【Articles】

Accounting for the Variability of L2 Article Choices
L2 冠詞選択におけるばらつきを解く

OGAWA Mutsumi

（小川睦美）

Abstract

This paper addresses the issue of second language (L2) learners' variability in English articles from the perspective of interaction of context and noun type. While the role of semantic contexts (i.e., definiteness, specificity) is commonly suggested to account for article choices (Ionin, Ko, & Wexler, 2004), some studies have documented different degrees of contextual effects according to the type of noun that follows articles (Snape, 2008; Trenkic, 2008; White, 2009). A forced-choice elicitation task on English articles (*the, a, ø*) involving count (singular, plural), mass, and flexible nouns (nouns that can equally be interpreted as count and mass) was administered with Japanese learners of English at three different proficiency levels. Results showed that L2 learners are less sensitive to semantics of the context where mass and plural nouns are involved, suggesting that the fluctuation between definiteness and specificity interacts with countability and number of nouns in context. In conclusion, this article argues that L2 learners experience a more prolonged learning difficulty with the count-mass distinction than with definiteness in the acquisition of articles.

1．Introduction

English articles have been known as a grammatical property which imposes great difficulty on second language (L2) learners, especially those who lack articles in their first language (L1), such Japanese, Russian, and Korean. The following examples illustrate the task facing those learners:

(1)　a. The cat is drinking the milk.
　　　b. A cat is drinking milk.
　　　c. The cats are drinking the milk.

> d. *A cats are drinking *a milk.
> e. Cats are drinking milk.
> f. *Cat is drinking milk.

Some nouns can take both *the* and *a*: *cat* in (1a) and (1b). Others can be preceded by *the*, but not *a*: *cats* and *milk* in (1c) and (1d). Some nouns can occur with a zero article *ø*, *cats* and *milk* in (1e), but some cannot as in *cat* in (1f). On top of this distribution of noun types, L2 learners must establish meaning contrasts determined by different choices of articles. A number of existing studies have documented variability in L2 article use, such as substitution of *the* for *a*, and *a* for *the*, as well as omission of articles. Although such errors are often observed until quite advanced stages of proficiency, they are not random, indicating that learners may posit certain non-target-like hypotheses during development.

Huebner (1985), Parrish (1987) and Thomas (1989) attempted to explain L2 development in English article acquisition on the basis of two semantic notions: *hearer knowledge* (i.e., whether an entity referred to is already known to the hearer) and *specific referent* (i.e., whether a specific entity is being referred to). Since the three classical studies, research has progressed in the direction of scrutinising effects of contextual semantics, focusing on *definiteness* and *specificity* (Ionin et al., 2004 and many others). However, some recent studies have remarked that there is a gap to fill in the domain regarding the effect of nominal properties such as the count-mass distinction and singular-plural contrast (Snape, 2008; Trenkic, 2002; White, 2009). The current paper investigates how relevant factors (e.g., definiteness, specificity, count-mass distinction, plurality) contribute to the selection of articles by Japanese learners of English. In particular, the main purpose of this study is to demonstrate the magnitude of those variables by statistically calculating their individual contributing power to each article choice: *the*, *a/an*, and *ø*.

2. Literature review

2.1. Fluctuation hypothesis
One of the typical errors found in the learner's article use is the overuse of definite articles in indefinite contexts, as in (2). Under the context where John does not know what Mary got for her birthday as well as the referent *bicycle* has never appeared in their conversation, the correct choice of article will be the indefinite article *a*.

> (2) John: What did you get for your birthday yesterday?
> Mary: I got lots of things, but my father bought me *the* bicycle and I love it the most.

This substitution error is likely to occur systematically in certain contexts, whereby a semantic notion *specificity* causes the definite article to be misused in specific indefinite contexts (Ionin et al., 2004; Thomas, 1989). Ionin et al. (2004, p. 5) define *definiteness* and *specificity* as follows:

  (3) [+definite][1]
       If a determiner phrase (DP) is definite, then the speaker and hearer presuppose the existence of a unique individual in the set denoted by the noun phrase (NP):
  (4) [+specific]
       If a DP is specific, then the speaker intends to refer to a unique individual in the set denoted by the NP and considers this individual to possess some noteworthy property.

Possible semantic contexts where articles are used can be characterised by combination of the two notions. For instance, there are specific and non-specific definite NPs, both of which are marked by the definite article *the* in English. In (5), the speaker intends to refer to a particular individual who wins the race, who is marked as [+definite] because there will be only one person who can be the winner of any race. The speaker also knows this person as a friend, which makes *the winner* [+specific] because the speaker considers him/her to possess a noteworthy property of being a friend. On the other hand, in (6), the speaker is simply referring to whoever happens to be the winner without intention to refer to a certain individual with noteworthy property, which then lets *the winner* defined as [−specific].

  (5) [+definite/+specific]
       e.g., I want to talk to *the winner* of the race. She is a good friend of mine.
  (6) [+definite/−specific]
       e.g., If you want to talk to *the winner*, wait until the end of the race.

In indefinite contexts, the indefinite article *a/an* is used for singular nouns, and the zero article *ø* for plural and mass nouns. There are also specific and non-specific indefinite NPs. Specificity in indefinite contexts can be distinguished by whether or not an article can be replaced with referential *this*. In colloquial English, when an NP refers to something specific as well as noteworthy to the speaker ([+specific]) but not identifiable to the hearer ([−definite]), referential *this* can replace articles, as shown in (7). In contrast, replacing is not infelicitous in non-specific contexts in (8), where the speaker does not intend to refer to a particular entity.

(7)   [–definite/+specific]

e.g., You should go and see *a film/this film* at the TOHO Cinemas tonight – it's superb and directed by Steven Spielberg.

(8)   [–definite/–specific]

e.g., You should go and see *a film/??this film* at the TOHO Cinemas tonight – don't just sit at home.

