

A Learner Corpus investigation of Filipino L2 English article use: The way forward for language teachers

Dr. Peter Crosthwaite and Miss Lavigne L.Y. Choy

Centre for Applied English Studies, University of Hong Kong¹

Bio: Crosthwaite

Peter's areas of research and supervisory expertise include second language acquisition, (learner) corpus analysis, language assessment, EFL materials preparation, Korean, Mandarin and South East Asian linguistics, and language teacher education. His current research involves the construction, annotation and analyses of learner corpora, second language acquisition of the English article system by Asian learners of English, the teaching of referential coherence in L2 academic writing, and using quantitative and think-aloud data to improve intra-rater reliability for criterion-based assessment.

Bio: Choy

Lavigne obtained her bachelor's degree from the University of Hong Kong. She is currently the Research Assistant of Dr. Peter Crosthwaite. Her research focuses on corpus linguistics such as the use of metadiscourse in academic writing of tertiary students as well as error analysis.

Abstract

It is well-documented that L2 English learners from article-less languages, such as Mandarin and Korean, have trouble appropriately marking definiteness/specificity in L2 English, causing numerous breakdowns in referential coherence. The status of Filipino / Tagalog as an article or article-less language is up for debate, with ANG, NG, SA markers used to signal definiteness, indefiniteness and genericity according to context. This paper presents a learner corpus analysis of L2 English definite article use by L1 Tagalog speakers, collected from the International Corpus of Learner English (ICNALE, Ishikawa, 2013), totalling 24,253 words from 94 essays. Using Pica's (1983) Target Language Use as a measure of article accuracy across the production of zero, indefinite and definite articles across four types of obligatory contexts (generic, specific definite, specific indefinite and non-specific indefinite), the results show that Filipino L2 English users struggle with the use of the definite article in generic contexts, much like the findings reported for Mandarin, Korean and Thai in the author's and others' previous research. Based on the results of this analysis, a number of suggestions for pedagogy are presented, looking at how Filipino learners use articles successfully, when they do not, and what can be done to improve Filipino L2 English learners' accuracy and range of definite article use.

Keywords: L2 article acquisition, learner corpora, Tagalog, Filipino, ICNALE

¹ **Peter Crosthwaite / Lavigne LY Choy**

Centre for Applied English Studies

University of Hong Kong

Room 6.38 Run Run Shaw Tower, Hong Kong SAR

E-mail: drprc80@hku.hk / lavignec2011@gmail.com

Introduction

It has been well attested that the English article system is one of the most difficult features of the language for L2 learners (Master, 1987; Thomas, 1989; Young, 1996; Robertson, 2000; Ionin, Ko & Wexler, 2004, Chuang & Nesi, 2006; Ekiert, 2004, 2007, 2010; Diez-Bedmar & Papp, 2008, Ionin, Baek, Kim, Ko & Wexler, 2012; Snape, Leung & Ting, 2006; Snape, García-Mayo and Gürel, 2013; Crosthwaite, 2013, 2014a, 2014b, 2016a, 2016b; Diez-Bedmar, 2015). English indefinite and definite articles have a variety of associated functions, with Ekiert (2007) describing the article system as ‘a complex set of abstract distinctions which are, to some extent, arbitrarily mapped onto surface forms’ (p.1). Their associated usages lie at the heart of the syntax/semantic/pragmatic interface, leading to problems of second language (L2) learnability (Sorace, 2011). Despite suggestions in second language acquisition literature that frequency is an essential factor in L2 learning (e.g. Ellis, 2010; Filipović & Hawkins, 2013) the high frequency of article use leads to a constant decision-making process on the part of the L2 learner during L2 production (Master, 2002; Ekiert, 2004, Świątek, 2013). In the case where form and function of definiteness between L1 and L2 is dissimilar, the opportunity for positive transfer is thus reduced, with L1s with an article or article-like system ([+ART] languages) finding article acquisition easier than those who come from L1s without an article or article-like system ([-ART] languages) (Master; 1987; Diez-Bedmar & Papp, 2008; Chrabaszcz & Jiang, 2014; Crosthwaite, 2016a), L2 learners from article-less languages frequently incorrectly encode definiteness and/or specificity of reference with the article system, overusing indefinite articles where definite articles are expected (Leung, 2001; Ionin, Ko & Wexler, 2004), overusing definite articles where indefinite articles are expected (Master, 1987; Young, 1996), or overusing indefinite/definite articles where zero articles are expected (Barrett & Chen, 2011, Author, 2016a)

In a recent study, AUTHOR (2016a) looked at the production of English articles by L2 English learners from Mandarin Chinese, Korean and Thai backgrounds (all considered article-less languages in the literature) at four L2 proficiency levels, using data sourced from the International Corpus Network of Asian Learners of English (ICNALE, Isihkawa, 2011, 2013). Speakers from all three L1 backgrounds experienced difficulties with the overuse of the definite article in generic contexts (e.g. *the students usually have part time jobs in college), although

L1-group-specific differences were found in the overall accuracy and accuracy order (definite, indefinite, zero) of L2 article use. Cross-learner group variation in accuracy was considered to be dependent on the relative grammaticalisation of an ‘article-like’ system of demonstrative and numeral + classifier NPs to locally encode definiteness, with Mandarin Chinese claimed to be the closest to an article-like system in this respect. Mandarin L2 English learners thus enjoyed significantly better L2 article accuracy than Korean and Thai L2 English learners, who had no such opportunity for positive transfer from their L1.

