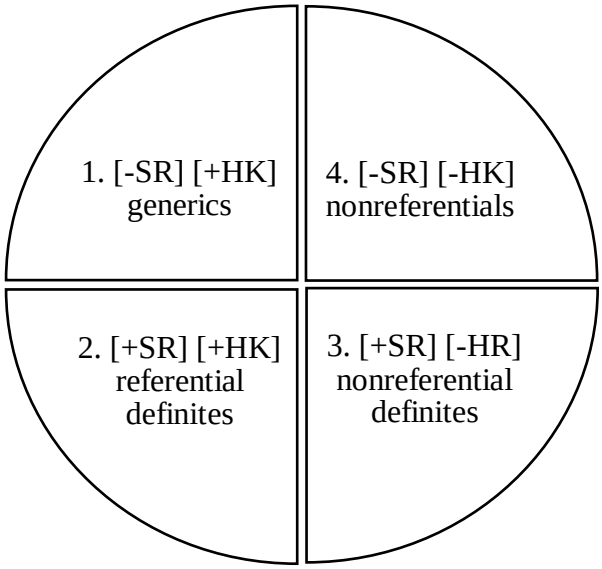
The notion of countability and its effect on the choice of articles have a controversial relationship with the noun type, namely, the distinction between common and proper nouns. Master (1994), comparing ***a*** *doctor* and ***Ø*** *Dr. Smith* as examples, treats noun type as an independent property of a noun, showing that proper nouns are more likely to be used with zero articles. White (2010), on the other hand, maintains that the article selection for proper nouns should be done in the same way as it is done for common nouns: "Through the framework, the basic principles of article use are the same for proper nouns as they are for common nouns. The key is accepting an unbounded non-individuated construal for proper nouns with Ø" (p. 75). Therefore, the effect of noun type on the use of article by NNS warrants more investigation.

It is perhaps not surprising, that this intricate notion of noun countability is merely one of the many factors governing the use of articles. This is clearly pointed out by Snape (2008), in

which he showed that the accurate distinction of countable and uncountable nouns did not necessarily equate with the accurate use of articles. He points to the importance of definiteness, which I will now turn to, given this paradox: "a continuing difficulty with definites may lie in the types of definites (e.g. anaphoric, encyclopaedic, situation) that co-occur with nouns in context" (Snape, 2008, p. 74).

2.3.2 Definiteness

The notion of Definiteness is central to the use of English articles. A common scheme for definiteness is Bickerton (1981)'s semantic wheel, which was subsequently revised by Huebner (1983, 1985) and is still widely used in corpus-based learner language research (e.g., Butler, 2002; Crosthwaite, 2016; Diez-Bedmar & Papp, 2008; Diez-Bedmar, 2015; Leroux & Kendall, 2018). This scheme categorizes NPs into four semantic contexts based on the presence and absence of Hearer's Knowledge (HK ±) and Specific Referent (SR ±). Figure 1.1 shows the graphical representation of this categorization scheme.



1. [-Specific Referent, +Assumed Known to the Hearer]: Generics
2. [+Specific Referent, +Assumed Known to the Hearer]: Referential Definites
    a. Unique or conventionally assumed unique referent;
    b. Referent physically present;
    c. Referent previously mentioned in discourse;
    d. Specific referent otherwise assumed common knowledge
3. [+Specific Referent, -Assumed Known to the Hearer]: Referential Indefinites
   First mention of NP [+SR] in a discourse and assumed not common knowledge.
4. [-Specific Referent, -Assumed Known to the Hearer]: Non-Referentials
    a. Equative noun phrases
    b. Noun phrases in the scope of negation
    c. Noun phrases in scope of questions, irrealis mode

Figure 2.1. Semantic wheel for noun phrase reference (Huebner, 1985)

The first two categories are both [+HK], meaning that the entity is assumed known to the hearer. The former (Category 1) is a known entity without specific referent (e.g., A **cat** likes mice), and the latter (Category 2) is a known entity with a specific referent (e.g., Pass me the **pen**). The other two categories are both [-HK], meaning that the entity is assumed unknown to the hearer. The former (Category 3) is an unknown entity with a specific referent, such as first mention (e.g., I saw a strange **man**), and the latter (Category 4) is an unknown entity without a specific referent

(e.g., He used to be a **lawyer**). All these examples were adopted from Butler (2002, pp. 478-479); for a more complete set of examples, see Butler (2002).

However, even after the idiomatic expressions and conventional uses were added as the fifth category by Thomas (1989), this 2 × 2 (+ 1) categorization scheme is not adequate, as it does not capture the full variability of the notion of definiteness. That is to say, differences *within* each of the five types of definiteness should also be taken into account; for example, in Figure 2.1, within the single category 2 [+SR, +HK], four types of examples are given. Lumping all these four types *within* category 2 would be problematic, if learners have varying degrees of problems among these four types.

This issue seems to be addressed in a more fine-grained coding scheme for communicative functions of definiteness (CFD; Bhatia, Simons et al., 2014; subsequently modified in Bhatia, Lin et al., 2014a). This coding scheme takes a hierarchical structure, as shown in Figure 2.2.

Nonanaphora

- Unique
  - Unique_Hearer_Old
    - Unique_Physical_Copresence
    - Unique_Larger_Situation
    - Unique_Predicative_Identity
  - Unique_Heaerer_New
- Nonunique
  - Nonunique_Hearer_Old
    - Nonunique_Physical_Copresence
    - Nonunique_Larger_Situation
    - Nonunique_Predicative_Identity
  - Nonunique_Hearer_New
  - Nonunique_Nonspecific
- Generic
  - Generic_Kind_Level
  - Generic_Individual_Level

Anaphora

- Basic_Anaphora
  - Same_Head
  - Different_Head
- Extended_Anaphora
  - Bridging_Nominal
  - Bridging_Event
  - Bridging_Restrictive_Modifier
  - Bridging_Subtype_Instance
  - Bridging_Other_Context
- Miscellaneous
  - Pleonastic
  - Quantified
  - Predicative_Equative_Role
  - Part_Of_Noncompositional
  - Measure_Nonreferential
  - Other_Nonreferential

Figure 2.2. Unified Annotation Scheme of CFD (adopted from Bhatia, Lin et al., 2014a)

As shown in Figure 2.2, CFD categorizes various types of definiteness based on its tree structure (examples of each category are given in Appendix A). The highest order distinction categorizes all noun phrases (NPs) into the following three intermediate nodes: *Nonanaphora, Anaphora,* and *Miscellaneous. Nonanaphora* refers to entities that are discourse-new, and it further ramifies into *Unique, Nonunique,* and *Generic. Unique* refers to uniquely identifiable entities such as Barack Obama, whereas *Nonunique* refers to unidentifiable entities. *Generic* refers to the entire genre rather than an individual case. *Anaphora* refers to entities that are previously mentioned or evoked in the discourse, and it further ramifies into *Basic* and *Extended Anaphora*. The former refers to the entities that have been mentioned in the discourse, whereas the latter refers to the entities that have not been directly mentioned, but evoked by indirect allusion. *Miscellaneous* refers to all other kinds of entities that do not fit into either *Nonanaphora* or *Anaphora*, such as a part of an idiomatic expression (e.g., in **fact**).

Even though, to my knowledge, this coding scheme has never been used in SLA research, it was deemed more informative and useful than the traditional semantic wheel because CFD overcomes the abovementioned problem that the semantic wheel lumps together different kinds of definiteness into one category. For example, in semantic wheel, different kinds of hearer's knowledge are all subsumed under [+HK], but CFD makes a finer distinction of hearer's knowledge, such as Nonunique_Physical_Copresence and Extended_Anaphora. The former is when the referred entity is known to the hearer because it is physically present at the moment of the speech, and the latter is when the referred entity is known to the hearer because it has been indirectly evoked in the previous discourse.

In addition to the variables central to the use of English articles that have been discussed in this section thus far, other factors, such as syntactic modification, are also reported to affect the use of English articles (e.g., Lee, 1999). I will now turn to those factors.

2.3.3 Other Factors

Other factors include syntactic modification of the NPs and the verb type that the noun is associated with. For syntactic modification, Lee (1999) hypothesized that the presence of a prenominal modification will increase the rate of ungrammatical omission of an article. Based on the Korean EFL learners' article use in written compositions, Lee (1999) concluded that the hypothesis was confirmed. Even though this result might not generalize for L1s other than Korean, syntactic prenominal and post-nominal modifications warrant more investigation, as Master (1994) makes a generalization that post-modified nouns are more likely to elicit a definite article than pre-modified nouns do. Ionin, Ko, and Wexler (2004) maintain that this use of the definite article associated with the presence of post-modification might be a widely accepted

strategy, and that it aligns with the overproduction of the definite article for post-modified nouns observed in their data (p. 54).

This effect of prenominal and post-nominal modifications is also shown to predict the seemingly random use of the definite article in organizational names, such as ***the*** *United States*. Tse (2001) conducted a logistic regression analysis on the presence and absence of the definite article preceding proper nouns, with various syntactic modifications as independent variables. Tse (2001) found that organizational names with pre-modification with a proper noun and the ones with post-modification were more likely to have no articles, whereas the ones with pre-modification with a common noun were more likely to have a definite article.

Another potential factor that affects the use of English articles is the verb the noun is associated with. Teng (2012) compared the accuracy of article use by Japanese EFL learners, and found that the sentences with the verb *be* led to a higher rate of ungrammatical omission of articles for countable nouns, compared to the sentences with other verbs. Teng (2012) hypothesizes that the verb *be* might have enhanced the perception of the associated noun as an abstract idea, rather than a concrete entity. This study did not control for the type of the association between the noun and the verb (e.g., subject, object); therefore, a more generalized investigation of the effect of verb type and syntactic Case needs to be conducted.

2.3.4 Multifactorial Approach

I have reviewed a number of semantic, morphological, and syntactic factors that affect the use of English articles. As has already been touched upon briefly in Section 2.2.2, however, it is impossible to investigate how strongly each of these factors contributes to the NNS' linguistic choice or how they interact with each other, without simultaneously including all of these factors into the analysis. Gries and Deshors (2014) championed the importance of this multifactorial

approach that "takes factors from many different levels of linguistic analysis into consideration" (p. 112), arguing that "such multifactorial statistical analyses of corpus data are yet to be widely adopted in SLA research" (p. 112).

Since this call for the use of a multifactorial approach in SLA research and the development of MuPDAR approach by Gries and Deshors (2014), various linguistic phenomena have been studied with this approach, such as English modal verbs *may* and *can* (Deshors, 2016), verb complementation patterns (Deshors & Gries, 2016), prenominal adjective order preferences (Wulff & Gries 2015), and an optional relativizer *that* (Lester, 2019). This approach has afforded a highly fine-grained analysis of NNS' linguistic choice that deviates from NS'. For example, Lester (2019) compared the use of an optional English relativizer *that* by two groups of NNS (Spanish and German L1) and English NS, and found that whereas NS were more likely to use the optional *that* in a linguistic context with structural complexity, the opposite was true about both groups of NNS.

Despite the increasing popularity of this powerful analytical tool, MuPDAR, it is not immediately applicable to the present study because all MuPDAR studies to date have investigated linguistic structures with binary choices, and its multinomial application is yet to be undertaken. In this context, because English articles require learners of English to make a choice from three options (DA, IA, and ZA), the present study makes the first attempt to apply MuPDAR to a linguistic structure with non-binary choices. The details of this multinomial application of MuPDAR will be further discussed in Section 3.

2.4 Research Questions

In light of all the above, the present study aims to answer the following three research questions:

1. How do various semantic, morphological, and syntactic factors affect the use of English articles by NNS?
2. How do such effects differentially affect three types of English articles; namely, definite, indefinite, and zero articles?
3. How do 1 and 2 differ for different L1 backgrounds, especially within article-less languages?

3 METHODOLOGY

The first part of this section (Section 3.1) consists of the introduction of corpus data, data extraction process, and the annotation scheme. The second part of this section (Section 3.2) will be dedicated to the explanations of the statistical evaluation.

3.1 Corpus Data and Annotation

In this section, I will first introduce the corpora used in this study, and then show the data extraction process; namely, how the noun phrases (NPs) were extracted from the corpora, with a particular focus on the inclusion/ exclusion criteria of the NPs. Thirdly, I will introduce the annotation process and the annotation scheme in tables.

3.1.1 Corpora

Because both native speaker (NS) data and non-native speaker (NNS) data are necessary, I used The EF-Cambridge Open Language Database (EFCAMDAT; Huang, Murakami, Alexopoulou, & Korhonen, 2018; Geertzen, Alexopoulou, & Korhonen, 2013) for NNS data, and The Louvain Corpus of Native English Essays (LOCNESS; Granger, 1998) for NS data. A brief overview and comparison of the two corpora is presented in Table 3.1.

Table 3.1. *The Two Corpora Used in the Study*

	EFCAMDAT	**LOCNESS**
Type	Learner Corpus	NS Corpus
Size	83,000,000+ words	324,304 words
L1	198 NNS varieties	2 NS varieties
Proficiency	16 levels (NNS)	College-level (NS)
Format	.xml	.txt
Access	Open	Open

As shown in Table 3.1, EFCAMDAT is a large-scale learner corpus with over 83 million words, in which 198 NNS varieties are represented. It is important to note, however, that this is based on

the self-reported nationalities, and that these NNS varieties are "the closest approximation to L1 background" (Huang, Geertzen, Baker, Korhonen, & Alexopoulou, 2017, p. 4). Furthermore, it includes various demographic information, such as proficiency levels. The 16 proficiency levels correspond to a widely-used proficiency measure, namely, Common European Framework of Reference (CEFR) levels, A1 – C2. As to the accessibility, EFCAMDAT is accessible upon completion of the user registration, and its online search engine allows for the downloading of relevant sub-corpora.[1] The data are stored in a mark-up language (XML), which allows for the use of a range of automatic processing.