English distinguishes [±definite] by articles; however, there are some languages which mark specificity by articles. For example, Samoan differentiates articles on the basis of [±specific]: *le* for specific or *se* for non-specific, marking neither definite nor indefinite by articles. Given such cross-linguistic contrasts, Ionin, Ko, and Wexler (2003, p. 248) proposed the Article Choice Parameter (ACP), under which both definiteness and specificity are universal semantic notions listed in the inventory of universal grammar (UG) and a language with two articles distinguishes them on the basis of either definiteness (e.g., English) or specificity (e.g., Samoan). Ionin et al. (2004) have adapted the idea of this parameter to the L2 article acquisition, advocating the Fluctuation Hypothesis (FH). According to the FH, L2 learners fluctuate between different parameter-settings of the ACP until the input leads them to set the parameter to the appropriate value. For instance, the FH predicts fluctuation to occur in contexts where the values of the two features conflict, such as [–definite/ +specific] and [+definite/–specific] contexts. L2 learners in the course of English article acquisition are then predicted to show a tendency to mark [±specific] by articles, resulting in using *the* more often in [–definite/+specific] than in [–definite/–specific] contexts, and more *a/an* or *ø* in [+definite/–specific] than in [+definite/+specific] contexts.

2.2. Problems in the FH

Although the FH seems to give a plausible account to the overuse of definite articles, there are some flaws in the analysis of the article distribution in Samoan. Tryzna (2009) illustrates that Samoan does not actually distinguish specificity within definite contexts, marking both specific and non-specific definite nouns with *le*, as shown in (9). Samoan articles mark non-specificity only in indefinite contexts.

(9)   a. [+definite/+specific]   *le* malo

e.g., I want to talk to *the winner* of the race. She is a good friend of mine.

b. [+definite/–specific]   *le* malo

e.g., If you want to talk to *the winner*, wait until the end of the race.

c. [–definite/+specific]   *le* pulou

e.g., I'm looking for *a hat*. I must have left it here yesterday.

d. [–definite/–specific]   *se* pulou

e.g., I'm looking for *a hat* to go with my new coat.
(Adapted from Table 1 in Tryzna, 2009, p. 71)

Another problem is that the distribution of articles was only reviewed with regard to singular count nouns, not with plural and mass nouns. In English, any noun can be preceded by *the* in definite contexts, whereas in indefinite contexts, *a* is given to singular count nouns and *ø* to plural and mass nouns. In Samoan, however, singular and plural nouns are differently marked in definite as well as in specific indefinite contexts. For instance, Samoan has no plural marker like English *–s*. A non-specific (indefinite) article *ni* is used for plural entities, whereas absence of an overt article indicates the NP being plural and specific, as illustrated in (10). Thus, nouns without an overt article refer to specific plural entities in contrast to presence of the article *ni* (Mosel & Hovdhaugen, 1992).

(10)   *le teine*   *ø teine*   *ni teine*
       girl [+specific]   girls [+specific] girls [–specific]

According to Mosel and Hovdhaugen (1992, p. 94), furthermore, there is no morphosyntactic distinction between mass and count nouns in Samoan. For example, *vai* 'water' can form the nucleus of both singular and plural NPs. Hence, the absence of an article, which is a plural form syntactically, indicates a specific denotation as a number of units or portions of mass property (i.e., *ø vai* interpreted as certain units of water like *glasses of water* or *bottles of water*), and the size and configuration of the units or portions are understood from the context (Mosel & Hovdhaugen, 1992, p. 269).

To sum up, some intra-linguistic differences exist in morphological marking between singular and plural NPs across specificity and definiteness both in Samoan and in English. Table 1 is the summary of article distributions in both languages and predicted fluctuation for L2 English articles.

Table 1   Summary of specificity and definiteness interaction

DP type	Specificity setting (e.g., Samoan)		Definiteness setting (e.g., English)		L2-English fluctuation			
	singular	plural	singular	plural/mass	singular		plural/mass	
[–def/–spec]	*se*	*ni*	*a*	*ø*	*a*		*ø*	
[–def/+spec]	*le*	*ø*	*the*		*a*	*the*	*ø*	*the*
[+def/–spec]								
[+def/+spec]								

The FH claims that fluctuation can be observed where the two settings (i.e., specificity and definiteness settings) give different results, predicting the erroneous use of *a/ø* in [−definite/+specific] contexts. If the singular-plural (mass) contrast is assumed to be part of universal semantics which prompts a morphological marking either through articles or plural markers, the two settings also differ in how the number contrast is realised in definite contexts: In Samoan, the number feature is incorporated in articles and specified by *le* or *ø*, whereas English uses a single morpheme *the* for both singular and plural NPs and the number is independently marked by the presence or absence of *−s*. If L2 learners are susceptible to different settings of universal semantics which require morphological distinction, the FH may need to predict fluctuation between singular and plural NPs in definite contexts in L2 English. Theoretically speaking, in other words, L2 learners could opt for making a distinction between singular and plural NPs by articles even in definite contexts.

2.3. Effects of noun types

Despite its explanatory power for common patterns in L2 article use, some studies have questioned the validity of the FH. For instance, it is not clear to what extent the FH can account for variability found with different types of noun: singular, plural, and mass nouns. Many studies which investigate the acquisition of English articles in relation to the FH have commonly used singular count nouns and produced similar results supporting the FH (Garcia Mayo, 2009; Hawkins et al., 2006; Kim & Lakshmanan, 2009; Snape, 2005, 2009; Tryzna, 2009). However, studies including plural count and mass nouns have reported divergent findings, suggesting no consensus regarding fluctuation with plural and mass nouns. A summary of the results of previous studies which investigated suppliance of *the* in indefinite contexts is

Table 2   Suppliance of *the* in indefinite contexts in previous studies (%)