With the above finding in mind, the goal of this paper is to consider L2 English article production by students from another Asian L1 context, namely Filipino L2 English learners. The choice to study this particular L1 group is because the status of the markers *ANG*, *NG* and *SA* in Tagalog/Filipino as equivalent to the English definite/indefinite article system is up for debate, with some researchers suggesting that Tagalog is an article-less language, with others claiming that Tagalog/Filipino is much closer to an article language than other Asian L1s. In addition, English is an official second language of the Philippines (although it is only spoken by just under 50% of the population) (Bolton, 2008) while Tagalog/Filipino, despite making up 90% of words in the national language (Filipino) is only spoken by 1/3 of the population, given that there are 180 languages spoken in the Philippines – creating a ‘linguistic power struggle’ (Wa-Mbaleka, 2014, p.17) in the Philippines which is likely to have a major impact on L2 English production. Thus, this paper presents a learner corpus analysis of L2 article use among L2 English learners from L1 Tagalog backgrounds at the four L2 proficiency levels specified in the ICNALE (A2 Waystage, B1-1 Threshold Lower, B1-2 Threshold Upper and B2+ Vantage or Higher), in order to answer the following research questions:

- 1) How do L2 English learners from L1 Tagalog/Filipino backgrounds manage the production of English articles in the L2 at the four proficiency levels specified in the ICNALE?
- 2) What is the accuracy order of L2 English article production by L2 English learners from L1 Tagalog backgrounds?
- 3) What are some possible implications of accuracy order and proficiency effects on L2 article production for English language teaching in the Philippines?

Definiteness in English

While definiteness is considered a linguistic universal, the encoding of definiteness varies across different languages. English is typically considered as the stereotypical ‘article’ language in terms of the article/article-less language divide, in that (in)definiteness is clearly marked over a range of definiteness contexts. Language such as Mandarin, Korean and Thai are labelled as ‘article-less’, in that while (in)definiteness may be marked via demonstratives or numerals, it is typically less marked than in English, leaving the task of understanding definiteness to the listener/reader.

As mentioned in the introduction, the status of Tagalog/Filipino as an article-less language is up for debate, with some suggesting that the status of certain grammatical markers share many similarities with those of English form/function mappings for definiteness encoding. However, recently, other researchers have noted subtle differences between Tagalog/Filipino and English form/function mappings for definiteness, which as seen in studies on other article- and article-less languages, may cause L2 English learners from particular L1 contexts to have difficulties with L2 articles. .

With this in mind, it is necessary to determine a universal method for the encoding of definiteness that covers all potential obligatory (and by extension, non-obligatory) article use contexts, so that one can compare how speakers of different languages encode definiteness across each context and to then make generalisations or predictions regarding L2 learnability of form and function. Previous accounts of definiteness in English such as Hawkins (1978) suggests a variety of syntax/pragmatic functions of definite article use, including visible situation (pass me *the water*), immediate situation (don’t go in there, *the dog* may bite you), associative (*the book* *the author*), and inclusive functions (bring (all of?) *the wickets* in after a game of cricket). Later approaches to definiteness marking are presented along a gradient of usage dependent on a) the familiarity (Prince, 1981), b) identifiability (Gundel, Hedberg & Zacharski, 1993) or c) accessibility (Ariel, 1991, 2008, 2010) of a given referent as a discourse unfolds. However, Hawkin's categories are related only to English article use and may not be consistent across all L1s, and Crosthwaite (2014a) and Hendriks (2003) have shown that L2 learners often struggle to produce texts that allow for a comparative givenness/accessibility-based account of form/function mappings for definiteness.

With this in mind, the approach to the encoding of definiteness taken in this study follows Ekiert (2004), Diez-Bedmar and Papp (2008), Winward (2012, 2014) amongst others in using Bickerton (1981) and Huebner's (1983, 1985) semantic / pragmatic approach to article use, dependent on whether the nominal element is being specifically referred to [+/-SR] or is known to the hearer [+/-HK] as highlighted in Table 1:

Table 1: Four contexts for article use in Bickerton (1981), Huebner (1983, 1985), as shown with examples from Ekiert (2004), Thomas (1989) and Goto Butler (2002)².