LOCNESS, on the other hand, is a corpus of essays written by NS, in which two English varieties, namely, American and British Englishes, are represented. As opposed to EFCAMDAT, LOCNESS provides the data in a text file format (.txt), and the range of preprocessing tools applicable to this data format is rather limited. This will be further discussed in Sections 3.1.2 and 3.1.3. Lastly, it is open to the public without user registration, and the whole corpus is downloadable without specifying sub-corpora through a search engine.[2]

3.1.2 Data extraction process

Target tokens (i.e., articles) were extracted from the abovementioned two corpora (EFCAMDAT and LOCNESS) in the following ways. For learner language, relevant essay topics, L1s, and proficiency levels were selected from EFCAMDAT and downloaded as an XML file. Initially, 5 L1s (English, Japanese, Chinese, Russian, and Korean) were to be included in the analyses; however, to ensure enough number of occurrences in each L1 group for subsequent statistical analyses, Russian and Korean were excluded from data annotation.

[1] https://corpus.mml.cam.ac.uk/efcamdat2/public_html/
[2] https://uclouvain.be/en/research-institutes/ilc/cecl/locness.html

For essay topics, EFCAMDAT has the total of 126 essay topics across 16 proficiency levels. As essay prompts are reported to affect the accuracy of certain article forms (Crosthwaite, 2016), a care was taken to ensure the comparability between the two corpora. Because the essays written by English native speakers in LOCNESS are argumentative essays, personal topics were excluded from EFCAMDAT. More specifically, once the list of 126 essay prompts across 16 levels and 6 proficiency groups (A1 – C2) was extracted, each of them was examined carefully based on (a) word count requirement, (b) writing format (e.g., email, letter, list, etc.), and (c) nature of the prompt. For (a), because essays in LOCNESS are 500 words or more, essays with fewer word counts were removed. However, because the longest word count requirement was 150 – 180 words in EFCAMDAT, an arbitrary cutoff point of at least 100 words was made. For (b), writing format that affects the discourse level variable was removed; for example, prompts that elicited bullet points were excluded. For (c), topics that are either non-argumentative or personal were excluded. A letter to a friend, email to a teacher, formal apology, and apartment lease are among the examples of the topics excluded from this study. For the final list of topics, see Appendix A.

As a result, levels A1–B1 and C2 were excluded because these levels did not have enough NPs for each L1 group once the exclusion criteria (a) – (c) were applied. For example, A1 and A2 did not meet the criteria because every single essay prompt in these levels were too short (either 20-40 or 50-70 words), and the topics were too personal (e.g., "write an email to your teacher to introduce yourself").

For NP extraction, due to the difference in file format, different steps were required for EFCAMDAT and LOCNESS. For EFCAMDAT, the most straightforward way was to extract all determiners and the NPs they precede. However, this approach cannot extract NPs with zero

articles. Hence, this study took a backward approach—all NPs and the preceding determiners were extracted, and irrelevant ones were identified and subsequently removed. This was done through Python syntax. Concretely, nouns that are either (a) preceded by quantifiers (e.g., **some** people, **any** reason), (b) preceded by demonstratives (e.g., **this** man, **these** people), (c) preceded by possessives (e.g., **his** car), or (d) functioning as a noun modifier (e.g., **credit** card, **bank** account) were all removed. Nouns that are irrelevant to the choice of determiners (e.g., **something, anything**) were also removed. These processes removed approximately a third of the NPs. After the extraction and removal processes, all the NPs were exported into an Excel sheet for the subsequent annotation. Each column in the Excel sheet corresponded to each variable that is described in the following section.

For LOCNESS, due to its file format (i.e., .txt), TagAnt (Anthony, 2016) was employed for automatic part of speech (POS) tagging. After the POS-tagged text file was generated, all NPs that are tagged as either NN, NNS, NNP, or NNPS were extracted and exported into an Excel sheet for further annotation. Because the automatic removal processes described above were not applicable to the native speaker data due to its file format, irrelevant cases of NPs were manually removed one by one. The exclusion criteria were the same as the ones used for the learner data. The descriptive statistics for the extracted tokens are presented in Table 3.2, with a breakdown of how many definite (DA), indefinite (IA), and zero articles (ZA) were used in each of the first language groups.

Table 3.2. *Descriptive Statistics for the Annotated Tokens of Articles*

L1	# Essays	Word Counts (per essay)	# Tokens	DA (%)	ZA (%)	IA (%)
English	15	4992 (332.8)	833	245 (29%)	479 (58%)	109 (13%)
Chinese	25	4822 (192.88)	795	221 (28%)	480 (60%)	94 (12%)
Japanese	24	4603 (191.79)	833	290 (35%)	471 (57%)	72 (9%)
Total	64	14417 (225.27)	2461	756 (31%)	1430 (58%)	275 (11%)

Table 3.2 shows that essays written by NS (M = 332.8) are substantially longer than the ones written by Chinese (M = 192.88) and Japanese (M = 191.79) speakers. There was no fix to this problem, as the word count requirements in EFCAMDAT and in LOCNESS were different, with LOCNESS requiring a much longer essay. This problem and its implications will be further discussed in the limitation section.

3.1.3 Data annotation scheme

Each of the 2,461 occurrences of articles was annotated[3] for the following 12 variables, presented in Table 3.3.

Table 3.3. *Overview of the Variables Used in Annotation*

Types of variables	Variables	Number of Levels
Semantics	NOUNCOUNT	2
	NOUNANIMACY	12
	NOUNTYPE	2
	DEFINITENESS	8
	VERBTYPE	5
Morphological	FORM	3
	NUMBER	2
Syntactic	MODIFICATION	4
	CASE	4
Data	L1	3
	PROFICIENCY	4
	ID	32

[3] I was the sole annotator of this annotation process, and I acknowledge that the untested interrater reliability is one of the limitations of this study.

In Table 3.3, the column "Types of variable" refers to the category to which each of the variables belongs. For example, variables labeled as semantics pertain to the meaning of the target NP. The column "Variables" refers to the name of the variable, and the column "Levels" indicates how many levels each of the variables has. In what follows, I present a detailed description of each of the following 12 variables in the order in Table 3.3. The first variable NOUNCOUNT is presented in Table 3.4.

Table 3.4. *The NOUNCOUNT Variable and its Levels*

Type of variables	**Variable Name**	**Levels**
Semantics	NOUNCOUNT	countable uncountable

The variable NOUNCOUNT was annotated in the following way. First, all NPs that are tagged as either "NNS" or "NNPS" were automatically annotated as "countable" because only countable nouns can be pluralized. For the rest of the NPs that are tagged as either "NN" or "NNP," as they can be either singular countable nouns or mass noun, this distinction was manually annotated. In this manual annotation process, whenever the countability was unclear, an online English dictionary[4] was used as a reference. Because the same noun can be countable *or* uncountable depending on the meaning it conveys in a particular context, the closest meaning was identified in the dictionary, and the corresponding countability was annotated in the data. Examples below show that the NP *life* in (1) is countable because it refers to a particular course of life, whereas the one in (2) is uncountable because it means general human existence.

(1) Today, having knowledge of how the computer operates is considered a necessary component of leading a successful **life** (ICLE-US-MICH-002.1).

[4] *Longman Dictionary of Contemporary English Online* (https://www.ldoceonline.com/) was used.

(2) Cars, telephones, and nuclear energy are just three examples of inventions and discoveries that have had profound effects on modern day **life** (ICLE-US-MICH-0035.1).

The variable NOUNANIMACY is presented in Table 3.5.

Table 3.5. *The NOUNANIMACY Variable and its Levels*

Type of variables	Variable Name	Levels
Semantics	NOUNANIMACY	non-human human natnl/group/socrole other abstract dynamic ling eff/state mental/emotional natural entity place/time social-conv

This variable NOUNANIMACY was adopted from Deshors (2016). The original variable had 23 levels; however, it was eventually conflated into 12 levels. For examples of each of these levels, see Appendix A. The conflation process is presented in Table 3.6.

Table 3.6. *The Conflation Process of the Variable* NOUNANIMACY

Original Levels	**Conflated Levels**
animal flora	non-human
human	human
nationals group social roles	natnl/group/socrole
object/ artifact scholarly work form/ substance imaginary beings absence measure	other
abstract	abstract
action process	dynamic
dummy 'it' (pseudo) cleft structure	ling
effect state	eff/state
mental/ emotional	mental/emotional
natural entity	natural entity
place/ time	place/time
social Convention	social-conv

As show in Table 3.6, the original variable with 23 levels was conflated into 12 levels (Deshors, 2016, p. 143). Examples of each of the NOUNANIMACY types are presented in Appendix A. For more details on the statistical and conceptual validity of this conflation, see Deshors (2016).

The variable NOUNTYPE is presented in Table 3.7.

Table 3.7. *The* NOUNTYPE *Variable and its Levels*

Type of variables	**Variable Name**	**Levels**
Semantics	NOUNTYPE	common proper

This variable NOUNTYPE was annotated automatically, based on the part-of-speech tag provided by EFCAMDAT. Because proper nouns are tagged as either NNP (singular) or NNPS (plural), and common nous as NN (singular) or NNS (plural), the first two were automatically annotated

as proper nouns, and the other two as common nouns. For example, the NP *America* in (3) was annotated as proper, and *humanness as* common.

(3) In **America** , this growing individualistic society, one no longer sees the realitive **humanness** between people (ICLE-US-MICH-0005.1).

The variable DEFINITENESS is presented in Table 3.8.

Table 3.8. *The* DEFINITENESS *Variable and its Levels*

Type of variables	Variable Name	Levels
Semantics	DEFINITENESS	Unique Hearer Old (uniq_hear_old) Unique Hearer New (uniq_hear_new) Non-Unique Hearer Old (nonuni_hear_old) Non-Unique Hearer New (nonuni_hear_new) Non-Unique Non-Specific (nonuni_nonspe) Generic (generic) Basic Anaphora (bas_anaph) Extended Anaphora (ext_anaph) Miscellaneous (misc)

This variable DEFINITENESS originally had 24 levels, but it was conflated into 9 levels for the ease of annotation. Examples for each of the 9 levels will require more than just a sentence as this discourse-level variable is suprasententially defined; for a simpler list of examples for each of the DEFINITENESS levels, see Appendix A. For the annotation, Bhatia, Lin et al. (2014a) developed an automatic classifier of this coding scheme; however, the classification accuracy was not sufficient and was therefore not adopted in this study. Table 3.9 summarizes the conflation process of the variable *DEFINITENESS.*

Table 3.9. *The Conflation Process of the Variable* DEFINITENESS

Original Levels	**Conflated Levels**
Unique_Physical_Copresence Unique_Larger_Situation Unique_Predicative_Identity	Unique Hearer Old (uniq_hear_old)
Unique_Hearer_New	Unique Hearer New (uniq_hear_new)
NonUnique_Physical_Copresence NonUnique_Larger_Situation NonUnique_Predicative_Identity	Non-Unique Hearer Old (nonuni_hear_old)
NonUnique_Hearer_New_Spec	Non-Unique Hearer New (nonuni_hear_new)
NonUnique_NonSpec	Non-Unique Non-Specific (nonuni_nonspe)
Generic_Kind_Level Generic_Individual_Level	Generic (generic)
Same_Head Different_Head	Basic Anaphora (bas_anaph)
Bridging_Nominal Bridging_Event Bridging_Restrictive_Modifier Bridging_Subtype_Instance Bridging_Other_Context	Extended Anaphora (ext_anaph)
Pleonastic Quantified Predicative_Equative_Role Part_Of_Noncompositional_MWE Measure_Nonreferential Other_Nonreferential	Miscellaneous (misc)

Originally, the variable DEFINITENESS had 25 levels, and they were conflated into nine levels as shown in Table 3.9, for the ease and accuracy of annotation. The conflation was mainly based on the original hierarchical structure proposed in Bhatia, Simons et al. (2014), which is shown in Figure 3.1. The nine underlined, boldfaced levels were kept after conflation.

Nonanaphora

- Unique
  - **Unique_Hearer_Old**
    - Unique_Physical_Copresence
    - Unique_Larger_Situation
    - Unique_Predicative_Identity
  - **Unique_Heaerer_New**
- Nonunique
  - **Nonunique_Hearer_Old**
    - Nonunique_Physical_Copresence
    - Nonunique_Larger_Situation
    - Nonunique_Predicative_Identity
  - **Nonunique_Hearer_New**
  - **Nonunique_Nonspecific**
- **Generic**
  - Generic_Kind_Level
  - Generic_Individual_Level

Anaphora

- **Basic_Anaphora**
  - Same_Head
  - Different_Head
- **Extended_Anaphora**
  - Bridging_Nominal
  - Bridging_Event
  - Bridging_Restrictive_Modifier
  - Bridging_Subtype_Instance
  - Bridging_Other_Context
- **Miscellaneous**
  - Pleonastic
  - Quantified
  - Predicative_Equative_Role
  - Part_Of_Noncompositional
  - Measure_Nonreferential
  - Other_Nonreferential

Figure 3.1. Unified Annotation Scheme (adopted from Bhatia, Lin et al., 2014a)

In addition to the hierarchical structure shown in Figure 3.1, I also took into account the distinction relevant to Bickerton (1981)'s semantic wheel; namely, Hearer Knowledge [HK±] and Specific Referent [SR±] during the conflation process. For example, because the [HK±] and [SR±] distinctions were present within *Nonanaphora*, this distinction was retained in the conflation process. Consequently, from the level *Nonanaphora* in Figure 3.1, the following six levels were retained: *Unique_Hearer_Old, Unique_Hearer_New, Nonunique_Hearer_Old, Nonunique_Hearer_New, Nonunique_Nonspecific*, and *Generic*. For *Anaphora*, because it is important to make a distinction between the anaphoric NPs that have actually been mentioned and the ones that have only been evoked by entities mentioned before, *Basic_Anaphora* and *Extended_Anaphora* were retained. Lastly, other types that fall under miscellaneous were conflated into one level *Miscellaneous* because they are not explicable by anaphoricity, [±HK], or [±SR].

The variable VERBTYPE is presented in Table 3.10.