L1	Proficiency	Singular		Plural		Mass	
		+spec	−spec	+spec	−spec	+spec	−spec
Japanese[1]	Inter.	40.0	15.0	35.0	10.0	0.0	5.0
	Adv.	20.0	5.0	20.0	5.0	5.0	10.0
Japanese[2]	Upper Inter. - Lower Adv.	50.0	8.0	58.0	19.0		
Polish	Inter.	37.0	40.0	20.0	20.0		
	Adv.	23.0	15.0	8.0	13.0		
Chinese	Adv.	22.0	5.0	18.0	18.0		
Russian	Upper Inter.	36.0	7.0				
Korean	Adv.	22.0	4.0				

Japanese[1] (Snape, 2005); Japanese[2] (Hawkins et al., 2006); Polish and Chinese (Tryzna, 2009); Russian and Korean (Ionin et al., 2004), whose proficiency levels are estimated in comparison to other studies.

shown in Table 2. Looking at singular nouns, fluctuation is observed in the form of more uses of *the* in [+specific] than in [–specific] contexts. On the other hand, studies with plural nouns demonstrated that *the* was used to mark [+specific] by Japanese learners, whereas no such overuse was found with leaners whose L1s were Polish or Chinese, both of which were article-less languages. Moreover, a study including mass nouns reported no overuse of *the* as a specificity marker.

In addition, some research pointed out that learners with lower proficiency spread *ø* across plural and mass definite contexts (Snape, 2005; Tryzna, 2009). Table 3 summarizes percentages of erroneous article uses in specific definite contexts reported in previous studies.

Table 3   Suppliance of *a* and *ø* articles in [+definite/+specific] contexts in previous studies (%)

L1	Proficiency	Singular		Plural		Mass	
		a	*ø*	*a*	*ø*	*a*	*ø*
Japanese	Inter	10.6	6.6	2.0	26.0	6.1	34.4
	Adv	7.2	3.9	1.0	18.0	1.7	32.8
Spanish	Inter	4.4	5.0	4.0	11.0	2.8	13.3
	Adv	1.1	1.7	2.8	6.1	0	6.1
Syrian Arabic	Inter	4.0	2.0	1.0	9.0	5.0	22.0
	Adv	1.0	0	2.0	4.0	0	17.0
French	Inter	6.0	0	8.0	4.0	1.0	25.0
	Adv	2.0	0	0	9.0	0	8.0

Japanese and Spanish (Snape, 2008), Syrian Arabic and French (Sarko, 2009)

If the misuse of *ø* with plural and mass nouns occurs due to misunderstanding of definite contexts as indefinite, the use of *a* with singular count nouns should equally be expected at similar rates. However, it is unlikely and even learners with L1s which have articles (i.e., Spanish, French, only definite articles in Syrian Arabic) exhibit slightly higher suppliance of *ø* with plural nouns and greater use with mass nouns, compared to *a* with singular nouns. Thus, this tendency cannot be explained solely by hypotheses regarding semantic contexts, such as the FH.

In short, there was a tendency of learners to extend the definite article into indefinite contexts (i.e., fluctuation), especially with singular count nouns (Ionin et al., 2004 among others), sometimes with plural count nouns (Hawkins et al., 2006; Tryzna, 2009), but never with mass nouns (Snape, 2005). Furthermore, the zero article is extensively used with plural and mass nouns in definite contexts, which casts a doubt in the assumption that it is only the semantic context that predominantly influences L2 learners' article choice. A potential rea-

son for the diverging findings in the literature regarding the specificity effect and noun types may be task effects. For instance, some studies only used singular count nouns, the majority of which were nouns referring to human (e.g., *student, friend*). In such a case, the countability and number of target nouns in the task are so obvious that learners can spare more cognitive resources to concentrate on contextual information, which may help them recognise subtle differences between specific and non-specific referents. These differences may possibly be overshadowed by judgment of countability and number when various types of nouns are involved in a task.

2.4. Other approaches to L2 variability of articles

English has some nouns which can occur equally well as count and mass depending on context, as in (11) and (12). Such nouns as *brick* and *fear* are called flexible nouns in this article.

> (11) a. Builder knows how to lay *bricks* properly.
>   b. This house is built of *brick*.
> (12) a. I have *a fear* of spiders.
>   b. She ran away from *fear*.

Native speakers rely on context to determine whether nouns are to be either count or mass, resulting in correct article choices between *a* and *ø*. For example, even though they perceive nouns in isolation from context as mass (e.g., *water*), they can use contextual information to interpret them as count when necessary (e.g., Could I get *a water*, please?). L2 learners are, however, reported to have certain preferences to regard nouns as either count or mass on the basis of what class they belong to (i.e., individual, material, abstract, proper nouns), irrespective of the context which actually provides one particular interpretation (Hiki, 1990). The biased countability judgment in L2 was also reported by Yoon (1993), who claimed that L2 learners would judge noun countability by intuition and stick with their intuitive countability to decide which article to use, *a* or *ø*.

In relation to the effect of nouns, other studies have taken different approaches to calculate and compare weights of influencing factors for each article choice (i.e., *the*, *a*, *ø*), such as noun type, countability and number, thematicity, the position of NP in the clause, and so on (White, 2009; Young, 1996). In particular, White (2009) examined effects of semantic contexts (i.e., [±definite] and [±specific]) and three noun types: imageable count (e.g., *book, creator, play*), abstract count (e.g., *life, style, environment*), and noncount (e.g., *bread, money, furniture*). As expected, results revealed that semantic context was the only significant predictor for the choice of *the* and that the choices of *a* and *ø* were significantly influenced by both semantic context and noun type. Interesting was that the effect of semantic context proved to be stronger than that of noun type for the use of *a*, whereas noun type was stron-

ger than context in the case of *ø*. In other words, L2 learners tend to rely more on context when they use *the* and *a*, but noun type when they use *ø*. It is therefore highly probable that they would use *ø* invariably with particular noun types, regardless of what kind of context nouns are embedded in.