<p>Context 1 - Generics [-SR, +HK]³ Ø Fruit flourishes in the valley Ø Elephants have trunks The Grenomian is an excitable person They say the elephant never forgets A paper clip comes in handy An elephant never forgets</p>	<p>Context 2 - Referential definites [+SR, +HK] Pass me the pen The idea of coming to the UK was... I found a book. The book was... The first person to walk on the moon...</p>
<p>Context 3 - Referential indefinites, first mentions [+SR, -HK] Chris approached me carrying a dog I've bought a new car A man phoned I keep sending Ø messages to him I've got Ø friends in the UK I've managed to find Ø work</p>	<p>Context 4 - Non-referentials – Attributive indefinites, non-specific indefinites [-SR, -HK] Alice is an accountant I need a new car I guess I should buy a new car A man is in the ladies, but I haven't seen him Ø Foreigners would come up with a better solution</p>

Thus, Table 1 summarises how English articles are used to encode (in)definiteness along these four contexts. The definite, indefinite and zero article may be used in generic contexts depending on whether the target referent is singular definite ('The Grenomian'), singular indefinite ('A paper clip') or plural/mass nouns ('Elephants'/'Fruit'). Only the definite article is used in referential definite contexts, and the indefinite and zero articles may be used in non-specific referential and non-referential contexts. The next section covers how speakers of Tagalog/Filipino handle the encoding of (in)definiteness across each of the four contexts outlined

² Later work such as Leńko-Szymańska (2012) includes a fifth context of phrasal/idiomatic usage of zero, indefinite or definite articles such as 'on *the other hand*', but these uses will not be referred to again in the present study.

³ [+/-SR] reference is made to a specific referent, [+/-HK] referent is known to the hearer [+/-HK]

above, via a grammar-based contrastive analysis in the absence of comparative L1 data to the L2 data analysed in the present study.

Definiteness in Tagalog/Filipino.

In the majority of cases, Tagalog/Filipino *appears* to clearly fall on the 'article language' side of the spectrum, with the earliest attribution of the *ANG* marker as a *de-facto* definite article suggested as far back as Humboldt (1836-1839). The *NG* non-topic/instrument/goal marker and *SA* dative / benefactor / locative markers can be used to signal both indefiniteness or definiteness depending on whether the target referent is topical or non-topical as with the following examples:

- (1) a. Magbibigay *ANG* babae *NG* bigas *SA* bata.
AF-will give *ANG* woman *NG* rice *SA* boy
'The woman will give rice to a/the boy.'
- b. Ibibigay *NG* babae *ANG* bigas *SA* bata.
GF-will give *NG* woman *ANG* rice *SA* boy
'A/The woman will give the rice to a/the boy.' (Adams & Manaster-Ramer, 1998, p.80).

In 1a, the *ANG*-marked actor (the woman) is the topic, while the *NG*-marked 'rice' is indefinite, leaving the *SA*-marked 'boy' either indefinite or definite depending on the prior context. In 1b, the *ANG*-marker is associated with the 'rice' (making this the topical referent), leaving the *NG*-marked non-topic 'woman' and the *SA*-marked 'boy' either indefinite or definite. In this respect, the *ANG*-marked topic referent is always considered definite (Schachter & Otones, 1972).

However, the *ANG* marker can also be used with non-definite generic reference:

- (2) Lubhang mapanganib *ANG* sawa [TOPIC, Ø NUM/IND], at napakalaki naman *ANG* elepante [TOPIC, Ø NUM/IND]

'A boa constrictor is a very dangerous creature, and an elephant is very cumbersome'
(Al-Malki, Majid & Omar, 2014:16)

NG-marked referents functioning as objects are unmarked for definiteness as with the storekeeper in (3) below, while *NG*-marked oblique referents in transitive clauses such as the rice (4) below are considered as necessarily indefinite (Schachter, 2013)

(3) Aalisin [NG tinder]erg. [AND bigas]abs. [SA sako.]loc.
will.take.out A storekeeper P rice sack
'A/The storekeeper will take the rice out of the sack.'

(4) Aalisan [NG tinder]erg. [NG bigas]obl. [ANG sako.]abs.
will.take.out.of A storekeeper rice P sack
'A/The storekeeper will take some rice out of the sack.' (Schachter, 2013, p.837-838).

In (1a), (1b) and (3), the reading of the locative SA marker is definite with respect to the discourse or situational context, but without such context, the (in)definiteness encoded by SA is unmarked:

(5) Mag-alis tayo NG bigas SA sako.
take.out abs.1.ipl. obl. rice loc. sack
'Let's take some rice out of a/the sack.' (Schachter, 2013, p.844).

While NG and SA can be used in specific and non-specific indefinite contexts similarly to the English indefinite article, the ANG, NG and SA markers in Tagalog can all be used in generic AND specific definite contexts where the discourse or situational context allows, unlike English where only the definite article can be used in specific definite contexts:

(6) Iniabót ng manggagamot sa sundalo ang itlóg
hand NG doctor SA soldier ANG egg
'The physician handed the egg to the soldier.' (Cortes, Milambiling & Paul, 2012:1)

From the above, it is still unclear as to whether Tagalog speakers will significantly benefit from positive transfer from L1 to L2 production. While the ANG, NG and SA markers are all used to encode various definiteness contexts, the form/function mappings of the three markers do not exactly follow those of English, leading to potential over/underuse of specific forms in non-obligatory (i.e. erroneous) contexts.