Table 3.10. *The* VERBTYPE *Variable and its Levels*

Type of variables	Variable Name	Levels
Semantics	VERBTYPE	stative activity achievement accomplishment n/a

This variable VERBTYPE is based on the taxonomy developed by Vendler (1957). For the nouns that were either subject (nominative case) or object (accusative case) of a verb, lexical aspect of the verb was annotated. For nouns that did not receive any syntactic case, its VERBTYPE was annotated as n/a.

The variable MODIFICATION is presented in Table 3.11.

Table 3.11. *The* MODIFICATION *Variable and its Levels*

Type of variables	Variable Name	Levels
Syntactic	MODIFICATION	Pre-modification with adjective (premod_a) Pre-modification with noun (premod_n) Post-modification with prepositional phrase (postmod_p) Post-modification with relative clause (postmod_rc) Post-modification with infinitival clause (postmod_ic) Post-modification with complement clause (postmod_cc)

Originally, these six levels were configured as a single variable "MODIFICATION". However, because noun phrases can have multiple modifications (e.g., a **big** house **in the city**), it was not ideal to annotate this variable with a six-level multinomial (i.e., single-label) variable. Instead, in order to treat this single variable "MODIFICATION" as a multi-label variable, it was separated into six variables, each of which was then treated as a two-level binary variable. The underlined NP in example (4) was annotated as premod_a and premod_n, (5) as postmod_p and postmod_ic, (6) as postmod_cc, and (7) as posmod_rc.

(4) As individuals we are constantly surrounded by racist and discriminative media **language** (ICLE-US-MICH-0004.1).

(5) This sudden burst of useful compounds not only improved the chances of a patient's survival in a hospital but also caused a great **need** for medical chemists to study and classify each new drug as it was discovered (ICLE-US-MICH-0015.1)

(6) We Chinese have a **saying** that men at their birth are naturally good (EFCAMDAT-writing-id-556256).

(7) An **invention** of the 20th century which I feel has significantly changed people's lives is the introduction of Bank-cash machines or Automatic teller machines (ICLE-US-MICH-0044.1).

The variable FORM is presented in Table 3.12.

Table 3.12. *The* FORM *Variable and its Levels*

Type of variables	Variable Name	Levels
Morphological	FORM	DA IA ZA

The variable FORM is the choice of the article made on each case. DA represents for definite article (i.e., the), IA represents for indefinite article (i.e., "a" and "an"), and ZA represents for zero article. For example, the NP *illustration* in (8) was annotated as IA, *work* as DA, and *computer* as ZA.

(8) A vivid **illustration** of this can be found by examining the **work**. Recently, an auto-pasts [*sic*] company put all of their inventory on **computer** (ICLE-US-MICH-0002.1).

The variable NUMBER is presented in Table 3.13.

Table 3.13. *The* NUMBER *Variable and its Levels*

Type of variables	Variable Name	Levels
Syntactic	NUMBER	singular plural

This variable NUMBER was annotated automatically based on the POS-tagging. NN and NNS were annotated as singular, and NNP and NNPS as plural. For example, in (9), the NP *saying* was annotated as singular, and *generations* as plural.

(9) Money is the root of all evil is an ancient **saying**-- but its truth applies to all **generations** (ICLE-US-IND-0015.1).

The variable CASE is presented in Table 3.14.

Table 3.14. *The* CASE *Variable and its Levels*

Type of variables	Variable Name	Levels
Syntactic	CASE	Accusative with preposition (acc_p) Accusative with verb (acc_v) Nominative (nom) neither

This variable CASE was annotated automatically for EFCAMDAT based on the syntactic structure it provides in the form of dependency relations (see Geertzen et al., 2013). For LOCNESS, it was annotated manually. For example, in (10), the NP *AIDS* was annotated as nom, *impact* as acc_v, and people as acc_p. In (11), the NP *money* was annotated as neither.

(10) **AIDS** has definately [*sic*] had an **impact** on **people** in the United States (ICLE-US-MICH-0013.1).

(11) The key is in the definition of term **money** (ICLE-US-IND-0015.1).

The variable L1 is presented in Table 3.15.

Table 3.15. *The L1 Variable and its Levels*

Type of variables	Variable Name	Levels
Data	L1	English Japanese Chinese

As has already been presented as descriptive statistics in Section 3.1.2, 833 occurrences of articles from LOCNESS were annotated as English, 833 from EFCAMDAT as Japanese, and the remaining 795 from EFCAMDAT as Chinese.

The variable PROFICIENCY is presented in Table 3.16.

Table 3.16. *The PROFICIENCY Variable and its Levels*

Type of variables	Variable Name	levels
Data	PROFICIENCY	B2 C1

The variable PROFICIENCY was only applicable to NNS data from EFCAMDAT. Based on the conversion chart between the proficiency level measures in EFCAMDAT and that of CEFR (Huang et al., 2017, p. 3), level 11 was converted into B2, and levels 13, 14, and 15 into C1.

Each of the 2,461 occurrences of articles was annotated based on the variables introduced thus far, and the annotated data were analyzed through a series of statistical evaluation, which I now turn to.

3.2 Statistical Evaluation

The statistical evaluation involves a two-step procedure that I will describe below. After an overview of the MuPDAR protocol (Section 3.2.1), I will discuss one possible way of validating MuPDAR findings, which, in my knowledge, remains unprecedented at this point.

3.2.1 MuPDAR

As has been briefly mentioned earlier in this paper, MuPDAR (Gries & Deshors, 2014) is a regression-based methodological protocol that enables the quantification of NNS' non-nativelikeness of the use of a certain linguistic structure. Conceptually, it predicts *what an NS would do in a given linguistic context that an NNS is in,* and this *given linguistic context* is

operationalized through a set of relevant linguistic features. Methodologically, MuPDAR consists of roughly four steps:

(1) train a logistic regression model (R_1) based on NS data,

(2) if the fit of R_1 is good, apply R_1 to NNS data to make predictions and obtain the probability distribution of the target linguistic form (i.e., what an NS *would do* at what probability in a given situation that an NNS is in),

(3) calculate the NNS' deviation based on the difference between the prediction made in (2) and the actual NNS data (i.e., what an NNS *actually did*), and

(4) create a regression model (R_2) to predict the deviation of NNS calculated in (3).

Gries and Deshors (2014), in their analysis of NNS usage of modals *may* and *can*, explain that (3) can be done in two different ways. The first approach is to calculate the deviation *categorically*; that is to say, whenever the predicted NS choice and actual NNS choice do not match, that case is marked as "for" (as in *foreign*), whereas it is marked as "nat" (as in *native*) when they match. The second approach is to calculate the deviation *quantitatively*. In this approach, a vector Dev (as in *deviation*) is created, and a numeric value is attached to each case of the target linguistic form. Whenever the actual NNS choice and the predicted NS choice match, the numeric value is set to 0 (no deviation). Whenever the choices do not match, the numeric value is set to $p - 0.5$, where p stands for the predicted probability of NS choice made by R_1 (for the complete explanation of the original MuPDAR approach, see Gries & Deshors, 2014). The second approach, or the quantitative one, is more commonly used because of the level of granularity it allows for (e.g., Lester, 2019).

When applying MuPDAR to a multinomial classification, a crucial difference between binomial and multinomial classification has to be noted. In binomial classification, one deviation

vector suffices because the probability of one class automatically determines the probability of the other class. For example, when the probability of *may* is 40%, then the probability of *can* is automatically 60%. However, this does not hold true in a multinomial classification like the one in the present study, because the probability of one class does not determine the probability of each of the remaining classes. For example, when the probability of DA is 40%, it only tells us that the sum of the probabilities of IA and ZA is 60%, but it does not tell us what the probabilities of IA and ZA are, respectively. Therefore, a modification has to be made to accommodate the number of classes of the response variable. Two possible alternative approaches and their pros and cons were considered.

3.2.1.1 Approach 1

The first approach is similar to the categorical approach to MuPDAR explained above. It consists of four (almost) identical steps:

(1) train a multinomial logistic regression model (R_1) based on NS data

(2) if the fit of R_1 is good, apply R_1 to NNS data to make predictions and obtain the probability distribution of the article choice (i.e., what an NS *would do* at what probability in a given situation that an NNS is in)

(3) create a vector that categorically represents whether or not the NNS' actual choice matches the NS prediction made in (2), and

(4) create a binary logistic regression model (R_2) to predict the deviation of NNS calculated in (3).

The biggest advantage of this approach is that the final step (4) can be taken in the exact same way as the original MuPDAR because of the categorical nature of the deviation vector.

On the other hand, this approach has (at least) two shortcomings: it cannot quantify the learner's deviation, and it does not distinguish different *kinds* of deviation. The former is inherent to the categorical approach, whereas the latter stems from the nature of the multinomial classification. That is to say, a dichotomous categorization of deviation (i.e., match vs. mismatch) will not tell us what an NNS chose and what an NS would choose in the same situation, because it only tells us if the responses were the same or not. For example, an NNS choosing zero article when an NS chooses definite article and an NNS choosing definite article when an NS chooses indefinite article are two very different scenarios (with probably two very different reasons for the deviation), but they are both recorded as "mismatch" in this approach.

3.2.1.2 Approach 2

Approach 2 addresses the first shortcoming; namely, the inability to quantify the deviation. In this Approach 2, steps (1) and (2) follow Approach 1, and steps (3) and (4) are different:

(1) train a multinomial logistic regression model (R_1) based on NS data
(2) if the fit of R_1 is good, apply R_1 to NNS data to make predictions and obtain the probability distribution of the article choice (i.e., what an NS *would do* at what probability in a given situation that an NNS is in)
(3) create a vector that numerically represents how much the NNS' actual choice matches the NS prediction made in (2), and
(4) create a multiple regression model (R_2) to predict the deviation score calculated in (3).

In step (3), instead of calculating the deviation dichotomously, a numeric value will be assigned to each instance of article use. If the NS and NNS choices were in alignment, a numeric value of

0 will be assigned (i.e., no deviation). If the NS and NNS choices diverge, then the deviation will be quantified as *how small the probability of NS making the same choice as NNS was*.

For example, in a given linguistic context X, NNS chose indefinite article, whereas NS had the probability distribution (as predicted by R_1) of (DA, IA, ZA) = (0.7, 0.1, 0.2). In this case, the probability of an NS making the same article choice as an NNS (i.e., indefinite article), is 0.1. However, this value is counterintuitive because it is to be interpreted as *the smaller the value is, the larger the deviation is*. To make this value more intuitively interpretable, the deviation will be operationalized as $0.5 - p$. Originally, the deviation was defined as $p - 0.5$ (Gries & Deshors, 2014); however, Lester (2019) flipped the equation to make the numeric value more intuitively interpretable. In this example, the deviation is 0.5 - 0.1 = 0.4. The reasoning behind this operation is that, when an NS and NNS choices do not match, the theoretical minimum of the predicted probability of an NS making the same choice as an NNS is 0 (i.e., maximum deviation), whereas the theoretical maximum is < 0.5 (i.e., minimum deviation) because the choices must have matched if the predicted probability exceeds 0.5 regardless of the probability distribution of the other two articles. Therefore, by subtracting this value from 0.5, the deviation value will fall under the range $0 \leq \text{dev} < 0.5$. Given the capacity to quantify learner deviation, Approach 2 was adopted. In what follows, I introduce a more detailed description of each step of Approach 2 with a particular focus on the software and packages used.

3.2.1.3 Software and Packages

A statistical software *R Studio* was used to run each of the four steps in Approach 2. Summarized below are the specifics of Approach 2 and the packages used in each step.

(1) A multinomial logistic regression model was built using a function *multinom()* in the R package *nnet*, with the choice of determiner as a three-level categorical response variable

and all other variables as categorical predictor variables. Following Lester (2019), cross-validation was conducted to ensure the generalizability of this classification accuracy. A commonly used five-fold cross-validation was employed in this study; in other words, 20% of the NS data were labeled as the test set and the other 80% the training set, and this data splitting took place five times, with a unique 20% assigned to the test set for each round of data splitting. Due to lack of available packages, I implemented the five-fold cross-validation code.

(2) The predictions of R_1 on NNS data were obtained through *predict()* function in the R package *nnet*.

(3) For the calculation of the deviation score, I wrote a code that followed the abovementioned calculation method.

(4) The mixed-effects multiple regression model R_2 was built with *lmer()* function in the R package *lme4*.

3.2.2 Validation of predicted NS judgment in NNS data

The whole idea of MuPDAR and its powerfulness is based on the assumption that R_1 (the regression model trained on NS data) will predict a native-like judgment when applied to NNS data. This is precisely what enables us to compare *what an NNS actually did in a given linguistic context* and *what an NS would do in the exact same linguistic context* (as represented by a vector of variables). This assumption is reasonable as long as R_1 fits the NS data well (as measured in the goodness-of-fit and classification accuracy); however, this assumption has never been, to my knowledge, tested empirically perhaps due to the lack of data that enable the empirical validation of such an assumption. That is to say, this validation necessitates data such that an NS choice and an NNS choice of English articles can be directly compared in the exact same linguistic context,

and one type of data that meet this requirement is an essay written by NNS and corrected by NS. In EFCAMDAT, one of the corpora used in the present study, learner essays are provided with professional feedback on grammatical and lexical errors by language teachers. According to recruiting information by Education First,[5] the teachers are all English native speakers with the minimum of 40 hours of training in TEFL. Therefore, the validation of the assumption is possible by comparing the prediction made by R_1 and the actual error correction (or non-correction) made by an NS.