White (2009) additionally elicited explanations from learners why they had chosen certain articles. The analysis of learners' self-monitoring showed that in indefinite contexts, they chose articles systematically on the basis of the countability they assigned to nouns; that is, they used *a* with nouns which they thought were count and *ø* with nouns which they regarded as mass. This is consistent with Hiki's (1990) claim. Moreover, a tendency was reported that learners used *ø* even in definite contexts when they considered nouns as mass, and Snape (2008) also documented the flooded use of zero articles with definite mass nouns. All the above findings imply that L2 learners' difficulty with article use lies not only in discriminating differences between semantic contexts (definite vs. indefinite), but also in recognising correct countability and number of nouns (singular, plural, vs. mass). It may be this interweaving aspect of several features that makes article choice problematic for L2 learners and perhaps brings different levels of learning tasks at different stages in L2 development.

3. Research

3.1. Research questions

This study examines the extent to which each factor related to articles (i.e., definiteness, specificity, count-mass distinction, number marking, noun flexibility) contribute to the choice of the definite, indefinite, and zero article respectively. It also aims to reveal how the magnitude of each factor may change in the course of L2 acquisition.

3.2. Participants

Fifty-three participants took part in the study and were all recruited in a university in the United Kingdom. Table 4 shows participant information. There were 15 native speakers for the control group (NC), and L2 learners were divided into three proficiency levels based on the scores of Oxford quick placement test (UCLES, 2001): 12 lower intermediate (LI), 15 upper intermediate (UI), and 11 lower advanced learners (LA). All the learner groups were significantly different from each other in terms of English proficiency ($F(2, 35) = 100.482$, $p < .0005$). Neither correlation between the onset age of their English learning and the proficiency score ($r = .204$, $p > .05$) nor between the length of stay in English speaking countries and the proficiency score ($r = .285$, $p > .05$) was found.

Table 4　Participant information

Groups	QPT score (range, %)	Age (range)	Onset age of English learning (range)	Length of stay in month (range)
Lower Inter. (n = 12)	35.8 (30-39, 59.7%)	26.1 (21-35)	11.8 (9-13)	11.3 (3-60)
Upper Inter. (n = 15)	43.7 (40-47, 72.9%)	23.7 (20-29)	12.1 (10-13)	21.5 (1-102)
Lower Adv. (n = 11)	50.0 (48-53, 83.3%)	25.5 (19-31)	12.2 (11-13)	35.7 (1-84)
Native (n = 15)	n/a	24.3 (19-36)	n/a	n/a

3.3. Instrument

A task was designed in a forced-choice elicitation format, where participants read short dialogues and chose an article from *the*, *a/an*, or *ø* described as '—' in the task. Semantic contexts were controlled according to the combination of definiteness and specificity defined by Ionin et al. (2004). Specificity was operationalized in terms of speaker knowledge, namely, whether the speaker explicitly showed or denied knowledge about the referent. Explicit speaker knowledge was described by the mention of the referent after the target NP was introduced in the dialogue. There were four contexts as in (13) – (16):

(13) [+definite/+specific]
　　Mark: We had so much fun in the zoo today.
　　Christina: Yeah, with lots of cute animals! Which animal did you like the best?
　　Mark: Well, I think I liked (a, the, —) pony. It was very friendly and fun riding on it.
　　Christina: Me too! It was so adorable.

(14) [+definite/−specific]
　　(Phone conversation)
　　Jim: Hi, is Max there?
　　Carrie: Hi, Jim. He's out at the moment.
　　Jim: Do you know where he's gone?
　　Carrie: He said he's ordered a replacement for (a, the, —) bowl of his food processor. I didn't know he had such a thing, but I think he's gone to get it in the town.

(15) [−definite/+specific]
　　Jane: Did you do anything fun this weekend?
　　Ian: Yes, I went to Sherwood Forest. And when I was walking there, I found (a, the,

—) castle deep in the forest. It seems to be known as the Haunted Tower by local people.

(16) [–definite/–specific]

Terry: Does your neighbour Mr Sadler have any pets?

Cindy: I don't know. But I think he used to have (a, the, —) cat. I never saw it myself, but my mom sometimes heard it meowing in our garden.

Target nouns were selected from the words used in Taler and Jarema's study (2007), which investigated English native speakers' processing of the count-mass distinction. The items were controlled for frequency, concreteness, and imageability. There were two major types of nouns: flexible (i.e., nouns which can be interpreted equally well as count and mass) or non-flexible type (i.e., nouns with a strong preference for either count or mass interpretation, typical count or mass nouns). Frequency rates of target nouns and their familiarity rates for Japanese learners of English were considered in the item selection. First, nouns were ranked based on a word familiarity list in Yokokawa (2009) as well as the frequency based on the British National Corpus. The entire procedure resulted in the selection of 48 nouns as shown in Table 5: 24 flexible nouns, and 24 non-flexible nouns.

Table 5   Target nouns

Flexible			Non-flexible		
			Count		
Singular	Plural	Mass	Singular	Plural	Mass
candy	onions	cane	beetle	bells	beauty
debt	pipes	cloud	bowl	buttons	beef
dessert	prayers	pepper	castle	clocks	damage
maple	pumpkins	talent	cat	dragons	fabric
rope	ribbons	thread	donkey	engines	pork
salad	stones	turkey	eagle	fountains	sugar
shadow	strings	wire	pony	helmets	traffic
steak	treasures	wonder	ticket	medals	wax

They were presented in singular, plural, or mass form in the task. In this study, singular means a countable form which requires the indefinite article in indefinite contexts, whereas mass indicates an uncountable feature which selects the zero article in indefinite contexts. Therefore, there were 8 singular, 8 plural and 8 mass forms in each noun type. The items were also tested in a lexical decision task conducted to the same participants in a different

study, so that it was proved that the participants were fairly familiar with the items: response accuracies of 95.67% by Lower Inter., 97.14% Upper Inter., 99.24% Lower Adv., and 99.40% Native.