Methodology

The L2 data are drawn from the written version of the International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2011, 2013). This corpus was preferred over similar large learner corpora such as the International Corpus of Learner English (ICLE - Granger, Dagneaux, Meunier and Paquot, 2009), the Asian Corpus of English (ACE, 2014) for two main reasons, namely that the ICNALE contains L2 data from L1 Tagalog speakers spread

over four L2 proficiencies but covering only two potential task types per proficiency. These advantages will be described in more detail in the following sections.

Proficiency levels

The ICNALE's design criteria (following Ishikawa, 2011) include texts drawn at four L2 English proficiency levels for the assessment of pseudo-longitudinal development. These proficiency groupings are *claimed* in Ishikawa (2011) to be equivalent to the levels A2-B2 of the Common European Framework (CEFR, Council for Europe, 2001). However, there are a number of different measures of proficiency used in the construction of the ICNALE, with some students' proficiency measured by standardized tests such as IELTS® and TOEFL®, but with other students (who had not previously taken a standardised test) having their proficiency measured via a converted score following Nation & Beglar's (2007) Vocabulary Size Test. For this reason, the present study makes **no** assumptions regarding ICNALE proficiency distinctions and CEFR equivalency, and has replaced the ICNALE-defined distinctions of proficiency with new labels, namely 'Beginner', 'Pre-Intermediate', 'Intermediate' and 'Upper Intermediate'.

Task types

Despite issues with the designation of proficiency levels, one advantage of the ICNALE is that the corpus is composed solely of discursive texts of just two types, which is preferable for inter- and intra-group comparison. The prompts for these tasks are shown below:

'Do you agree or disagree with the following statements? Use reasons and specific details to support your answer.'

[Part-time job] 'It is important for college students to have a part-time job'

[Smoking ban]: 'Smoking should be completely banned at all the restaurants in the(country' ⁴

Each essay averages about 300 words, and the essays are largely equivalent to that of a high school essay in that there are no citation or referencing including and the works are generally of an informal, non-academic tone.

⁴ taken from <http://language.sakura.ne.jp/icnale/about.html>

Corpus Sample

The following table describes the corpus sample of the Filipino ICNALE dataset analysed:

Table 2 – Corpus Sample of Filipino ICNALE dataset analysed

Prompt/Level	Texts	Words	Mean words per composition	Obligatory contexts ⁵
Part-time job				
Pre-Int.	11	2664	242.18	317
Int.	25	6166	246.64	792
Upper-Int.	11	3092	281.09	389
(Total)	47	11922	256.63	1498
Smoking ban				
Pre-Int.	12	2792	232.67	387
Int.	24	5865	244.38	904
Upper-Int.	11	2874	261.27	450
(Total)	44	11531	246.10	1741

The ICNALE sample for Tagalog is heavily skewed towards Intermediate level, with 50 texts per task in the Tagalog subcorpora with much smaller numbers at other levels. In order to make the L2 proficiency subcorpora more equal, a maximum of 25 ICNALE texts were randomly selected from the Intermediate level dataset for each task. The number of texts for Pre-Intermediate and Upper-Intermediate level reflects the total ICNALE sample for those levels, which, while relatively small, still represent over 600 (Pre-Int.) / 700 (Upp. Int.) obligatory article contexts, and thus are still worthy of investigation. There were too few ICNALE texts at Beginner levels (2 per task) and so these have not been included in the analysis.

⁵ Includes all obligatory article context types 1-4 as shown in Table 1.

Annotation

All source texts were compiled into a searchable corpus using UAMCorpustool (O'Donnell, 2008), version 3.2. For corpus analysis to take place, it is necessary to code or 'annotate' the corpus data so that one can extract and quantify the different linguistic features one is coding for across the four L2 proficiencies. Annotation for each text followed the scheme created in Diez-Bedmar & Papp (2008) and modified in Diez-Bedmar (2015), which has been used to code for the four article contexts (generics, referential definites, referential indefinites and non-referentials) under investigation in the present study.:

Table 3 - Tagging system for correct uses of articles (Diez-Bedmar & Papp, 2015)

	Article used by the learner			
	Generics	Referential definites	Referential indefinites	Non-referentials
Definite article (DA)	1DA	2DA		
Indefinite article (AI)	1IA		3IA	4IA
Zero article (ZA)	1ZA		3ZA	4ZA

Table 4 - Tagging system for incorrect uses of articles (Diez-Bedmar, 2015)⁶

	Article used by the learner			
	Generics	Referential definites	Referential indefinites	Non-referential
Definite article	1GAIA 1GAZA	2GAIA 2GAZA		
Indefinite article	1GADA 1GAZA		3GADA 3GAZA	4GADA 4GAZA
Zero article	1GADA 1GAIA		3GADA 3GAIA	4GADA 4GAIA

⁶ GA=grammatical article, 1GAIA = grammatical article error with incorrect use of indefinite article in context 1 (generics), 1GADA = grammatical article error with incorrect use of definite article in context 1 (generics), 2GAIA, grammatical article error with incorrect use of indefinite article in context 2 (referential definites), etc.