Of the 1628 tokens of articles in 49 learner essays extracted from EFCAMDAT, 34 tokens had error corrections across 13 essays. However, because EFCAMDAT does not claim that error corrections have been provided to all essays, and only "*a substantial portion* of scripts comes with error corrections" (Huang et al., 2017, p. 7, emphasis added), it is dangerous to assume that the other 35 essays, which had no error correction on article, were reviewed by NSs and judged to be completely error-free. Therefore, the only way to distinguish articles that were judged to be correct by an NS from the ones that were simply not reviewed by an NS is to restrict the scope of this analysis to the essays that have at least one error correction. It is reasonable to assume that all the articles with no error correction in an essay that has at least one error correction elsewhere within it were judged by an NS to be correct. Consequently, 13 essays with at least one error correction were deemed appropriate for the assumption validation. These 13 essays included 34 error-corrected tokens and 428 error-free tokens of article, the following two versions of data were created out of these 462 tokens of article:

1. Original NNS production of 462 tokens (428 error-free tokens + 34 tokens before error-correction)

[5] http://www.englishtown.com/teachonline/

2. Error-corrected NNS production of 462 tokens (428 error-free tokens + 34 tokens after error-correction)

If the assumption of MuPDAR is correct, then the predicted NS choice (by R_1) should align closer to the error-corrected NNS production than to the original NNS production.

Following this procedure, R_1 was applied to both versions 1 and 2 of the 462 tokens, and the classification accuracy of the version 2 was higher (73%) than the version 1 (70%), although not statistically significant ($z = 0.87$, $p = .38$). Given that the model R_1 predicts Japanese data and Chinese data at a good accuracy of 70% and 71%, respectively, this is not a cogent piece of evidence to validate the assumption of R_1's native-like judgment on NNS data. This will be further discussed in the limitation section.

4 RESULTS

4.1 MuPDAR

4.1.1 Regression on NS data (R_1)

Overall, the model fit of the multinomial logistic regression model R_1 was excellent[6] (C= .96, well beyond the threshold level C = .8 as proposed by Gries & Deshors, 2014; as well as C = .93 in Lester, 2019; $R^2_{McFadden}$ = .68). The overall classification accuracy, which is the number of correct predictions divided by the total number of predictions, was 89%. This means that the model trained on NS data was able to predict with the accuracy of 89% which one of the three article options (i.e., ZA, IA, DA) a native speaker would use in a given linguistic context represented by a vector of variables. Its generalizability was also good; the result of five-fold cross-validation showed that the mean accuracy score was 84% (*SD* = 2%). Given the little decrease in the accuracy score, this model is reasonably generalizable to different datasets.

Initially, learner ID was to be included as a random effect in R_1; however, because no R packages allow the inclusion of random effects in a multinomial logistic regression model as far as I know, it was first included as a fixed effect. The classification accuracy was higher with learner ID (90%), but the generalizability decreased substantially, as we can see in the mean accuracy score of five-fold cross-validation (77%). Therefore, for the first regression model R_1, a decision was made to not include the variable learner ID. Also, at this stage of MuPDAR, AIC was not considered because the purpose of R_1 is not to construct a parsimonious model; rather, it solely aims at making the most accurate prediction on NNS data that approximates what an NS would do.

[6] C statistics was defined as the area under the receiver operating characteristic (ROC) curve.

Because the fit of R_1 was shown to be good, the NS model R_1 was applied to NNS data. As has already been explained in the methodology section, NS prediction based on R_1 was made on each of the cases in NNS data, and Tables 4.1 and 4.2 show the confusion matrices for the actual article choice by Chinese and Japanese learners of English and the NS prediction by R_1, respectively.

Table 4.1. *Confusion Matrix of R_1 prediction on NNS data (L1 = Chinese)*

		NS choice predicted by R_1			
		DA	IA	ZA	**Total**
Actual NNS choice (L1 = Chinese)	DA	**153**	32	36	221
	IA	23	**64**	7	94
	ZA	76	54	**350**	480
	Total	252	150	393	795

Table 4.2. *Confusion Matrix of R_1 prediction on NNS data (L1 = Japanese)*

		NS choice predicted by R_1			
		DA	IA	ZA	**Total**
Actual NNS choice (L1 = Japanese)	DA	**213**	29	48	290
	IA	12	**56**	4	72
	ZA	91	66	**314**	471
	Total	316	151	366	833

In Tables 4.1 and 4.2, each of the rows correspond with the actual NNS choices, whereas each of the columns correspond with the predicted NS choices. The boldfaced, underlined figures indicate the match between the two; namely, the number of occurrences of articles, in which the actual NNS choice and the predicted NS choice were the same. Overall, the proportion of Chinese speakers' article choice that matched with the predicted NS choice (71%) was not significantly higher than the proportion of Japanese speakers' article choice that matched with the predicted NS choice (70%; $z = 0.59$, $p = .56$).

4.1.2 Regression on dev score (R_2)

Approach 2 was adopted first in the calculation of deviation score and the construction of the final regression model (R_2) because it allows for a more fine-grained quantitative analysis of the deviation. Following the procedure described in the methodology section, the deviation score ($0 \leq dev < 0.5$) was calculated for each of the cases in NNS data. A generalized linear mixed-effect model was built with a function *glmer()* in the R package *lme4*. All the independent variables included in R_1 were entered as predictors, and the variable FORM was also included in R_2, as we would like to see how three types of article uniformly or differentially affect the learner deviation, and how they interact with other predictors. Also, all of them were allowed to interact with the variable L1, as their effects on the deviation are expected to differ based on the learners' first language. Consequently, the model included main effects, two way interactions with FORM (FORM : everything), two way interactions with L1 (L1 : everything), and three-way interactions with both FORM and L1 (FORM : L1 : everything). To avoid overparameterization, AIC and BIC scores were calculated for the models with (1) only main effect, (2) main effect and two-way FORM interaction, (3) main effect and two-way L1 interaction, and (4) main effect, two-way FORM interaction, two-way L1 interaction, and three-way FORM : L1 interaction. This model selection process is summarized in Table 4.3.

Table 4.3. *Model Selection of R_2*

Model	AIC	BIC	$\mathbf{R^2_C}$
1	-1259.08	-1005.51	.16
2	**-3665.61**	**-3055.96**	.81
3	-1243.44	-801.04	.21
4	-3630.61	-2535.404	**.83**

Note. $\mathbf{R^2_C}$ = Conditional R_GLMM^2, which measures the variance explained by both fixed and random effects.

In Table 4.3, AIC and BIC represent numerical measures of the model fit and model parsimony, which penalize the inclusion of additional terms. AIC is more useful for detecting

type II error (false negatives), whereas BIC is more sensitive to type I error (false positives). Based on the table, Model 2 has the smallest AIC and BIC, and Model 4 has a slightly higher R^2_C. The contrast between the models with smaller AIC and BIC (Model 2 and Model 4) and the ones with larger AIC and BIC (Model 1 and Model 3) is most likely due to the inclusion of interaction terms with the variable FORM. The slightly higher R^2_C of Model 4 is not surprising, given that Model 4 is Model 2 + three-way interaction (FORM : L1 : everything). The AIC and BIC for Model 4 are lower than those of Model 2; however, given the conceptual importance of investigating how different L1 speakers receive different influences of other variables, Model 4 was selected as the initial model of R_2. The model was highly significant ($F(192, 1435) = 35.46$, $p < .001$) without any sign of multicollinearity (all VIFs < 2).[7]

Because the inclusion of all of these categorical variables involves the generation of dummy variables, a reference level had to be set. A reference level is interpreted as the baseline level, to which all other levels will be compared to. The reference levels of the 12 variables are presented in Table 4.4.

[7] VIF was tested on a model with only main effects, as Friedrich (1982) posited: "Though a multiplicative term and its constituent variables are often highly correlated, this multicollinearity does not pose problems for the interpretation of the regression results" (p. 803).

Table 4.4. *Reference Level for Each of the Categorical Independent Variables*

	Number of Levels	Reference Level	Rationale	
			less marked	most frequent
FORM	3	ZA	✓	✓
L1	2	Chinese		
NOUNCOUNT	2	singular	✓	✓
NOUNANIM	10	other		✓
NOUNTYPE	2	common	✓	✓
VERBTYPE	5	n/a		✓
PREMOD_A	2	no	✓	✓
PREMOD_N	2	no	✓	✓
POSTMOD_P	2	no	✓	✓
POSTMOD_RC	2	no	✓	✓
POSTMOD_IC	2	no	✓	✓
POSTMOD_CC	2	no	✓	✓
NUMBER	2	singular	✓	✓
CASE	5	nom	✓	
DEFINITENESS	8	misc		✓

In table 4.4, for the variables that are linguistically more marked, the least marked level was set as the reference level (e.g., no prenominal adjectival modification is less marked than modified ones). For the variables for which linguistic markedness was difficult to define, the most frequent level was set as the reference level, following Gries and Deshors (2014). In addition to these fixed-effect independent variables, the variable ID was also included in this model as a random effect.

Significant predictors of the regression model are summarized in Table 4.5.

Table 4.5. *Significant Predictors of Deviation Score*

	df	***F***	***p***
L1 : NOUNANIM	(9, 1628)	2.32	.013
FORM : NOUNCOUNT	(2, 1628)	383.26	< .001
FORM : NOUNANIMACY	(18, 1628)	5.60	< .001
FORM : VERBTYPE	(8, 1628)	6.20	< .001
FORM : PREMOD_N	(2, 1628)	3.65	.026
FORM : POSTMOD_P	(2, 1628)	33.91	< .001
FORM : POSTMOD_CC	(2, 1628)	5.88	.003
FORM : NUMBER	(2, 1628)	410.10	< .001
FORM : CASE	(6, 1628)	5.84	< .001
L1 : FORM : NOUNTYPE	(2, 1628)	3.54	.029
L1 : FORM : POSTMOD_IC	(2, 1628)	4.46	.012
L1 : FORM : DEFINITENESS	(13, 1628)	3.22	< .001

In Table 4.5, only the statistically significant effects with no significant higher interaction effects were included. For example, even though it was significant, the main effect for FORM was not included in the table because its main effect was overridden by the significant interaction term FORM : NOUNTYPE. This interaction was not included either because it was overridden by the highest order interaction term L1 : FORM : NOUNTYPE, which was highly significant. In this sense, it is this highest order interaction (three-way interaction) that is particularly noteworthy in Table 4.5. That is to say, the three highest order interactions at the bottom of Table 4.5 indicate that each of these three independent variables (i.e., NOUNTYPE, POSTMOD_IC, and DEFINITENESS) differentially affected the accuracy of article use, and that such differential effects further varied for Chinese and Japanese learners of English. This will be presented graphically later in this section.

Each of the F-statistics in Table 4.5 can be interpreted as the amount of change in the model fit when the full model is compared against another model without one variable of interest. For example, the row L1 : NOUNCOUNT represents how much improvement the inclusion of the interaction term L1: NOUNCOUNT would make in terms of the model fit, when all other terms are already included in the model. Considering that all the independent variables

are categorical variables, this is in principle identical to factorial ANOVA with type-III sum of squares. In the following sections, the significant predictors will be further investigated graphically and statistically.

For each graph, the dots (or other shapes, such as triangles and squares) represent the means of the deviation score at a given level of the variable of interest. Because all other variables not on the graph are held constant, the differences between such means represent the marginal effects of the variables of interest. Error bars represent 95% confidence intervals. The value on *y*-axes is always the deviation score, whereas the value on *x*-axes and legends are the predictor variables that construct the interaction term of interest (as no main effects are analyzed here, every graph will have at least two predictor variables).

For the statistical analyses, because everything to everything comparison will lead to an extremely conservative α-level adjustment, every level of the predictor variable was only compared against the reference level, which has already been discussed earlier. This means that for each predictor variable with *k* levels, (k – 1) tests were conducted. Therefore, the α-level was adjusted accordingly based on Bonferroni correction, a simple and conservative type of adjustment. More specifically, a conventional threshold level ($\alpha = .05$) was divided by the number of tests (k -1). This calculation is a simplification of a more complex formula, but it can be used satisfactorily in most cases (Walters, 2016). For two-way interaction terms, comparisons were conducted based on the second predictor, within a given level of the first predictor. For example, when analyzing FORM : NOUNCOUNT interactions in Figure 4.1, comparisons were conducted based on the second predictor (i.e., NOUNCOUNT; countable vs. uncountable) within a given level of the first predictor (i.e., FORM); which is, say, DA. Because the second predictor

only has two levels and only one comparison will take place, the critical α-level will not be adjusted for this particular example.

4.1.2.1 Two-Way Interactions

In this section, marginal effects of each of the significant two-way interactions will be presented graphically, in the same order as presented in Table 4.5. Figure 4.1 shows the marginal effect of the interaction term L1 : NOUNANIMACY.

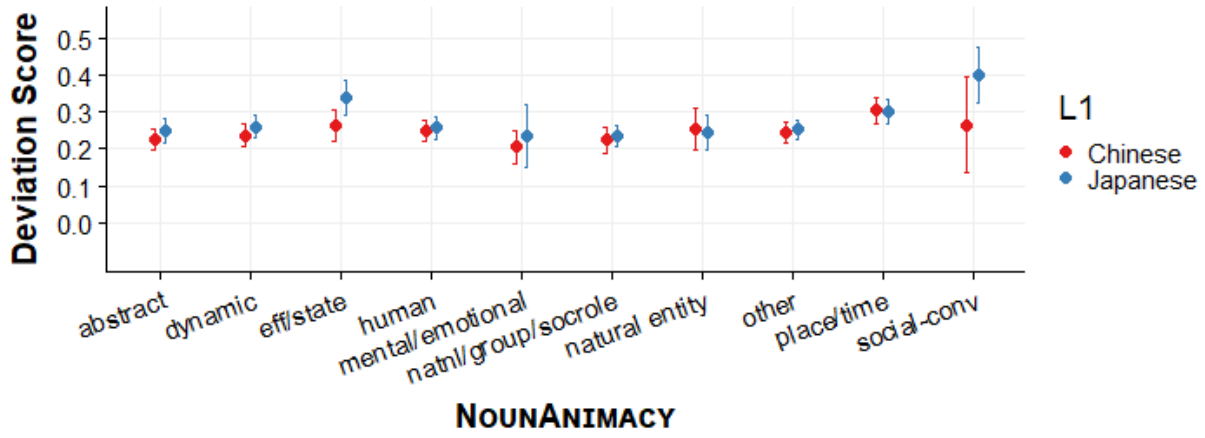


Figure 4.1. Marginal Effect of the Interaction Term L1 : NOUNANIMACY

Figure 4.1 shows that, for most of the noun animacy types, Chinese and Japanese learners of English seem to have little difference in the deviation score. However, it is noteworthy that Japanese learners of English seemed to have larger deviation scores with noun animacy types eff/state and social-conv.