The distribution of semantic contexts and noun types along with the choice of correct articles is shown in Table 6. There were four types of sematic contexts, each containing two types of nouns (flexible and non-flexible) in three forms (singular, plural, and mass). It in total made six different noun types in one context. There were two tokens for each combination of semantic context and noun type. In total, the task contained 48 dialogues. The environment of the target NPs was set under the following criteria: no adjectives were attached, most of them were in the object position of the sentence, and some were a predicate of prepositional phrases. The task was checked by two native speakers of English prior to the experiment and they provided target answers correctly.

Table 6   Distribution of correct articles

Semantic settings	Flexible			Non-flexible		
	sing.	plur.	mass	sing.	plur.	mass
+definite / +specific	*the*	*the*	*the*	*the*	*the*	*the*
+definite / −specific	*the*	*the*	*the*	*the*	*the*	*the*
−definite / +specific	*a*	*ø*	*ø*	*a*	*ø*	*ø*
−definite / −specific	*a*	*ø*	*ø*	*a*	*ø*	*ø*

4. Results

The overall accuracy of each group is as follows: LI 61.7%, UI, 70.6%, LA, 77.5%, NC 89.6%. The accuracy of the native control may arguably be considered low. An individual item analysis revealed that there were nine items to which the native participants showed less than 80% of agreement. Although exclusion of those items improved the accuracy to 94.6%, it did not remarkably change interpretations of statistical robustness and trends that will be discussed in the following sections. In addition, there was one participant whose accuracy rate was 68.1%. Exclusion of this person increased the native group's accuracy to 91.2%; however, the total variability of the group was unlikely to stem from this participant. Therefore, all the test items as well as participants were included in the following analyses.

Generalized estimating equations (binomial logistic regression with repeated data) were performed in order to estimate how much each independent variable contributes to article choice. The analysis was conducted for each article by each group. There were three dependent variables: *the, a/an,* and *ø*. Independent variables were named as follows: *definiteness* (definite or indefinite), *specificity* (specific or non-specific), *flexibility* (flexible or non-flexible),

mass (count or mass), and *plurality* (plural or singular). In the regression analysis, the reference category was set to be the latter of the pair. For example, indefinite is the reference category for the variable *definiteness*. Therefore, positive statistical values indicate an effect for definite over indefinite, while negative ones for indefinite over definite. For the convenience to create dummy variables for noun types, mass nouns were set as the reference category against singular count nouns, although the variable name *mass* was chosen instead of count. The following sections report statistical results of unstandardized coefficients (*B*), standard errors (*SE*), standardized coefficients (*β*), and exponentiated beta coefficients (*Exp (B)*). Important in this analysis are standardized coefficients and exponentiated beta coefficients. The former is an indicator with which the degree of contribution of each variable to the choice of an article can be measured and directly compared against that of other variables. The latter shows odds ratio that tells you how much a change in the independent variable increases or decreases the probability of article choice. For instance, if *Exp (B)* of definiteness is 3 for the choice of *the*, then it means that *the* is three times more likely to be chosen when the context is definite than when it is indefinite.

4.1. Binomial logistic regression <*the*>

Table 7 shows results of the definite article choice. *Definiteness* was the most significant determinant for the choice of *the* in all groups. The two intermediate groups, LI and UI, depended primarily on [+definite] feature of the context to use *the*, which was in fact the only

Table 7　Binomial logistic regression for the choice of the definite article (*the*)

LI	*B (SE)*	*β*	*Exp (B)*	UI	*B (SE)*	*β*	*Exp (B)*
Flex.	0.13 (0.21)	.015	1.14	Flex.	0.22 (0.15)	.024	1.24
Plural	0.54 (0.28)	.055	1.71	Plural	−0.05 (0.22)	−.005	0.96
Mass	0.43 (0.25)	.045	1.54	Mass	−0.16 (0.19)	−.017	0.85
Def.	1.75 (0.26)***	.191	5.78	Def.	2.89 (0.41)***	.315	17.98
Spec.	0.02 (0.21)	.002	1.02	Spec.	0.14 (0.24)	.016	1.15
Constant	−2.16 (0.22)			Constant	−2.59 (0.42)		

LA	*B (SE)*	*β*	*Exp (B)*	NC	*B (SE)*	*β*	*Exp (B)*
Flex.	−0.31 (0.14)*	−.035	0.74	Flex.	0.18 (0.22)	.022	1.19
Plural	−0.34 (0.36)	−.036	0.71	Plural	0.48 (0.37)	.056	1.61
Mass	−0.91 (0.22)***	−.098	0.40	Mass	−1.09 (0.30)***	−.128	0.34
Def.	4.16 (0.39)***	.445	63.85	Def.	5.00 (0.34)***	.554	148.07
Spec.	0.18 (0.20)	.021	1.20	Spec.	0.03 (0.23)	.004	1.04
Constant	−2.94 (0.37)			Constant	−2.20 (0.36)		

*$p < .05$, **$p < .01$, ***$p < .001$

significant factor. The advanced and native groups showed the variable *mass* to be significant along with *definiteness*. However, it was presumably caused by two items in mass condition (*sugar, pepper*) which happened to produce fewer uses of *the* in [+definite/+specific] contexts than expected. This explanation is supported by the fact that *mass* in these two groups exhibited negative values, indicating that *the* was less likely to be used when nouns were mass. It is therefore an item-specific error, rather than a systematic pattern.

Looking at standardized coefficient values *β*, furthermore, *definiteness* weighed more than the other variables, and the effect became stronger as L2 proficiency increased (.191 → .315 → .445). This is also evident in exponentiated beta coefficients *Exp (B)*, such that the probability of choosing *the* in definite contexts is 5.78 times as much as that in indefinite contexts in LI, 17.98 times in UI, and 63.85 times in LA, clearly indicating that learners become capable of distinguishing [+definite] from [−definite] and correctly draw on [+definite] for the use of definite articles. In contrast, another contextual factor *specificity* was not significant and made very little contribution. None of the three nominal properties (i.e., *flexibility*, *plural*, *mass*) had meaningful effects. One exception was the advanced learners, who showed a significant main effect of *flexibility*. Although its contributing power was much less than *definiteness*, the result indicated that *the* was less likely to be used with flexible nouns than with non-flexible nouns.