Both tables above thus cover all of the obligatory and non-obligatory (error) uses of English articles across the four definiteness contexts outlined in the literature review. All obligatory and non-obligatory article uses in all texts were manually annotated by the researcher (a native speaker of English) and a non-native speaking research assistant with an M.A. in applied linguistics. The researcher manually double-checked all coding for accuracy. Two native English speaking raters then analysed a random sample of 50 texts for correct/incorrect codings, producing an intraclass correlation coefficient (ICC) of .836, of which a figure of greater than .740 is considered 'excellent' (Fleiss, 1981).

Target Language Use

After coding for appropriate and inappropriate use, it is possible to determine an overall measure of article accuracy. The present study follows Diez-Bedmar & Papp (2008) and Crosthwaite (2016a) in adopting Pica's (1983) measure of article accuracy in the form of *Target Language Use*, as shown in the equation below:

$$TLU = \frac{\text{No. of correct supliances in obligatory contexts}}{(\text{No. of obligatory contexts}) + (\text{No. of supliances in non-obligatory contexts})} \times 100$$

Under an approach that considers (non)obligatory contexts only, there is no need to normalise word counts across/between subcorpora, given that language users of different L1/L2 groups may use other kinds of NPs (such as demonstrative or quantitative NPs) that are, by their nature, non-obligatory contexts for articles. .

Results and Discussion

The following tables describe the Target Language Use (TLU) scores for the Filipino L2 English data for each task. Inter-level comparison was performed using Kruskal-Wallis comparison, with post-hoc pairwise comparison using Dunn's correction for adjusted *p* values, while inter-task comparison was performed using Mann-Whitney U tests. As the data is subject to non-parametric analyses, the medians and median absolute deviations are reported rather than the typical mean/standard deviations:

Table 5 – Target Language Use ratings by level and context for Filipino L2 English group [Part time Job]

Level	Texts	Generic contexts			Referential definite contexts	Referential indefinite contexts	Non-referential contexts		
		1DA	1IA	1ZA	2DA	3IA	3ZA	4IA	4ZA
Pre-Int	13	M=0 AD=0	M=.80 AD=.80	M=.83 AD=.17	M=1 AD=0	M=.83 AD=.33	M=.86 AD=.14	M=.83 AD=.07	M=1 AD=0
Int.	49	M=0 AD=0	M=1 AD=1	M=1 AD=0	M=1 AD=0	M=1 AD=.25	M=1 AD=0	M=.96 AD=.04	M=.91 AD=.09
Upp-Int.	22	M=0 AD=0	M=1 AD=1	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0

Table 6 – Target Language Use ratings by level and context for Filipino L2 English group [Smoking Ban]

Level	Texts	Generic contexts			Referential definite contexts	Referential indefinite contexts	Non-referential contexts		
		1DA	1IA	1ZA	2DA	3IA	3ZA	4IA	4ZA
Pre-Int	13	M=0 AD=0	M=1 AD=1	M=.94 AD=.06	M=1 AD=0	M=.63 AD=.36	M=.89 AD=.11	M=1 AD=0	M=1 AD=0
Int.	49	M=0 AD=0	M=1 AD=0	M=.91 AD=.09	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0
Upp-Int.	22	M=0 AD=0	M=1 AD=1	M=.89 AD=.11	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0	M=1 AD=0

From the tables, the Tagalog/Filipino L2 English learners have little difficulty with article use from Pre-Intermediate level onwards with one exception - the use of the definite article in generic contexts, which remains at 0 median TLU across each L2 proficiency and in both tasks. This suggests that these learners are either overextending the use of the zero article into definite singular generic contexts (*[the] lion is dangerous'), or providing false specific readings with inappropriate definite article use (i.e. *the students need part time jobs). Given the very low

frequency of definite singular generics in English discourse (e.g. Biber et al., 1999), the issue is likely to be the latter - namely massive over-production of definite article in generic contexts where the zero article (plural/mass NP) is appropriate. This trend was also seen in Crosthwaite (2016a) for Mandarin, Korean and Thai L2 English learners, although unlike those groups, the Filipino L2 English group in the present study do not appear to have any difficulty with the indefinite article in generic contexts. These findings suggest severe optionality of L2 definite article use in generic contexts which lasts at least until the highest L2 proficiency surveyed. However, overall article accuracy is an average 15%-20% higher in all article contexts for the Filipino L2 English group than found for Mandarin, Korean and Thai L2 English groups in Crosthwaite (2016a), which suggests that despite the problems the Filipino L2 English learners experience with definite article overuse, their L1 affords them significant opportunities for positive transfer compared to speakers of 'true' article-less languages, presumably because the ANG, NG, and SA markers function similarly (although not identically) to English indefinite/definite articles.

Mann-Whitney U tests employed on the data showed no effect of task on TLU scores for any of the article contexts ($\alpha=0.00625$, tests=8). There was also no significant effect of L2 proficiency on the pseudo-longitudinal development of any article form/context after Kruskal-Wallis tests ($\alpha=0.00625$, test=8) across both tasks or within each task.