The marginal effect of the interaction term FORM : NOUNCOUNT is shown in Figure 4.2.

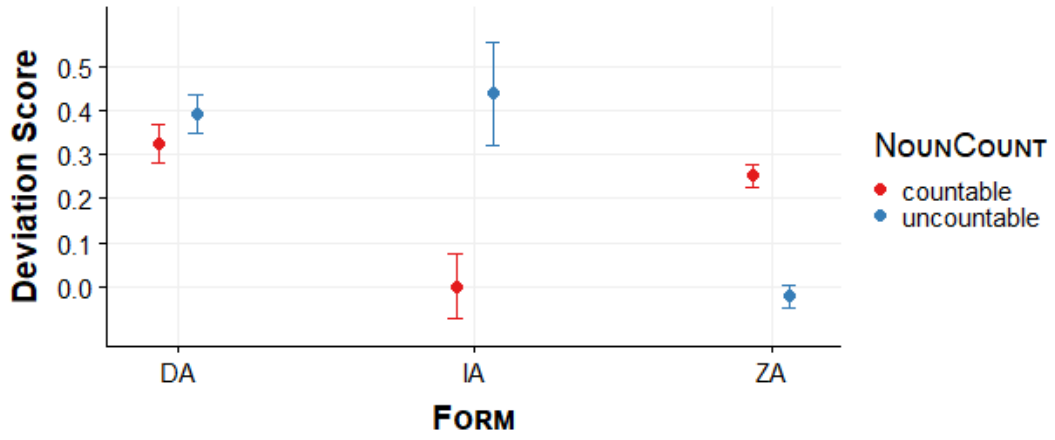


Figure 4.2. Marginal Effect of the Interaction Term FORM : NOUNCOUNT

Figure 4.2 shows that, regardless of the noun countability, both Chinese and Japanese learners had problems using DA accurately, and this inaccuracy of DA was more obvious with uncountable nouns. This difference becomes much more pronounced for the use of IA; learners had great difficulty using IA with uncountable nouns, whereas their use of IA with countable nouns was almost native-like. Most strikingly, the relative ease of countable nouns does not hold true for ZA; that is to say, the use of ZA with uncountable nouns was more nativelike, whereas the use of ZA with countable nouns was more deviant.

Figure 4.3 shows the marginal effect of the interaction term FORM : NOUNANIMACY.

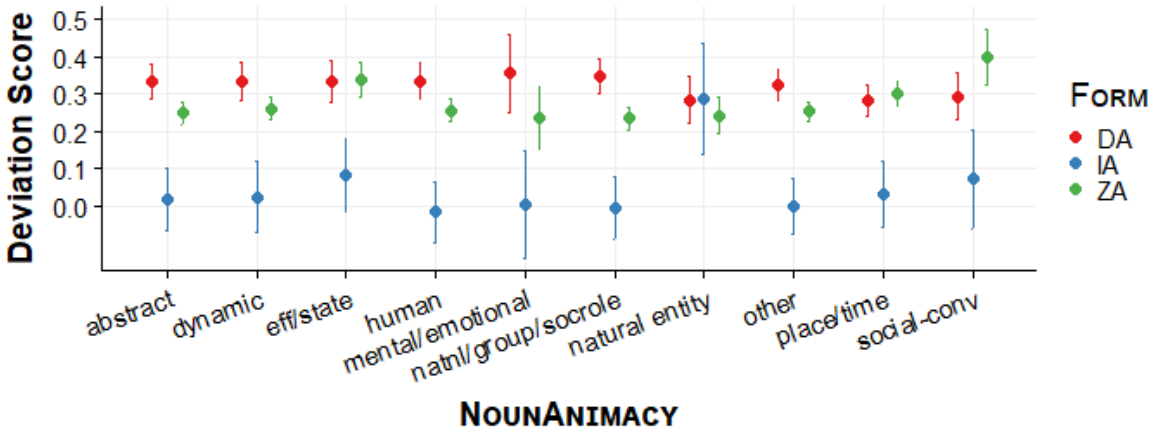


Figure 4.3. Marginal Effect of the Interaction Term FORM : NOUNANIMACY

Figure 4.3 shows that learners tended to have higher nativelikeness in the use of IA, less nativelike use of ZA, and more deviant use of DA with most of the noun animacy types. However, for the animacy type natural entity, this general pattern does not hold true. IA, which is used with the highest nativelikeness in all other animacy types, has the highest deviation score with the animacy type natural entity. However, the wide error bar indicates that we need more sample size to be more certain about this observation. Indeed, of 1628 cases of article use, the use of IA with natural entity noun was very few (n = 2). Another observation made in the figure was that the nativelikeness of the use of ZA on animacy types eff/state, place/time, and social-conv was lower compared to other animacy types.

Figure 4.4 shows the marginal effect of the interaction term FORM : VERBTYPE.

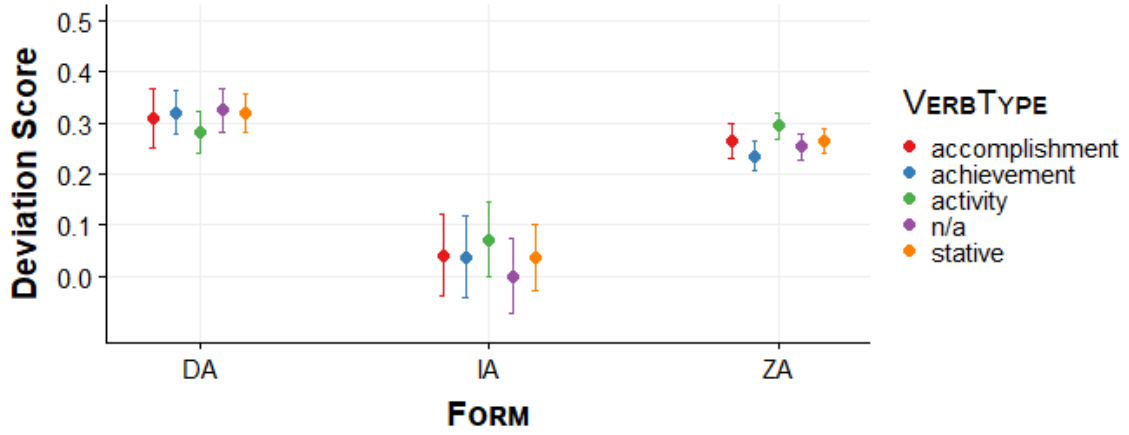


Figure 4.4. Marginal Effect of the Interaction Term FORM : VERBTYPE

Figure 4.4 shows that the verb type activity affects the nativelikeness of the use of DA, IA, and ZA differently. Whereas the use of IA and ZA was more deviant with activity type verbs, the use of DA was more nativelike with activity type verbs, compared to the reference level (i.e., no verb).

The interaction terms FORM : POSTMOD_P, FORM : PREMOD_N, and FORM : POSTMOD_CC showed similar interaction patterns, and are therefore presented altogether in Figure 4.5.

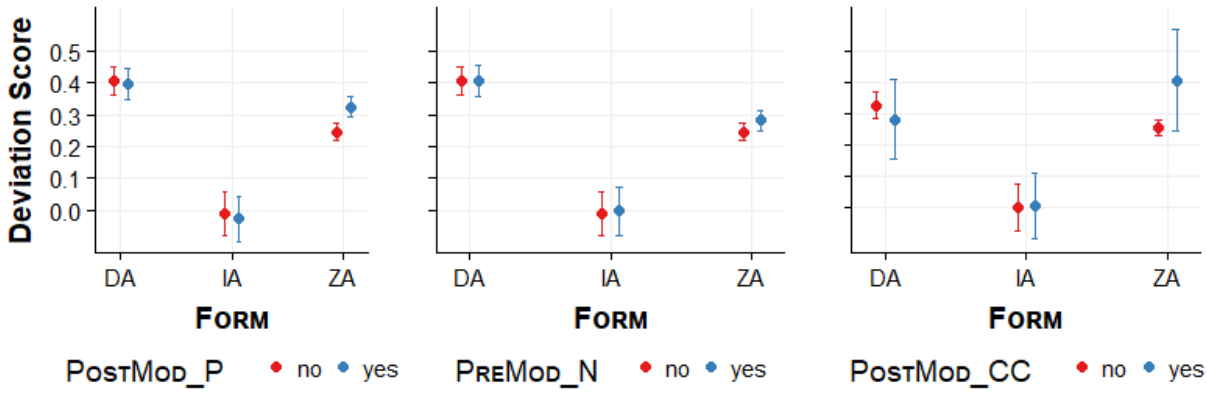


Figure 4.5. Marginal Effect of the Interaction Term FORM : MODIFICATION

Figure 4.5 shows that regardless of the presence or absence of the modification (whether it is post-modification with a prepositional phrase or complement clause, or a pre-modification with a noun), the use of DA tended to be deviant. On the other hand, the use of IA was more nativelike regardless of the modification. As opposed to DA and IA, whose deviation scores seemed to receive little influence by noun modification, ZA seemed to be used inaccurately overall, and even more so with the presence of modification. This effect of modification on the higher level of deviation of ZA use was more pronounced with the presence of the post-modification with prepositional phrase and with complement clause. However, the wider error bars that correspond to the presence of post-modification with complement clause are due to the small sample size and should be interpreted with caution.

The marginal effect of the interaction term FORM : NUMBER is presented in Figure 4.6.

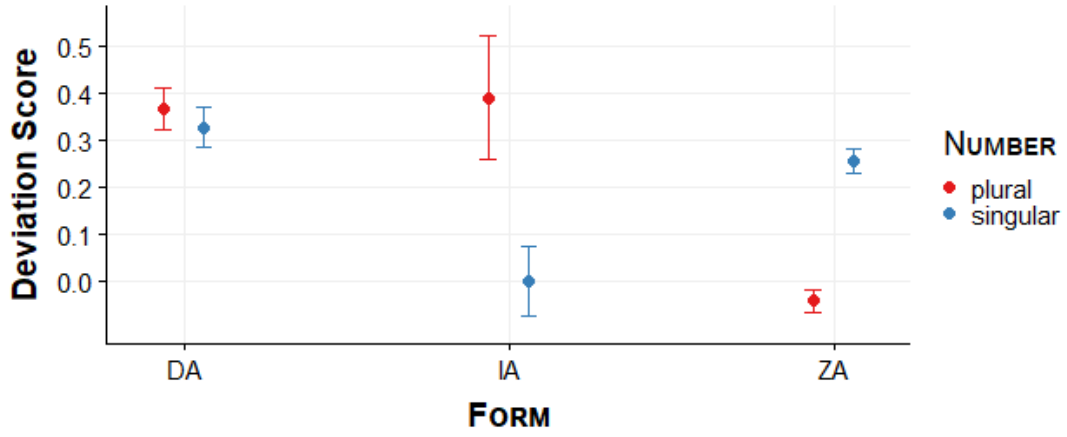


Figure 4.6. Marginal Effect of the Interaction Term FORM : NUMBER

Figure 4.6 shows that, on the one hand, the use of DA is deviant for both plural and singular nouns, with plural nouns associated with a higher deviation score. On the other hand, the nativelikeness of the use of IA was strongly affected by the variable NUMBER. Whereas the use of IA was far off with plural nouns, it was much more nativelike with singular nouns. The

opposite was true about ZA; the use of ZA was more nativelike with plural nouns than with singular nouns.

The marginal effect of the interaction term FORM : CASE is presented in Figure 4.7.

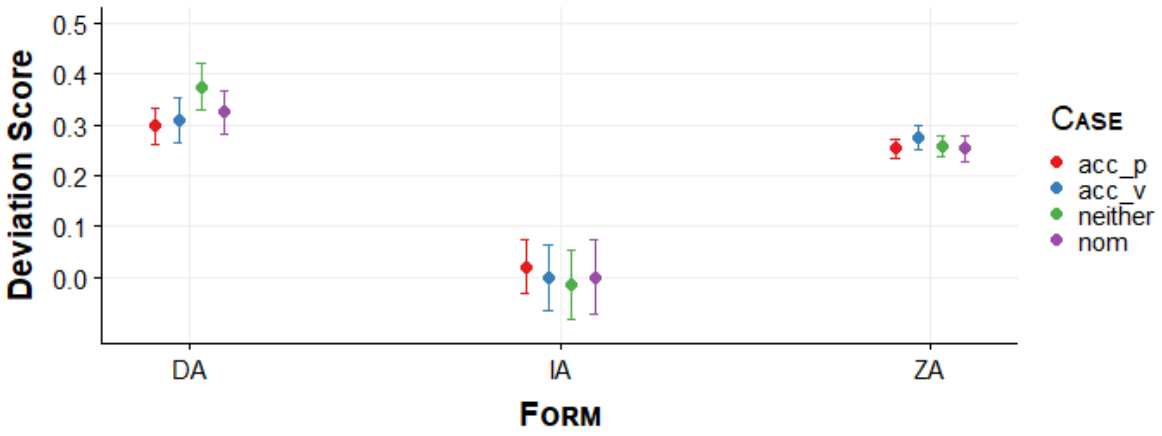


Figure 4.7. Marginal Effect of the Interaction Term FORM : CASE

Figure 4.7 shows that, whereas the deviation scores for the use of IA and ZA seem to be similar across different syntactic Cases, this was not the case for the use of DA. Specifically, the deviation score was higher when the noun had neither accusative nor nominative CASE, while the deviation scores were similar for other types of CASE.

4.1.2.2 Three-Way Interactions

In this section, marginal effects of each of the three-way interaction terms will be presented graphically, in the same order as in Table 4.5. The marginal effect of the interaction term L1 : FORM : NOUNTYPE is presented in Figure 4.8.