4.2. Binomial logistic regression <*a/an*>
Table 8 presents results for the choice of *a/an*. First of all, *a/an* was used with plural nouns

Table 8　Binomial logistic regression for the choice of the indefinite article (*a/an*)

LI	B (SE)	β	Exp (B)	UI	B (SE)	β	Exp (B)
Flex.	−0.06 (0.17)	−.007	0.94	Flex.	−0.35 (0.17)*	−.035	0.71
Plural	−4.54 (0.41)***	−.458	0.01	Plural	−5.06 (0.56)***	−.455	0.01
Mass	−1.61 (0.31)***	−.172	0.20	Mass	−1.64 (0.22)***	−.154	0.19
Def.	−1.21 (0.31)***	−.138	0.30	Def.	−1.65 (0.25)***	−.165	0.19
Spec.	−0.59 (0.25)*	−.068	0.55	Spec.	−0.61 (0.24)*	−.061	0.54
Constant	1.83 (0.27)			Constant	1.65 (0.21)		

LA	B (SE)	β	Exp (B)	NC	B (SE)	β	Exp (B)
Flex.	−0.36 (0.28)	−.041	0.69	Duality	−0.22 (0.34)	−.020	0.80
Mass	−1.88 (0.32)***	−.209	0.15	Mass	−4.52 (0.78)***	−.405	0.01
Def.	−2.06 (0.33)***	−.228	0.13	Def.	−4.83 (0.67)***	−.431	0.01
Spec.	−0.91 (0.31)**	−.102	0.40	Spec.	−0.51 (0.26)*	−.047	0.60
Constant	1.66 (0.28)			Constant	2.20 (0.43)		

*p < .05, **p < .01, ***p < .001

by neither the advanced nor native groups. Zero instances cannot be analysed in regression analysis; therefore, the variable *plural* was excluded from the analyses of the two groups. This removal itself has proved that NPs with plural marker *–s* is a significant indicator to eliminate the choice of *a/an* from article options that L2 learners may select.

Plural was also significant in the two intermediate groups, and *mass* and *definiteness* were significant predictors in all the groups; however, they appeared differently in terms of degrees of contribution in each group. Negative values in the table indicate the direction of the effect, such that a negative value in the case of *plural* means that the singular form has a strong effect. The LI learners relied on nominal properties (–.458 for *plural*, –.172 for *mass*) more strongly than on *definiteness* (–.138). The value of *mass* became slightly lower in the UI group (–.154), whereas *definiteness* increased its value (–.165). These changes in the values indicate that learners in early stages of L2 learning tend to use information such as nouns being pluralised or countable to select *a/an*, regardless of the fact that the context is indefinite. As they increase knowledge in the L2, they may shift focus to contextual information instead of keeping dominant attention to nominal properties. *Definiteness* in fact gained more explanatory power as proficiency got higher (–.138 → –.165 → –.228). Consequently, the advanced group showed a contribution pattern close to that of native speakers, as indicated in the *β* weights of both groups.

As for *specificity*, it proved to be one of the significant determinants, but its predictive power was not as large as the other variables. The reliance of [–specific] is presumably due to the fact that L2 learners selected *a/an* in a fluctuation context [+definite/–specific] more frequently than in [+definite/ +specific], indicating that they used [–specific] as a cue for the choice of *a/an*. Lastly, a significant effect was detected with *flexibility* in the UI group. It technically implies that *a/an* was more likely to be used with non-flexible than flexible nouns. However, it did not have a particular pattern indicating more uses of *a/an* with all sorts of non-flexible nouns. It was therefore merely an accumulated result of *a/an* being used slightly more often with some non-flexible nouns than with flexible nouns.

4.3. Binomial logistic regression <ø>

The last analysis shows results for the choice of *ø* in Table 9. In the native control group, *plural*, *mass*, and *definiteness* were the most influential factors, and they contributed to a similar extent. The L2 groups also showed significances with those variables; however, the strengths of the variables varied. For instance, *plural* weighed the most of the three in all the L2 groups, indicating that *ø* is more likely to be used with plural nouns than singular nouns (.370 → .374 → .399). *Mass*, which signifies the use of *ø* with mass over singular count nouns, was the second strongest (.245 → .264 → .335). Interestingly, although *definiteness* was not as dominant as *plural* or *mass* especially in the intermediate groups, it generated the most remarkable change, that is, the largest increase in contributing power in the course of

Table 9　Binomial logistic regression for the choice of the zero article (ø)

LI	B (SE)	β	Exp (B)	UI	B (SE)	β	Exp (B)
Flex.	−0.10 (0.22)	−.011	0.90	Flex.	0.07 (0.13)	.008	1.07
Plural	3.76 (0.21)***	.370	43.08	Plural	3.51 (0.24)***	.374	33.50
Mass	2.43 (0.23)***	.245	11.36	Mass	2.42 (0.23)***	.264	11.23
Def.	−1.10 (0.13)***	−.119	0.33	Def.	−1.82 (0.19)***	−.212	0.16
Spec.	0.53 (0.26)*	.058	1.70	Spec.	0.36 (0.24)	.043	1.43
Constant	−2.97 (0.38)			Constant	−2.10 (0.20)		

LA	B (SE)	β	Exp (B)	NC	B (SE)	β	Exp (B)
Flex.	0.55 (0.19)**	.067	1.74	Flex.	−0.08 (0.25)	−.009	0.92
Plural	3.67 (0.46)***	.399	39.16	Plural	4.86 (0.55)***	.468	129.21
Mass	3.02 (0.37)***	.335	20.59	Mass	5.74 (0.58)***	.539	311.56
Def.	−2.56 (0.17)***	−.303	0.08	Def.	−4.35 (0.44)***	−.447	0.01
Spec.	0.54 (0.25)*	.066	1.72	Spec.	0.24 (0.28)	.026	1.27
Constant	−2.12 (0.39)			Constant	−3.47 (0.55)		

*$p < .05$, **$p < .01$, ***$p < .001$

L2 development (−.119 → −.212 → −.303). It can therefore be surmised that when learners select ø, they start out looking for clues in nouns and gradually rely on context at later stages of L2 acquisition. As seen with the advanced group, the balance of the three variables got closer to the pattern seen with the native speakers.