The following table describes the median TLU ratings and orders of accuracy across all article contexts for the Filipino L2 English group, using Friedman's test ($\alpha=0.166$ for Pre-intermediate, Intermediate and Upper-Intermediate levels, tests=3, $\alpha=1$ for 'Across all levels' statistics). Where a significant Friedman's test occurred, post-hoc comparison is performed using Holm-Bonferonni correction.

Table 6 –Target Language Use ratings and orders of accuracy across all article contexts for Filipino L2 English group

Level	Definite article(DA)	Indefinite article(IA)	Zero article(ZA)	Result	Result (Friedman/Pairwise)
Pre-Intermediate	M=.72 AD=.25	M=.75 AD=.16	M=.92 AD=.07	ZA=IA=DA	$F_r(2)=1.014, p=.602$
Intermediate	M=.77 AD=.22	M=.94 AD=.05	M=.92 AD=.04	ZA=IA=DA	$F_r(2)=5.722, p=.057$
Upper-Intermediate.	M=.78 AD=.22	M=1 AD=0	M=.96 AD=.04	ZA=IA=DA	$F_r(2)=7.385, p=.025$ <i>No pairwise sig. after correction</i>
Across all levels	M=.75 AD=.25	M=.93 AD=.06	M=.93 AD=.07	ZA=IA>DA	$F_r(2)=9.022, p=.011$ (DA<ZA $t(2)=-.375, p=.039$)

The findings suggest that the overall order of accuracy for the Filipino L2 English group has zero articles as the most accurate (Median=.93, AD=.07), and definite articles as the least (Median=.75, AD=.25), following the findings of Crosthwaite (2016a) for Korean and Thai L2 English learners. Accuracy of indefinite article use is comparable to that of zero article use, and so the overall order of article accuracy follows the order ZA=IA>DA found for Mandarin L2 English learners in Diez-Bedmar & Papp (2008) and for Korean and Thai L2 English learners in Crosthwaite (2016a). Thus, despite the accurate use of the definite article in referential definite contexts, the oversuppliance of definite articles in generic plural/mass contexts reduces the overall accuracy of definite article use for these L2 English learners.

Discussion and Conclusion

The present study is the largest such study of L2 article use by Tagalog/Filipino L2 English speakers, highlighting the difficulties (and successes!) experienced by these learners as they attempt to produce grammatical articles in written discursive essays.

In terms of difficulties, the use of the definite article in generic contexts is particularly problematic, with Filipino L2 English learners overproducing the definite article where the zero article is expected for plural/mass NPs. This leads to numerous false definite readings (with inappropriate definite article use, i.e. *the students need part time jobs). However, in terms of

success, the overall accuracy of L2 article use by L1 Tagalog/Filipino speakers is substantially higher than that reported in the literature for article-less language such as Korean and Thai at equivalent L2 proficiencies. Thus, regarding the status of Tagalog/Filipino as an 'article' language and the L1 transfer effect such a label might suggest for L2 English acquisition, one might suggest from these findings two possible conclusions.

Firstly, if we take into account the potential (or lack of potential) for positive transfer effects regarding L2 English article acquisition that have been shown in previous studies on article vs. article-less languages, the status of Tagalog/Filipino as an 'article' language is not entirely clear cut. That is, one would not expect Tagalog/Filipino L2 English learners to achieve an average median definite article accuracy of 75% even at Upper-Intermediate level if their L1 configuration of articles was the same as that of English. It appears as though the L2 learners fluctuate between readings of definiteness/specificity for generic mass/plural NPs, incorrectly providing the definiteness marker where definiteness is not assumed. This, however, then appears to be more of an issue with mass/plural vs. singular distinctions on the NP rather than definiteness marking, given that the *only* area where the Tagalog/Filipino L2 English learners struggle is with definite article use in generic mass/plural contexts, but not with indefinite/zero article use in generic contexts, and with other article forms in their respective definite/indefinite/non-referential contexts.

However, the second conclusion one can make is that the overall article accuracy is *higher* than that reported for L2 English learners from article-less L1s in Author (2016). Thus, L1 Tagalog/Filipino speakers certainly have the advantage over speakers of true 'article-less' languages. Thus, it is increasingly apparent that one needs to consider the relative grammaticalisation of form/function relationships for article acquisition, considering a continuum of syntactic to pragmatic approaches to definiteness marking over the labelling of languages as either 'article' or 'article-less' if one is to make any transfer-based claims regarding relative L2 article learnability across different L1 speakers.

Pedagogical Implications

Given the above finding that Filipino L2 English learners experience problems with the inappropriate use of definite articles in generic contexts, English teachers in the region need to

devote special attention to driving improvements in this area. Snape García-Mayo and Gürel (2013) suggest that EFL instruction has neglected article use in generic reference in favour of a one-form one-function mapping of the definite article to specific definite readings. Thus, the first hurdle to overcome is to create materials that ask students to focus on form to make them aware of whether a referent is being referred to generically or specifically, and the implications that count/mass nouns have on article assignment, given that the majority of article errors occur when the writer provides a definite reading to a generic count/mass NP. Before producing a nominal element, students should ask themselves whether they are referring to ‘all’ entities in a set (i.e. [All] ‘students’), a previously mentioned entity (‘the students’ [that I mentioned already]), a single specific entity ([there was] ‘a student’ [who I will mention again]), or a description ([I was] ‘a student’), covering Bickerton’s article contexts 1-4 respectively.