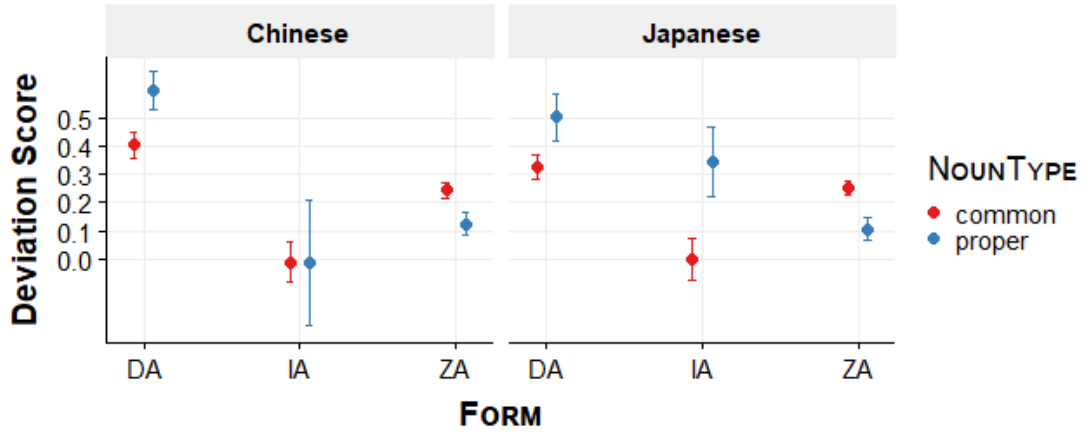


Figure 4.8. Marginal Effect of the Interaction Term L1 : FORM : NOUNTYPE

Figure 4.8 shows that, the use of DA is more nativelike with common nouns than with proper nouns, and the opposite is true about the use of ZA; its use is more nativelike with proper nouns than with common nouns. This pattern holds true for both Chinese and Japanese learners; however, the patterns diverged for the use of IA. For Chinese speakers, the use of IA is equally nativelike with almost no deviation for both common and proper nouns, but for Japanese speakers, the use of IA was nativelike only for common nouns. It was largely deviant for proper nouns.

The marginal effect of the interaction term L1 : FORM : POSTMOD_IC is presented in Figure 4.9.

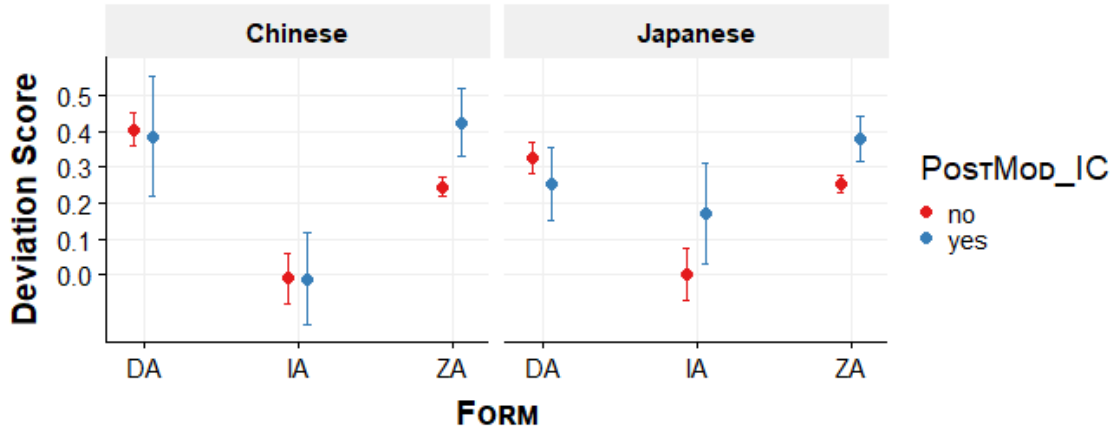


Figure 4.9. Marginal Effect of the Interaction Term L1 : FORM : POSTMOD_IC

Figure 4.9 shows that for the use of ZA, both Chinese and Japanese speakers were more nativelike with the absence of no post-modification with infinitival clause. For the use of DA and IA, different patterns were observed. For Chinese speakers, the presence or absence of the post-modification with infinitival clause did not affect the nativelikeness of the use of DA and IA, although DA was used with more deviation overall. However, for Japanese speakers, the use of DA was more nativelike with the absence of the modification than without it, while the opposite was true for the use of IA; it was more nativelike when there was no post-modification with infinitival clause.

The marginal effect of the interaction term L1 : FORM : DEFINITENESS is presented in Figure 4.10.

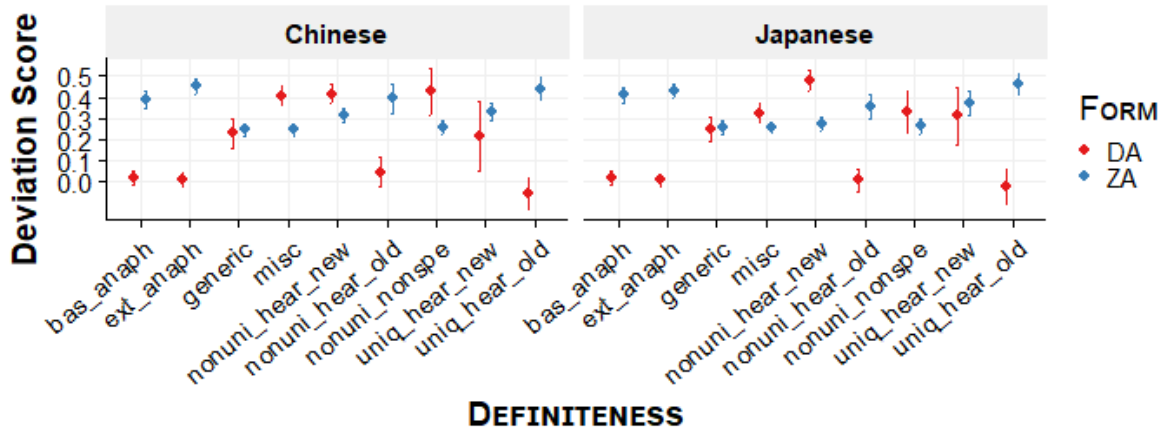


Figure 4.10. Marginal Effect of the Interaction Term L1 : FORM : DEFINITENESS

In Figure 4.10, one level (i.e., IA) of the variable FORM was omitted because the distribution of the article choice across all levels of DEFINITENESS was not completely crossed; in particular, IA was not used at all for the levels uniq_hear_new and uniq_hear_old. Figure 4.8 shows that Japanese and Chinese learners of English follow a similar deviation pattern across different kinds of definiteness, with the following three exceptions. First, for the level misc (miscellaneous), even though both Chinese and Japanese learners have more trouble using DA than ZA, this difference is larger for Chinese than for Japanese. Secondly, for the level nonuni_hear_new (non-unique hearer new), similarly, both Chinese and Japanese speakers have higher deviation scores for DA than for ZA. However, the difference was larger for Japanese learners, and their use of DA to express the definiteness of nonuni_hear_new seems to be more deviant. Lastly, for the level nonuni_nonspe (non-unique nonspecific), again, the use of ZA is more nativelike than DA for both Chinese and Japanese learners. This difference was, however, larger for Chinese speakers and the use of DA for the definiteness of nonuni_nonspe seems to pose a particular challenge for them.

5 DISCUSSION

5.1 Implications

This study set out to explore the effect of various contextual factors on the article use by Chinese and Japanese learners of English, and how the first language background differentially influences such effects. Most notably, three independent variables (i.e., NOUNTYPE, POSTMOD_IC, and DEFINITENESS) interacted with both L1 and FORM, constituting significant three-way interactions. This level of analytical granularity is unique to the multifactorial approach this study has adopted, and it is difficult to grasp without a proper comparison with monofactorial approach. Hence, in what follows, I first present what MuPDAR approach was able to capture, that a monofactorial approach would not have been able to capture. Let us take the interaction term Form: L1 as an example. A simple comparison of group means is shown in Figure 5.1.

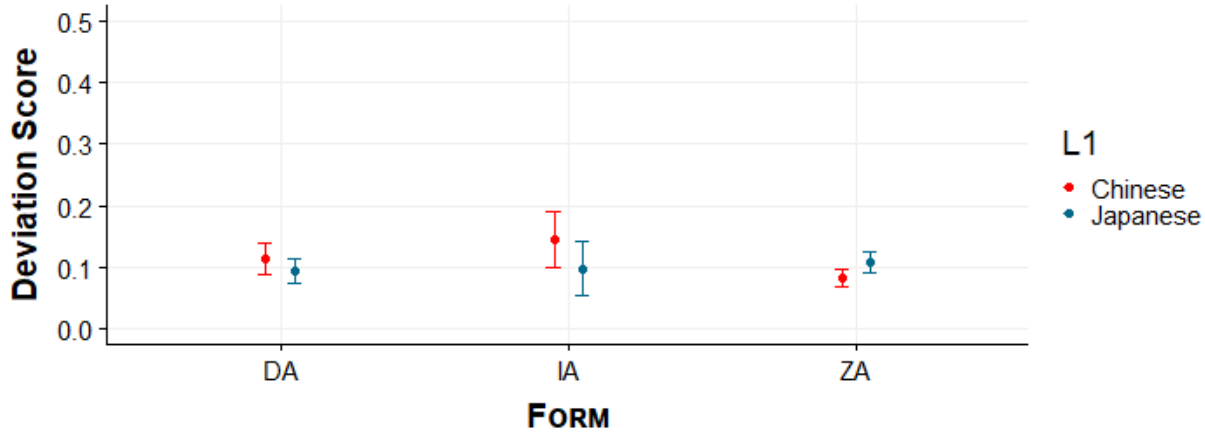


Figure 5.1. Mean Comparison across L1 and FORM

As shown in Figure 5.1, a simple group comparison shows that the deviation score for each of the three article types seems to be somewhat similar to each other. Within each article type, Japanese speakers seem to outperform Chinese speakers for DA and IA, whereas the opposite is

true about ZA. However, when all other independent variables are taken into account, we obtain a very different picture, as shown in Figure 5.2.

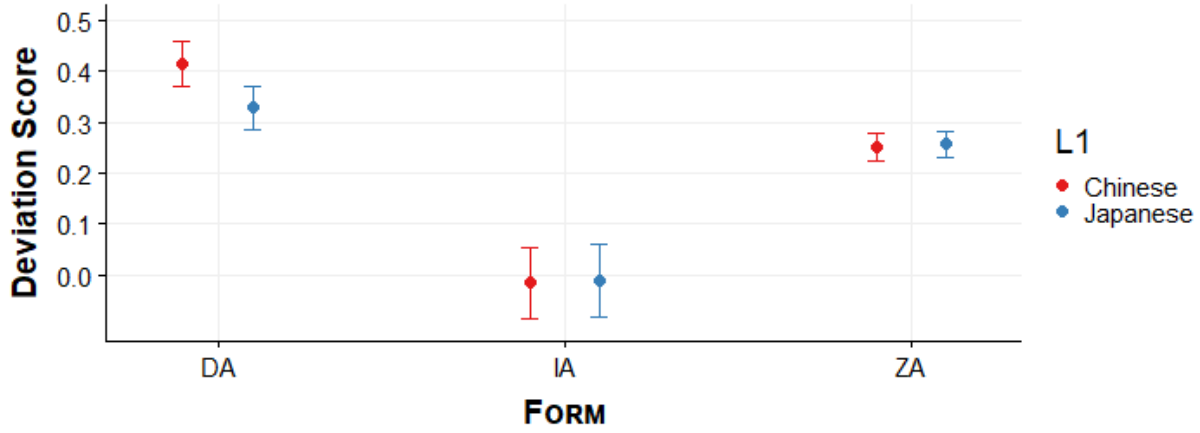


Figure 5.2. Marginal Effect of the Interaction Term L1 : FORM

Figure 5.2 is identical to Figure 5.1, except that all other independent variables are held constant. In other words, the difference observed in Figure 5.2 is purely (at least to the extent the independent variables considered in this study cover relevant variables affecting the use of English articles) due to the difference of L1 backgrounds and the article types. The amount of difference observed in Figure 5.2 suggests that, a mere comparison of group means in Figure 5.1 masks quite a bit of the actual difference associated with the variable(s) of interest. With this in mind, let us now turn to what this multifactorial approach was able to find.

Overall, the results were congruent with the previous studies in that noun countability and number significantly affected the accuracy of the NNS use of English articles. Specifically, FORM : NUMBER interaction ($F(2, 1628) = 410.10$, $p < .001$) and FORM : NOUNCOUNT interaction ($F(2, 1628) = 383.26$, $p < .001$) had the largest F statistics among all the significant variables. Post-hoc analyses on the FORM : NUMBER interaction showed that the use of IA on uncountable nouns and the use of ZA on countable nouns were deviant from the NS predictions. This particular pattern suggests that, speculatively, these errors could be attributed to the

misconception of noun countability. This is in alignment with the observations made by Liu and Lu (2020), in which they concluded that the misconception of noun countability was the most important factor leading to the misuse of English articles by Chinese speakers.

Post-hoc analyses on the FORM : NOUNCOUNT interaction also showed an expected pattern—the use of IA on plural nouns and the use of ZA on singular nouns were deviant. There were only four cases of IAs used on plural nouns, and this seems to be a rare type of mistake rather than a systematic kind of error. On the other hand, the use of ZA on plural nouns had the lowest deviation score (i.e., the highest nativelikeness). This result is congruent with Lourex and Kendall (2018), in which they reported that the use of articles was most nativelike in plural contexts across the two corpora of Chinese EFL learners.