　　Moreover, the effect of *flexibility* was only evident with the advanced learners, indicating that they accepted ø more often with flexible nouns than non-flexible nouns. Although *specificity* showed a significant effect with the lower intermediate and advanced groups, it did not contribute as much as the other significant variables did.

5.　Discussion

The statistical results in Table 7, Table 8, and Table 9 are summarised in Table 10. Standard-ized coefficients β were modified in terms of direction of values according to the features presented in the table, and exponentiated beta coefficients *Exp (B)* were also recalculated against the opposite value.

　　First, all the groups mostly agreed about influential factors specified for each article choice: [+definite] was specified for the definite article, [−plural], [−mass], and [−definite] for indefinite *a/an*, and [+plural], [+mass], and [−definite] for zero ø. Interestingly, the degree to which L2 learners rely on each factor differed by each article as well as by proficiency. In the two intermediate groups, [−definite] is considerably less contributing than [−plural] for *a/an*

Table 10　Summary of standardized coefficients β of influential factors for article choice (*Exp (B)*)

Group	*the*[2]	*a/an*	*ø*
Lower Inter.	[+definite] .191 (5.78)	[–plural] .458 (93.51) [–mass] .172 (4.98) [–definite] .138 (3.36) [–specific] .068 (1.81)	[+plural] .370 (43.08) [+mass] .245 (11.36) [–definite] .119 (3.00) [+specific] .058 (1.70)
Upper Inter.	[+definite] .315 (17.98)	[–plural] .455 (158.25) [–mass] .154 (5.15) [–definite] .165 (5.23) [–specific] .061 (1.84)	[+plural] .374 (33.50) [+mass] .264 (11.23) [–definite] .212 (6.14)
Lower Adv.	[+definite] .445 (63.85)	[–plural] * [–mass] .209 (6.57) [–definite] .228 (7.82) [–specific] .102 (2.49)	[+plural] .399 (39.16) [+mass] .335 (20.59) [–definite] .303 (12.98) [+specific] .066 (1.72)
Native	[+definite] .554 (148.07)	[–plural] * [–mass] .405 (91.48) [–definite] .431 (124.67) [–specific] .047 (1.67)	[+plural] .468 (129.21) [+mass] .539 (311.56) [–definite] .447 (77.12)

* This is the strongest variable, as *a/an* is never used with plural nouns.

and [+plural] for *ø*. It indicates that the learners relied on the nominal property [±plural] to choose *a/an* and *ø* rather than semantic context [–definite]. Furthermore, it can be assumed that the smaller the factor contribution is, the easier it is to be overridden by others. For instance, the intermediate learners showed [+definite] as a dominant feature for *the*, but it is not a factor as strong as [–plural] for *a* or [+plural] for *ø*. Therefore, *the* may be interfered with by *a/an* when nouns are definite singular, or by *ø* when definite plural. The connection between articles and the relevant factors are then assumed to become stronger along with increased L2 knowledge and experience. Consequently, for example, the feature [+definite] of context gets more valued for the use of *the* and starts to compete with the other choices. This is presumably how the use of *a* and *ø* may retreat from definite contexts over the time. In fact, the use of indefinite articles with definite singular nouns decreased (LI 50.00% → UI 33.33% → LA 20.45%). Although some decrease was observed in the use of *ø* with plural nouns, they were still supplied with zero articles in definite contexts at comparably high rates throughout all proficiency levels (LI 42.71%, UI 36.44%, LA 31.82%). Therefore, the difficulty using *the* with plural nouns may stem from other unknown factors.

As a matter of course, there were correct uses of *the* with singular and plural nouns. Even though the contributing power of [+definite] is relatively weaker in the intermediate groups, learners' focus on the semantic contrast between [+definite] and [–definite] may let *the* win over *a* and *ø*. The concentration on contextual information may be hindered by com-

plex judgments of such as whether nouns are countable or not, and singular or not; therefore, L2 learners may probably experience relative easiness to supply definite articles to typical count or mass nouns, whose countability is clearer than flexible nouns. Although such speculation was not empirically supported by the statistical analysis in this study, it may be interesting to clarify it in a carefully designed experiment which controls the burden of interpreting semantic contexts and nominal properties simultaneously.

Moreover, the indefinite and zero articles differ within each group in terms of the strength of [–definite] in relation to other nominal properties (i.e., [±plural], [±mass]). In the LI group, the strengths of [–definite] for both *a* and *ø* are not hugely distinctive, and they are smaller than the effects of nominal properties. The UI group relates [–definite] more closely with the zero article than with the indefinite article, and the gap becomes even greater in the LA group. Within the factors specified for the zero article, however, [–definite] is still smaller than nominal factors of [+plural] and [+mass] even at an advanced proficiency level. This finding is in line with the previous study by White (2009), which noted more reliance on noun type than on context for the choice of zero articles by advanced learners. As to the features specified for the indefinite article, by contrast, White (2009) found that semantic contexts (i.e., combinations of [±definite] and [±specific]) were more influential than noun type, although his study did not include plural nouns. The present study similarly observed that a semantic context [–definite] contributed to the choice of *a/an* more strongly than a nominal property [–mass] in the advanced group, let alone the non-plural form to be the most significant factor.