Once students have mastered recognising when they are or are not referring generically and in what capacity, L2 learners should continue to associate the definite article with definite referential contexts only, given that the TLUs for this context is almost at 100% for these Tagalog/Filipino speakers. One way that this can be stressed is via frequency of input, given that the definite singular generic is very rare in English production (Biber et al, 1999). However, outside of input considerations, the quickest and easier improvements can be achieved by converting any singular generic reading (definite or indefinite) to a bare plural/mass generic, i.e. switching from ‘*the potato* was first cultivated in South America’ / ‘It is usual for *a person* from Italy to drink wine with his/her meal’) to ‘*potatoes* were first cultivated in South America’ / ‘It is usual for *people* from Italy to drink wine with their meals’ (Snape, García-Mayo and Gürel, 2013). This would then leave the use of the definite article reserved for referential definite contexts, and the indefinite article reserved for referential indefinite / non-referential contexts, potentially improving L2 learnability of all three article forms across their respective contexts. Given the frustration the majority of EFL teachers experience when trying to teach the English article system (and the even greater frustration experienced by EFL students when trying to learn it), teachers can do themselves and their students a great service by minimizing the variety of functions associated to individual article forms, at least in the initial and intermediate stages of L2 acquisition. By recognising that appropriate article use in generic contexts appears to be a criterial feature of advanced L2 proficiency levels (and apparently regardless of whether the learners’ L1 is a ‘article’ or ‘article-less’ language), teachers should also be able to tailor their

assessment of student performance in article use, and not penalise students who make inappropriacies in generic article use too heavily. The value of the present study is in increasing awareness of the difficulties encountered by Tagalog/Filipino L2 English learners during the acquisition of the English article system, and this, hopefully, should lead to tailored materials design and pedagogy, and ultimately increased accuracy in article production by learners from this region.

References

- ACE. (2014). *The Asian Corpus of English*. Director: Andy Kirkpatrick; Researchers: Wang Lixun, John Patkin, Sophiann Subhan. <http://corpus.ied.edu.hk/ace/> (12/02/16)
- Adams, K.L. & Manaster-Ramer, A. (1988). Some questions of topic/focus choice in Tagalog. *Oceanic Linguistics*, 27(1/2): 79-101.
- Al-Malki, E.A., Majid, N.A. & Omar, N.A.M. (2014). Generic reference in English, Arabic and Malay: A cross-linguistic typology and comparison. *English Language Teaching*, 7(11): 15-27.
- Andersen, R. (1990). Models, processes, principles and strategies: Second language acquisition inside and outside of the classroom. In Van Patten, B. & Lee, J. (Eds.) *Second Language Acquisition – Foreign Language Learning* (pp.45-68). Clevedon: Multilingual Matters.
- Barrett, N.E. & Chen, L. (2011). English article errors in Taiwanese college students' EFL writing. *Computational Linguistics and Chinese Language Processing*, 16 (3-4): 1-20.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor MI: Karoma Press.
- Bolton, K (2008). English in Asia, Asian Englishes, and the issue of proficiency. *English Today*, 24: 3-12 doi:10.1017/S026607840800014X
- Chrabaszcz, A., & Jiang, N. (2014). The role of the native language in the use of the English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research*, 30(3): 351-379.
- Cortes, K., Milambiling, L. & Paul, I. (2012). On Tagalog determiners. *Proceedings of the 2012 Annual Conference of the Canadian Linguistic Association*: 1-12.
- Crosthwaite, P. (2013). An error analysis of L2 English discourse reference through learner corpora analysis. *Linguistic Research*, 30(2): 163-193.
- Crosthwaite, P. (2014a) Differences between the coherence of Mandarin and Korean L2 English learner production and English native speakers: An empirical study. Unpublished doctoral dissertation, University of Cambridge.
- Crosthwaite, P. (2014b) Definite discourse-new reference in L1 and L2: A study of bridging in Mandarin, Korean and English. *Language Learning*, 64 (3): 456-492.
- Crosthwaite, P. (2016a). L2 English article use by speakers of article-less languages: A learner corpus study. *International Journal of Learner Corpus Research*, 2(1), in press.
- Crosthwaite, P. (2016b). Definite article bridging relations in L2: A learner corpus study. *Corpus Linguistics and Linguistic Theory*, 12(1), in press.
- Diez-Bedmar, M. B. (2015). Article use and criterial features in Spanish EFL writing. In Callies, M. & Gotz, S. *Learner Corpora in Language Testing and Assessment* (pp.163-190). Amsterdam: John Benjamins.
- 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* (pp.147-175). New York: Cambridge University Press.
- Ekiert, M. (2004). Acquisition of the English article system by speakers of Polish in ESL and EFL settings. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 4(1): 1–23.
- Ekiert, M. (2007). The acquisition of grammatical marking of indefiniteness with the indefinite article 'A' in L2 English. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 7(1).
- Ekiert, M. (2010). Linguistic effects on thinking for writing: The case of articles in L2 English. In Z. Han & T. Cadierno (Eds.), *Linguistic Relativity in SLA: Thinking for Speaking* (pp. 125–153). Bristol: Multilingual Matters.