The graphical analysis of the interaction term Form : NOUNANIMACY provides a deeper insight into the NNS' tendency to misperceive the noun countability. High deviation scores of the use of ZA on noun animacy types eff/state, place/time, and social-conv suggest that NNS might have misclassified nouns within these animacy types as uncountable. This is understandable, given the observation by Doetjes (2017) that "both agents and cohesive objects are normally denoted by count nouns, as in two dogs or three pens" and the analysis by Butler (2002) that Japanese ESL learners with relatively low English proficiency had "fixed notion of noun countability" (p. 458). In other words, NNS in the present study might have automatically labeled eff/state, place/time, and social-conv types as uncountable based on their intangible meanings, without considering the fact that the countability can vary from context to context.

Interestingly, the misconception of the noun countability seems to be susceptible to L1 transfers. The graphical analysis of the interaction term L1 : NOUNANIMACY shows that, while the deviation scores of Chinese and Japanese speakers are mostly similar across different

animacy types, the accuracy of the article use with the eff/state type seems to be higher for Chinese speakers. This might be due to the presence of the richer variety of measure words in Chinese. In Chinese, seemingly intangible nouns within eff/state type can be counted with a specific measure word. For example, *an accident* can be counted as 一場意外 (*yi chang yi wai*; one occurrence of accident/ unexpected situation) in Mandarin Chinese, whereas this is not the case in Japanese.

This presence of richer measure words in Mandarin Chinese may account for the Chinese speakers' slight outperformance in the accuracy of the English article use in this study. Overall, Chinese speakers were more accurate (71%) than Japanese speakers (70%) in their article use, though not significantly different from each other ($z = 0.59$, $p = .56$), as presented in Tables 4.1 and 4.2 in Section 4. This equivocal result, on the one hand, supports the observation by Crosthwaite (2016) that presence of rich demonstratives is one of the factors that contribute to the difference within article-less languages. Chinese and Japanese languages both have similarly rich demonstratives, and the similar performance between the two groups of NNS in this study is congruent with this hypothesis.

However, on the other hand, the Chinese speakers' slight outperformance also seems to replicate the trend observed in Han et al. (2006), in which Chinese speakers produced slightly lower number of article errors compared to Japanese speakers. Given that the rich demonstratives are present in both languages, it may be reasonable to attribute this slight difference in the richer measure words in Mandarin Chinese.

5.2 Limitations

Throughout the post-hoc graphical analyses of each significant effect, it appeared that the use of IA, in general, was associated with lower deviation scores (i.e., more nativelikeness).

However, it is important to note that this does not mean Chinese and Japanese learners of English have a good understanding of the distribution of IA; rather, it seems like the low deviation score of IA is attributed to its overly restricted use by NNS. With a closer look at the confusion matrices of actual NNS article choice and predicted NS article choice presented earlier, it becomes clear that NS (as predicted by R_1) is far more likely to use IA (19%) than Chinese (12%). This pattern was also true for the Japanese data, in which NS prediction of IA (18%) was much more frequent than Japanese (9%). It, then, follows naturally that the seeming nativelikeness of IA use by NNS was due to the high precision of IA use (Chinese: 68%, Japanese: 78%), and the overall infrequency of IA use resulted in a low recall of IA use (Chinese: 43%, Japanese: 37%). This difference in the precision score and recall score of IA use is particularly large for Japanese. In other words, the problem lies in the fact that the marginal effect of FORM (and its interactions with other independent variables) on the deviation score only considers the non-nativelikeness of IA that was *actually used* by NNS, and does not consider the non-nativelikeness due to the non-use of IA in an obligatory context (as predicted by R_1).

One possible fix for this problem is to define FORM as what the NS prediction by R_1 is, instead of what the actual NNS choice is. By doing so, we can observe how the deviation score is differentially affected by other variables in each of the three obligatory contexts for DA, IA, and ZA. However, this is merely an ad-hoc fix for the problem described above, and is not effective when learners overuse a certain target form (for the exact same reason for which TLU was proposed in place for SOC). This problem of defining the variable FORM is unique to a multinomial MuPDAR like the present study.

Another aspect of this study that warrants further investigation is the calculation of the deviation score. Conceptually, it is reasonable to operationalize the deviation score as $p - 0.5$,

where *p* stands for the probability of an NS *not* choosing the article chosen by an NNS. However, as has already been pointed out in Section 3, the deviation score defined in this way does not tell us whether the deviation is due to an overproduction or an underproduction of one level of the target structure. The deviation score for binary linguistic structures in original MuPDAR (Gries & Deshors, 2014) contains more information, because it ranges from -0.5 to 0.5, with the absolute value and ± sign indicating the magnitude of deviation and the direction of the deviation (underproduction vs. overproduction), respectively. Therefore, a way to operationalize the deviation score in a multinomial setting, such that the direction of the deviation is also included in the score, would advance the instrumental convenience of multinomial MuPDAR.

As to the validation of the assumption that R_1 in fact makes a nativelike judgment on NNS data, the results remained inconclusive. This is mainly due to the small number of corrections on the article errors available in the data used in this study. As has already been mentioned in Section 3, the number of occurrences of articles used for this validation was 461, of which 34 were error-corrected and 418 were error-free. Because the validation relies on the difference between R_1's classification accuracy on the 461 occurrences before error-correction and the same 461 occurrences after error-correction, the number of occurrences of articles that *were actually error-corrected* has to be big enough for the two classification accuracies to be different enough to validate the assumption. One way to address this problem is to selectively extract essays with a large number of error corrections on articles from EFCAMDAT.

6 CONCLUSION

The present study is the first to (i) apply MuPDAR to a multinomial target structure and to (ii) take a multifactorial approach to the investigation of the article use by NNS. Conceptually, the results showed relative importance of each of the relevant semantic, syntactic, and morphological factors governing the use of English articles. Methodologically, the first attempt to extend MuPDAR to a multinomial linguistic structure was not without problems, but it would potentially open up MuPDAR to a wider range of linguistic phenomena, as it is no longer restricted to the ones that have binary choices.

APPENDIX

APPENDIX

Table A.1. *List of Topics Extracted from EFCAMDAT*

Level	Unit	Title	Topic
10	1	Extreme activities	Helping a friend find a job
10	2	Gender differences	Doing a survey about discrimination
10	3	The cost of living	Requesting a bank loan
10	4	Health and fitness	Applying to be a fitness trainer
10	5	Lifestyles	Finding a home for a wealthy client
10	6	Telling stories	Describing a terrifying experience
10	**7**	**Presenting information**	**Presenting trends**
10	8	Competition and cooperation	Giving feedback about a colleague
11	**1**	**Talking about films**	**Writing a movie review**
11	2	Fears and phobias	Helping a coworker deal with a phobia
11	*3*	*Technology*	*Writing an advertising blurb*
11	**4**	**Beliefs and convictions**	**Writing up survey findings**
11	**5**	**Career paths**	**Reviewing a self-help book**
11	6	Computers and the Internet	Setting rules for social networking
11	7	Law and order	Dealing with a breach of contract
11	8	Listening skills	Improving your study skills
12	1	Manners and etiquette	Turning down an invitation
12	2	Books and stories	Entering a writing competition
12	3	Mysterious phenomena	Buying a painting for a friend
12	4	Corporate culture	Writing a report on staff satisfaction
12	5	World English	Proofreading an article
12	6	Leadership qualities	Attending a leadership course
12	7	Soft skills	Conducting a performance appraisal
12	8	Awkward situations	Writing an apology note
13	1	Politics	Writing a campaign speech
13	2	Home design	Renting out a room
13	*3*	*Market research*	*Comparing two demographic groups*
13	4	Fair trade	Giving advice about budgeting
13	**5**	**Contributing to society**	**Writing about a disaster relief effort**
13	6	Art and design	Writing a brochure for a museum
13	7	Mother nature	Making an educational product for kids
13	8	Reaching your potential	Reaching your potential
14	*1*	*Advertising*	*Writing advertising copy*
14	**2**	**The environment**	**Choosing a renewable energy source**
14	3	Good and bad news	Writing a rejection letter
14	4	Health and well-being	Attending a seminar on stress reduction
14	5	Taking a risk	Talking a friend out of a risky action
14	6	Education and training	Applying for sponsorship

Table A.1. (cont'd)

14	7	Making a speech	Writing a wedding toast
14	8	Jokes and humor	Delivering a punch line
15	1	In the news	Covering a news story
15	2	Communication	Hosting a group of foreign buyers
15	3	The power of the mind	Writing an article about NLP techniques
15	4	The entertainment industry	Making a movie
15	**5**	**E-commerce**	**Comparing two online retailers**
15	6	Urban issues	Writing an article about a superstore
15	7	Quality of life	Writing about future lifestyles
15	8	Meaning and symbols	Interpreting a prophecy
16	1	Science and technology	Attending a robotics conference
16	2	National identity	Writing about a symbol of your country
16	3	Tough choices	Following a code of ethics
16	4	Fame and fortune	Criticizing a celebrity
16	5	Creative thinking	Using creative writing techniques
16	6	Financial planning	Applying for a home loan
16	7	Dealing with stress	Writing a visualization script
16	8	Doing research	Researching a legendary creature

Note. This is the list of topics that require more than 100 words (i.e., criterion (a) described in page 22 of this paper). From this list, based on the criteria (b) and (c) described on page 22, seven topics (underlined and boldfaced) were adopted. Italicized topics were also included after the exclusion criteria (b) and (c), but not used in this study due to the time constraint.
Note. The topic list is extracted from EFCAMDAT (Huang et al., 2018; Geertzen et al., 2013), and it contains actual writing instructions. It is available from https://corpus.mml.cam.ac.uk/efcamdat2/public_html/task_screenshots/EFwrittenTasks.xml

Table A.2. *The Annotation Scheme of the Variable* NOUNANIMACY *(Deshors, 2016, pp. 110 – 111, 143)*

ID Tag levels	**Examples**	**Conflated Tag Levels**
Animal	*Birds*	non-human
Flora	*Plant*	
Human	*People, guy*	human
Nationals	*Americans, Europeans*	natnl/group/socrole
Group	*Parliament, committees*	
Social roles	*Shop owners, scientists*	
Object/ artifact	*Computers, missiles*	other
Scholarly work	*Essay, chapter*	
Form/ substance	*Drugs, radioactive materials*	
Imaginary beings	*Fictional beings, character*	
Absence	*Nothing*	
Measure	*Majority, doses*	
Abstract	*Cultural differences, problems, power*	abstract
Action	*Reading, prayer*	dynamic
Process	*Changes, progress*	
Dummy 'it'	*It may not sound very patriotic*	ling
(pseudo) cleft structure	*It may be predicted that*	
Effect	*Consequences, results*	eff/state
state	*Existence, knowing*	
Mental/ emotional	*Consciousness, imagination*	mental/emotional
Natural entity	*Crops, eggs*	natural entity
Place/ time	*1993, countries*	place/time
Social Convention	*Constitution, tax rates*	social-conv

Table A.3. *The Annotation Scheme of the Variable DEFINITENESS (Adopted from Bhatia, Lin et al., 2014b)*

ID Tag levels	Examples	Conflated Tag Levels
Unique_Physical_Copresence	***John** here is an investment banker.*	Unique Hearer Old (uniq_hear_old)
Unique_Larger_Situation	*In the days since **Hillary Clinton** unburdened herself in an interview with The Atlantic's Jeffrey Goldberg …*	
Unique_Predicative_Identity	*Clark Kent is **Superman**.*	
Unique_Hearer_New	*A restaurant chain named **Shoney's***	Unique Hearer New (uniq_hear_new)
NonUnique_Physical_Copresence	***The podium** is too high.*	Non-Unique Hearer Old (nonuni_hear_old)
NonUnique_Larger_Situation	***The chair** (at a conference) / **today***	
NonUnique_Predicative_Identity	*He is **the manager**.*	
NonUnique_Hearer_New_Spec	*I am looking for **a nurse**. Her name is Sara.*	Non-Unique Hearer New (nonuni_hear_new)
NonUnique_NonSpec	*I am looking for **a nurse** [any nurse would do].*	Non-Unique Non-Specific (nonuni_nonspe)
Generic_Kind_Level	***Dinosaurs** are extinct.*	Generic (generic)
Generic_Individual_Level	***Cats** have fur.*	
Same_Head	*I'm going to tell you a quick story. It's **a true story**.*	Basic Anaphora (bas_anaph)
Different_Head	*I adopted a cat this weekend. **The animal** is so cute.*	
Bridging_Nominal	*Ilooked at <u>an apartment</u> yesterday. **The kitchen** was really large.*	Extended Anaphora (ext_anaph)
Bridging_Event	*My friend's son <u>got married</u> this weekend. **The bride** looked beautiful.*	
Bridging_Restrictive_Modifier	***the house** <u>next door</u>/ <u>John's</u> **daughter***	
Bridging_Subtype_Instance	*I collect <u>coins.</u> I have **a 1943 steel penny**.*	
Bridging_Other_Context	*I want to focus on <u>what many of you have said you would like me to elaborate on</u>. What can you do about **the climate crisis?***	
Pleonastic	***It is raining.***	Miscellaneous (misc)
Quantified	***All the people / no motorcade***	
Predicative_Equative_Role	*He's **a teacher**. / This is **an opportunity.***	
Part_Of_Noncompositional_MWE	*Ole's Charlie kicked **the bucket** today.*	
Measure_Nonreferential	***Hours** later / **miles** away*	
Other_Nonreferential	***Global warming/ concern/** the topic of **energy***	

Table A.4. *Relevant Excerpts from Bhatia, Simons et al. (2014, p. 912, emphasis added)*[8]

Excerpt	Page
The three main communicative functions in CFD are **Anaphora** vs. **Nonanaphora** (whether the entity is old in the discourse or not), **Hearer-old** vs. **Hearer-new**, and **Unique** vs. **Nonunique** (annotated for **Nonanaphoric** only in the current scheme).	912
Anaphoric NPs include pronouns and nouns that have been mentioned previously. Previously-mentioned nouns do not need to be identical in form to their antecedents, e.g. the child can be an **anaphoric** reference to a girl. NPs whose existence is evoked by previous NPs or events are also treated as **anaphoric** with the subheading of **bridging anaphora** (in analogy with and extending the notion **bridging** introduced by (Clark, 1977). These include mentioning the kitchen after talking about a house or mentioning the victims after using the verb attack. A special case of **bridging** is NPs that contain a modifier that evokes them as in the woman who lives next door, which can be used in a conversation where the woman has not been previously mentioned. Next door is used deictically relative to the speaker, making the referent of the whole noun phrase identifiable.	912
Non-anaphoric NPs are those that have not been mentioned or evoked by something that was mentioned. They can be **specific** (She wants to marry an Irishman. His name is Paul.) or **non-specific** (She wants to marry an Irishman. She should go and find one). Some **nonanaphoric** nominals are known to the addressee because they are physically present or because of the situation that the speaker and hearer are in. For example, you can talk about the hotel or the program chair at a conference even when those things have not been previously mentioned. **Non-anaphoric** NPs also include those with unique, common-knowledge referents such as the Empire State Building, Barack Obama.	912
Aside from **anaphoric** and **non-anaphoric** nominals, other categories are **pleonastic**, **quantified**, **predicative**, **nonreferential**, and **part of non-compositional multi-word expression**.	912

[8] The excerpts were re-organized into bullet points, and quotation marks were not used for readability.