Lastly, the learners showed a gradual increase in the strength of [±mass] in the use of indefinite (.172 → .154 → .209) and zero articles (.245 → .264 → .335). The findings suggest that they incorporate count-mass information into the choice of articles in indefinite contexts, resulting in singular count and mass nouns to be correctly distinguished in article use at a later stage in L2 acquisition. However, the development of [±mass] is still minor when compared to the increase in the contribution of semantic context [–definite] for indefinite (.138 → .165 → .228) and zero articles (.119 → .212 → .303). This comparison may ultimately imply that L2 acquisition of the count-mass distinction may pose more prolonged and complicated learning tasks than that of definiteness.

6. Conclusions

The analyses of the effects of nominal properties and semantic contexts in this paper have empirically revealed an intertwined picture of several factors underlying the variability in L2 article use. More specifically, the study has demonstrated L2 learners' initial concentration on nominal properties and increased attention to semantic contexts at late phases in the L2 article acquisition, showing that learners with low proficiency draw on noun types when they

choose the indefinite and zero articles, whereas advanced learners focus more on semantic contexts. It has verified that the fluctuation between definiteness and specificity interacts with countability and number of the noun in context. Furthermore, contextual information of definiteness is gradually considered important as a distinction between the definite and indefinite articles; however, the zero article remains to be closely linked to nominal properties instead of indefiniteness of context. The findings of this research in conclusion proposes that L2 learners may experience more persistent problem with the count-mass distinction than with definiteness in the acquisition of English.

Notes

1) Linguistic features are conventionally indicated in the form of [± *feature*]. For instance, *the man* is described as bearing [+definite] [–mass] [+singular] features, *men* as [–definite] [–mass] [–singular], and *water* as [–definite] [+mass].

2) There are other factors which appeared to be significant for the LA (flexibility, mass) and Native groups (mass); however, they are excluded here because they seemed to have been caused by some item-specific errors.

Acknowledgements

References

Garcia Mayo, M. d. P. (2009). Article choice in L2 English by Spanish speakers. In M. d. P. Garcia-Mayo & R. Hawkins (Eds.), *Second language acquisition of articles: empirical findings and theoretical implications* (Vol. 49, pp. 13-35): John Benjamins Publishing Company.

Hawkins, R., Al-Eid, S., Almahboob, I., Athanasopoulos, P., Chaengchenkit, R., Hu, J., Rezai, M., Jaensch, C., Jeon, Y., Jiang, A., Leung, I., Matsunaga, K., Ortega, M., Sarko, G., Snape, N. and Velasco-Zarate, K. (2006). Accounting for English article interpretation by L2 speakers. *EUROSLA Yearbook, 6*(1), 7-25. doi: 10.1075/eurosla.6.04haw

Hiki, M. (1990). The judgment of noun countability by Japanese college students : Where is the difficulty? *JACET bulletin*(21), 19-38.

Huebner, T. (1985). System and variability in interlanguage syntax. *Language Learning, 35*(2), 141-163. doi: 10.1111/j.1467-1770.1985.tb01022.x

Ionin, T., Ko, H., & Wexler, K. (2003). *Specificity as a grammatical notion: Evidence from L2-English article use.* Paper presented at the WCCFL 22 Proceedings.

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

Kim, L., & Lakshmanan, U. (2009). The processing role of the Article Choice Parameter. In M. d. P. Garcia-Mayo & R. Hawkins (Eds.), *Second language acquisition of articles: empirical findings and theoretical implications* (Vol. 49, pp. 87-114): John Benjamins Publishing Company.

Mosel, U., & Hovdhaugen, E. (1992). *Samoan Reference Grammar*: Scandinavian University Press Oslo.

Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning, 37*(3), 361-384. doi: 10.1111/j.1467-1770.1987.tb00576.x

Sarko, G. (2009). L2 English article production by Arabic and French speakers. In M. d. P. Garcia-Mayo & R. Hawkins (Eds.), *Second language acquisition of articles: empirical findings and theoretical implications* (Vol. 49, pp. 37-66): John Benjamins Publishing Company.

Snape, N. (2005). The use of articles in L2 English by Japanese and Spanish learners. *Essex Graduate Student Papers in Language and Linguistics, 7*, 1-23.

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. (2009). Exploring Mandarin Chinese speakers' L2 article use. In N. Snape, Y. I. Leung & M. S. Smith (Eds.), *Representational deficits in SLA: Studies in honor of Roger Hawkins* (Vol. 47, pp. 27-51).

Taler, V., & Jarema, G. (2007). Lexical access in younger and older adults: The case of the mass/count distinction. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 61*(1), 21-34. doi: 10.1037/cjep2007003

Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics, 10*(3), 335-355. doi: 10.1017/S0142716400008663

Trenkic, D. (2002). Formmeaning connections in the acquisition of English articles. *EUROSLA Yearbook, 2*(1), 115-133. doi: 10.1075/eurosla.2.09tre

Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition, 11*(1), 1-18. doi: 10.1017/s1366728907003185

Tryzna, M. (2009). Questioning the validity of the Article Choice Parameter and the Fluctuation Hypothesis: Evidence from L2 English article use by L1 Polish and L1 Mandarin Chinese speakers. In M. d. P. Garcia-Mayo & R. Hawkins (Eds.), *Second language acquisition of articles: empirical findings and theoretical implications* (Vol. 49, pp. 67-86): John Benjamins Publishing Company.

UCLES (2001). Oxford Quick Placement Test. Oxford University Press.

White, B. (2009). Accounting for L2-English learners' article choices. *MSU Working Papers in Second Language Studies, 1*(1), 14-37.

Yokokawa, H. (2009). *Database for L2 Research & Pedagogy: English Word Familiarity for Japanese EFL Learners -Sounds-*: Kuroshio Publisher.

Yoon, K. K. (1993). Challenging prototype descriptions: Perception of noun countability and indefinite vs. zero article use. *International Review of Applied Linguistics in Language Teaching (IRAL), 31*(4), 269-289. doi: 10.1515/iral.1993.31.4.269

Young, R. (1996). Form-function relations in articles in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (Vol. 10, pp. 135-175): John Benjamins Publishing Company.