- Ellis, N.C. (2010). Constructing a second language: Learning as a function of frequency, frequency distribution, form and function. In M. Pütz & L. Scola (Eds.) *Cognitive Processes in Second Language Acquisition: Inside the Learner's Mind* (pp.27-48). John Benjamins.
- Filipović, L., & Hawkins, J. A. (2013). Multiple factors in second language acquisition: The CASP model. *Linguistics*, 51(1), 145–176.
- Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions, 2nd Edition*. New York: Wiley.
- Goto Butler, Y. (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): 451-480.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). The international corpus of learner English. Version 2. Handbook and CD-ROM.
- 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.
- Humboldt, W. von. (1836-1839). *Über die Kawi-Sprache auf der Insel Java, nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin 1832, vol. 2. Berlin, Druckerei der Königlichen Akademie der Wissenschaften. 3 vols.
- Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2012). That's not so different from the: definite and demonstrative descriptions in second language acquisition. *Second Language Research*, 28: 69–101.
- Ionin, T., Ko, H. & Wexler, K. (2004). Article semantics in L2-acquisition: The role of specificity. *Language Acquisition*, 12(1): 3–69.
- 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* (pp.3-11). Glasgow, UK: University of Strathclyde Press.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In Ishikawa, S. (ed.), *Learner Corpus Studies in Asia and the World*, 191-118.
- Leńko-Szymańska, A. (2012). The role of conventionalised language in the acquisition and use of articles by EFL learners. In Tono, Y., Kawaguchi, Y. & Minegishi, M. (eds) *Developmental and Crosslinguistic Perspectives in Learner Corpus Research* (pp.83-103). Amsterdam: John Benjamins.
- Leung, Y. I. (2001). The initial state of L3A: Full transfer and failed features? In X. Bonch-Bruевич, W. Crawford, J. Hellerman, C. Higgins, & H. Nguyen (Eds.), *The Past, Present, and Future of Second Language Research: Selected Proceedings of the 2000 Second Language Research Forum* (pp. 55–75). Somerville, M.A.: Cascadilla Press.
- Master, P. (1987). *A Cross-Linguistic Interlanguage Analysis of the Acquisition of the English Article System*. Ph.D. Thesis. University of California at Los Angeles.
- Master, P. (2002). Information structure and English article pedagogy. *System*, 30: 331-348.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- O'Donnell, M. (2008). Demonstration of the UAM CorpusTool for text and image annotation. *Proceedings of the ACL-08: HLT Demo Session (Companion Volume)*, pages 13–16, Columbus, June 2008.
- Pica, T. (1983). Methods of morpheme quantifications: their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, 6: 69-78.
- Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, 16(2): 135-172.
- Schachter, P. (2013). Tagalog. In C. Bernard (Ed.), *The World's Major Languages* (pp. 833-855). New York, NY: Routledge.
- Schachter, P., and Otones, F. (1972). *A Tagalog Reference Grammar*. Berkeley and Los Angeles, University of California Press.
- Snape, N., Leung, I. & Ting, H. C. (2006). Comparing Chinese, Japanese and Spanish Speakers in L2 English Article Acquisition: Evidence against the Fluctuation Hypothesis? In Grantham O'Brien, M., Shea, C. and Archibald, J., (eds), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*. Somerville, MA: Cascadilla Proceedings Project, 132–39.
- Snape, N., Mayo, M. D. P. G., & Gürel, A. (2013). L1 transfer in article selection for generic reference by Spanish, Turkish and Japanese L2 learners. *International Journal of English Studies*, 13(1), 1-28.
- Sorace, A. (2011). Pinning down the concept of interface in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1): 1–33.

- Świątek, A. (2013). The acquisition of the English article by Polish learners in different proficiency groups juxtaposed with a case study. In Piechurska, E. & Szymańska-Czaplak, E. (eds) *Language in Cognition and Affect. Second Language Learning and Teaching* (pp.151-170). Berlin: Springer. DOI:10.1007/978-3-642-35305-5_9
- Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics* 10: 335–355.
- Wa-Mbaleka, S. (2014). English teachers' perceptions of the mother tongue-based education policy in the Philippines. *European Journal of Research and Reflection in Educational Science*. 2(4): 17-32.
- Winward, J. (2012). The role of semantic and pragmatic factors in article production by advanced Thai learners. *Linguistics Journal*, 6(1): 104-126.
- Winward, J. (2014). The state of the article. *Linguistic Variation*, 14(1): 46-68. DOI: 10.1075/lv.14.1.03win
- Young, R. (1996). Form-function relations in articles in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 135–175). Amsterdam: John Benjamins.