Table A.5. *Relevant Excerpts from Bhatia, Lin et al.* (2014a, p. 1061, emphasis original)[9]

Excerpt	Page
At the highest level, the distinction is made between **Anaphora**, **Nonanaphora**, and **Miscellaneous** functions of an NP (the annotatable unit). **Anaphora** and **Nonanaphora** respectively describe whether an entity is old or new in the discourse; the **Miscellaneous** function is mainly assigned to various kinds of nonreferential NPs. The **Anaphora** category has two subcategories: **Basic_Anaphora** and **Extended_Anaphora**. **Basic_Anaphora** applies to NPs referring to entities that have been mentioned before. **Extended_Anaphora** applies to any NP whose referent has not been mentioned itself, but is evoked by a previously mentioned entity. For example, after mentioning a wedding, the bride, the groom, and the cake are considered to be **Extended_Anaphora**.	1061
Within the **Nonanaphora** category, a first distinction is made between **Unique**, **Nonunique**, and **Generic**. The **Unique** function applies to NPs whose referent becomes unique in a context for any of several reasons. For example, Obama can safely be considered unique in contemporary political discourse in the United States. The function **Nonunique** applies to NPs that start out with multiple possible referents and that may or may not become identifiable in a speech situation. For example, a little riding hood of red velvet in fig. 2 could be annotated with the label **Nonunique**. Finally, **Generic** NPs refer to classes or types of entities rather than specific entities. For example, Dinosaurs in Dinosaurs are extinct. is a **Generic** NP.	1061
Another important distinction CFD makes is between **Hearer_Old** for references to entities that are familiar to the hearer (e.g., if they are physically present in the speech situation), versus **Hearer_New** for nonfamiliar references. This distinction cuts across the two subparts of the hierarchy, **Anaphora** and **Nonanaphora**; thus, labels marking **Hearer_Old** or **Hearer_New** also encode other distinctions (e.g., **Unique_Hearer_Old**, **Unique_Hearer_New**, **Nonunique_Hearer_Old**).	1061

[9] The excerpts were re-organized into bullet points, and quotation marks were not used for readability.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, R, W. (1978). An Implicational Model for Second Language Research, *Language Learning 28*, 221-282.

Bhatia, A., Lin, C., Schneider, N., Tsvetkov, Y., Talib Al-Raisi, F., Roostapour, L., Bender, J., Kumar, A., Levin, L., Simons, M., & Dyer, C. (2014a, August). Automatic Classification of Communicative Functions of Definiteness [Technical papers]. In the Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland, pp. 1059-1070. Retrieved from https://www.aclweb.org/anthology/C14-1.pdf

Bhatia, A., Lin, C., Schneider, N., Tsvetkov, Y., Talib Al-Raisi, F., Roostapour, L., Bender, J., Kumar, A., Levin, L., Simons, M., & Dyer, C. (2014b, August). Automatic Classification of Communicative Functions of Definiteness [Poster presentation]. In the Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland. Retrieved from http://people.cs.georgetown.edu/nschneid/p/definiteness-poster.pdf

Bhatia, A., Simons, M., Levin, L., Tsvetkov, Y., Dyer, C., & Bender, J. (2014, May). Unified Annotation Scheme for the Semantic/Pragmatic Components of Definiteness. In the Proceedings of the 9th Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, pp. 910-916. Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/pdf/1194_Paper.pdf

Bickerton, D. 1981. *Roots of Language*. Ann Arbor MI: Karoma Press.

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

Butler, Y. G. (2002). Second Language Learners' Theories on the Use of English Articles: an Analysis of the Metalinguistic Knowledge Used by Japanese Students in Acquiring the English Article System. *Studies in Second Language Acquisition, 24*(3), 450–481. https://doi.org/10.1017/s0272263102003042

Cancino, H, E. Rosansky, J. and Schumann, J.H., 1978. The acquisition of English negatives and interrogatives by native Spanish speakers. In E. Hatch (Ed.) *Second Language Acquisition* (Rowley, Mass.: Newbury House Publishers), pp. 207-230.

Carlsen, C. 2012. Proficiency level – A fuzzy variable in computer learner corpora. Applied Linguistics 33(2): 161–183. DOI: 10.1093/applin/amr047

Celce-Murcia, M., Larsen-Freeman, D., & Williams, H. A. (1999). *The grammar book: An ESL/EFL teacher's course*. Boston : Heinle & Heinle.

Collins, P. (2007). Can/could and may/might in British, American and Australian English: a corpus-based account. *World Englishes, 26*(4), 474-491. doi:10.1111/j.1467-971X.2007.00523.x

Crosthwaite, P. (2016). L2 English article use by L1 speakers of article-less languages. *International Journal of Learner Corpus Research, 2*(1), 68–100. https://doi.org/10.1075/ijlcr.2.1.03cro

Deshors, S. (2016). Multidimensional perspectives on interlanguage: Exploring may and can across learner corpora. Corpora and Language in Use. Presses Universitaires de Louvain.

Deshors, S., & Gries, St. Th. (2016). Profiling verb complementation constructions across New Englishes: A two-step random forests analysis of ing vs. to complements. *International Journal of Corpus Linguistics, 21*(2), 192-218. doi:10.1075/ijcl.21.2.03des

Diez-Bedmar, M. B. & Papp, S. 2008. "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. Amsterdam: Rodopi, 147–175. doi: 10.1163/9789401206204_007

Diez-Bedmar, M. B. 2015. "Article use and criterial features in Spanish EFL writing". In M. Callies & S. Götz (Eds.), Learner Corpora in Language Testing and Assessment. Amsterdam: John Benjamins, 163–190. doi: 10.1075/scl.70.07die

Doetjes, J. (2017). The count/mass distinction in grammar and cognition. *The Annual Review of Linguistics, 3*, 199-217. doi:10.1146/annurev-linguistics-011516-034244

Dulay, H., M. Burt, and S. Krashen (1982). Language Two. New York: Oxford University Press.

Durham, M. (2011). I think (that) something's missing: Complementizer deletion in nonnative emails. *Studies in Second Lanuage Learning and Teaching 1*(3), 421–445. doi:10.14746/ssllt.2011.1.3.6

Friedrich, R. J. (1982). In Defense of Multiplicative Terms in Multiple Regression Equations. *American Journal of Political Science*, *26* (4), 797-833. Retrieved from http://links.jstor.org/sici?sici=0092-5853%28198211%2926%3A4%3C797%3AIDOMTI%3E2.0.CO%3B2-A

Fuchs, R. (2017). Do women (still) use more intensifiers than men? Recent change in the sociolinguistics of intensifiers in British English. *International Journal of Corpus Linguistics, 22*(3), 345–374. doi:10.1075/ijcl.22.3.03fuc

Geertzen, J. , Alexopoulou, T., & Korhonen, A. (2013). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). Selected Proceedings of the 31st Second Language Research Forum (SLRF), Cascadilla Press, MA.

Goldschneider, J., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*(1), 1-50. doi: 10.1111/1467-9922.00147

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In Granger, S. (ed.) Learner English on Computer. Addison Wesley Longman : London & New York, 3-18.

Granger, S., E. Dagneaux and F. Meunier (2002). International Corpus of Learner English, Louvain: UCL.

Gries, St. Th., & Deshors, S. (2014). Using regressions to explore deviations between interlanguage and native language: Two suggestions. *Corpora 9*(1): 109-136. doi: 10.3366/cor.2014.0053

Gries, St. Th., & Wulff, S. (2013). The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics 18*(3), 327–356. doi:10.1075/ijcl.18.3.04gri

Han, N. R., Chodorow, M. & Leacock, C. 2006. "Detecting errors in English article usage by non-native speakers", *Natural Language Engineering 12*(2), 115–129. doi:10.1017/S1351324906004190

Huang, Y., Geertzen, J., Baker, R., Korhonen, A., & Alexopoulou, T. (2017). The EF Cambridge Open Language Database (EFCAMDAT): Information for Users. Retrieved from https://corpus.mml.cam.ac.uk/efcamdat2/public_html/EFCamDat-Intro_release2.pdf

Huang, Y., Murakami, A., Alexopoulou, T., & Korhonen, A. (2018). Dependency parsing of learner English. International Journal of Corpus Linguistics, 23(1), 28-54.

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

Huebner, T. (1985). System and variability in interlanguage syntax, *Language Learning 35*, 141–163. doi: 10.1111/j.1467-1770.1985.tb01022.x

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150–169. doi:10.1017/S0267190512000037

Ionin, T., Ko, H., Wexler, K. (2004). Article Semantics in L2 Acquisition: The Role of Specificity. *Language Acquisition, 12*(1), 3-69. doi: 10.1207/s15327817la1201_2

Ishikawa, S. 2011. "A new horizon in learner corpus studies: The aim of the ICNALE project". In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), Corpora and Language Technologies in Teaching, Learning and Research. Glasgow: University of Strathclyde Press, 3–11.

Ishikawa, S. 2013. “The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English”. In S. Ishikawa (Ed.), Learner Corpus Studies in Asia and the World. Kobe: Kobe University School of Languages and Communication, 91–118.

Lee, H. (1999). Variable article use in Korean Learners of English. *University of Pennsylvania Working Papers in Linguistics, 6*(2), 35-47. Retrieved from https://repository.upenn.edu/pwpl/vol6/iss2/4/

Leroux, W., Kendall, T. (2018). English article acquisition by Chinese learners of English: An analysis of two corpora. *System, 76*, 13-24. doi: 10.1016/j.system.2018.04.011

Lester, N. A. (2019). That’s hard: Relativizer use in spontaneous L2 speech. *International Journal of Learner Corpus Research, 5*(1), 1–32. doi:10.1075/ijlcr.17013.les

Liu, Y., Lu, X. (2020). Chinese EFL learners’ misconceptions of noun countability and article use. *System, 90.* 1-12. doi:10.1016/j.system.2020.102222

Master, P. (1988, March). Acquiring the English article system: A cross-linguistic internalnguage analysis. Paper presented at the Annual Meeting of the Teachers of Speakers of Other Languages, Chicago, IL.

Master, P. (1994). “Efiect of Instruction on Learning the English Article System.” In T. Odlin (ed.), Perspectives on Pedagogical Grumman 229-252. New York: Cambridge University Press.

Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: a learner corpus study. *Studies in Second Language Acquisition, 38*(3), 365-401. doi:10.1017/S0272263115000352

Pallotti, G. (2007). An Operational Definition of the Emergence Criterion. *Applied Linguistics, 28*(3), 361-382. doi: :10.1093/applin/amm018

Pica, T. (1984). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition, 6*, 69-78.

Pienemann, M. (1998). Language processing and second language development: Processability Theory. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/sibil.15

Snape, N. (2008). Resetting the Nominal Mapping Parameter in L2 English: Definite article use and the count–mass distinction. *Bilingualism: Language and Cognition, 11*(1), 63-79. doi: 10.1017/S1366728907003215

Snape, N., García-Mayo, M. D. P., & Gürel, A. (2013). L1 transfer in article selection for generic reference by Spanish, Turkish and Japanese L2 learners. *IJES (Internal Journal of English Studies), 13*(1), 1–28. Retrieved from http://revistas.um.es/ijes

Song, E., & Sung, M. (2017). A corpus-based study of contextual factors influencing Korean EFL learners' dative alternation: Lexical verbs, syntactic weights, and information structures. *Australian Review of Applied Linguistics, 40*(1), 19–39. doi:10.1075/aral.40.1.03son

Teng, X. (2012). How Japanese Learners Use English Articles in Sentences with Be in Contrast to Sentences with Other Verbs. *US-China Foreign Language, 10*(8), 1401-1404.

Thewissen, J. (2013). Capturing L2 Accuracy Developmental Patterns: Insights From an Error-Tagged EFL Learner Corpus. *The Modern Language Journal, 97*(S1), 77–101. doi:10.1111/j.1540-4781.2012.01422.x

Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics, 10*, 335–355. doi:10.1017/s0142716400008663

Tse, G. W. (2001). The Grammatical Factors Influencing the Choice between the Use and Omission of the Definite Article Preceding Multi-Word Organization Names: A Statistical Analysis. *Journal of Quantitative Linguistics, 8*(1), 13-32. doi: 10.1076/jqul.8.1.13.4090

Vendler, Z. (1957) Verbs and Times. *The Philosophical Review, 66*(2), 143-160. doi: 10.2307/2182371

Walters, E. (2016). The *P*-value and the problem of multiple testing. *Reproductive Biomedicine Online, 32*(4), 348-349. doi:10.1016/j.rbmo.2016.02.008

White, B. J. (2010). In search of systematicity: A conceptual framework for the English article system. Unpublished doctoral dissertation. East Lansing, MI: Michigan State University.

Wulff, S., & Gries, St., Th. (2015). Prenominal adjective order preferences in Chinese and German L2 English: A multifactorial corpus study. *Linguistic Approaches to Bilingualism, 5*(1), 122-150. doi:10.1075/lab.5.1.05wul