

# How definite are we about articles in English? A study of L2 learners' English article interlanguage during a University Preessional English course

---

Richard Nickalls  
English for International Students Unit (EISU)  
University of Birmingham

## Abstract

This paper presents the preliminary findings of a study which investigated the changing patterns of overuse and underuse of English articles among 30 L1 Mandarin learners of English using a small error-tagged learner corpus developed during a three month University preessional course in 2010. The study was designed to examine the nature of such learners' article interlanguage and their reaction to explicit grammar teaching as a preceding step to future research into the effects of various pedagogical interventions. New insights into the exact nature of English article underuse and overuse among this learner population were gained while the effects of explicit grammar teaching were neither sustained nor greater than a control group which had had only their attention focussed upon the error.

## 1. Introduction

The difficulties with which many international students use *a/the/Ø* articles is an interesting area for research for both linguistic and pedagogical reasons. For linguists, since English learners from different backgrounds may have an L2 with or without a comparable system of article determiners, this language area is ideal for research into L1 transfer. Moreover, such highly frequent words are useful for any linguist who wishes to examine the effects of L1 upon the acquisition of an L2 using smaller corpora.

In pedagogic terms, their high frequency of use also supports my own view that more effective ways must be found of helping learners acquire greater 'Target Like' accuracy. Indeed, it was pointed out by Roger Berry (1991) that *the/a/an* together account for one in every ten words in the average academic text. Moreover, when  $\emptyset$  articles in which noun phrases take neither article or other determiner (or a proper noun takes a 'null article') are included, this choice of article can be seen to be the most frequent and confusing decision facing learners of English. For example, in the 40,000 word learner corpus presented in this paper, in which every single noun phrase was tagged for determiner choice, the learners were confronted with a choice of *a/the/Ø* articles in around 1 in every 5 words.

The perfectly accurate use of *a/the/Ø* articles will arguably never be the first priority in any academic English writing classroom since few article errors cause the reader to misunderstand the message. Yet, although teachers of English for Academic Purposes also have more confidence in their ability to teach other areas related to other language areas, vocabulary, and academic skills, a text with article errors in every sentence becomes more

difficult to read and many teachers would like to help their learners achieve greater accuracy of use. While few teachers attempt to ‘teach’ the complex and often contradictory rules contained in the pedagogic English grammars, many teachers continue to ‘mark’ article use errors in their students’ written work, perhaps hoping that acquisition will occur when students notice the English article system. Ever more conscious of their inaccuracies, the students in turn ask for help with articles, and many teachers send their students off to read further reference materials and do grammar exercises even when they have little faith in the pedagogic merits of such learning activities.

This study formed part of the preliminary stages of doctoral research into the effects of L1 differences upon the acquisition of the L2 English article system and the impacts of various pedagogical interventions on the learners’ use of English articles. In addition to comparing the learners’ ‘Target Like Use’ of articles in a small learner corpus with the use of their own English Teachers’ use in essays on the same subject, this study was designed to longitudinally analyse the same learners’ changing accuracy throughout their 3 months of study.

Previous research relating to English article description, pedagogy and learner acquisition patterns will be outlined before the research questions are presented. After presenting the participants and the final tagging framework and process chosen for the research, the study’s preliminary findings will then be compared to previous research and discussed in relation to both this field of research and the PhD’s research questions.

## 2. Literature Review

This research builds on previous research from a diverse range of fields and approaches, as discussed below.

### 2.1 Background

Although most Indo-European languages have some form of article system, their ancestors (such as Latin and Sanskrit) did not. According to some sources (Bybee, 1998) the definite article originates from the Old English demonstrative *that* (Old English did not have an article system) while the indefinite *article a/an* came from the numeral *one*. According to the OED, like most parts of English, the definite article’s use has been in a constant state of flux – in the 17<sup>th</sup> century all dates were preceded by *the* (e.g. the 1685). Mandarin Chinese does not have a comparable article system, even if it does have other markers of definiteness.

### 2.2 Pedagogic grammar approaches to prescribing article use.

Many commentators have noted an overemphasis in EFL materials given to the anaphoric referential function of the definite article (Berry, 1991, Whitman, 1974, Yoo, 2009), in which the first mention of nouns occurs with the indefinite article and the second mention requires the definite article. As Whitman (1974) pointed out, this juxtaposition of *a/an* and *the* falsely implies that article choice is a simple dichotomy, when in reality the choice of no determiner or quantifier or alternative determiners and quantifiers is a far more complex issue. However, more advanced grammars such as Quirk and Crystal’s (1985) *A*

*comprehensive grammar of the English language* provide a more comprehensive guide to article use.

### 2.2.1 Quirk and Crystal's (1985) description of the definite article for specific reference

Quirk and Crystal (1985) identified eight functions of *the* as a marker of specific reference: the immediate situation (*the roses are beautiful*), unique reference (*the sun/the moon*), anaphoric reference (second mention), cataphoric reference (post-modified noun phrases and of phrases), sporadic reference (*my sister goes to the theatre every month*), logical use with adjectives (*the same, the only, superlatives*), and reference to body parts (*the mind*). Six of these functions, forgetting *sporadic reference* and *body parts*, seem most important for academic writing. Of these six written functions of the definite article, it is cataphoric reference which evidence suggests learners will most need in academic English since, according to Biber (1999), 40% of definite articles used in academic writing have this function.

### 2.2.2 Generic use of articles

Another function of the English article system is to indicate the more general nature of the noun as a class, and this generic function is often equally served by plurals, *a/an* or *the*, as an example from Langendoen (1970, as cited in Master, 1987) below shows:

i. *An elephant never forgets*

ii. *The elephant never forgets*

iii. *Elephants never forget*

Master (ibid.) studied a corpus of the journal 'Scientific American' and reported that the  $\emptyset$  article was the most frequent generic article form (54%), followed by generic *the* (38%) and finally the generic *a/an* article (8%).

### 2.2.3 Article use governed by convention

Reflecting what Sinclair (1991) framed as the 'idiom principle' – that much of our language is framed not by 'open choice' but by the norms of convention, it should be remembered that article use often appears totally arbitrary to learners. On the one hand, many discourse markers in academic English (e.g. *the first point, on the other hand*) could be argued to fit into what Quirk and Crystal (1985) called 'logical uses' of definite article. However, it might be more difficult to explain the vast majority of discourse idioms (e.g. *on the whole, on the rise, in the main*). In their study in the 'natural order' of article acquisition (excluding generic uses), Liu and Gleason (2002) found that learners had the most problems inserting the obligatory use of *the* in texts where the definite article had been omitted in contexts of conventional, or what they termed as the 'cultural' use of the definite article.

For example, confusion is often caused by Proper nouns, which normally take 'null' article (e.g. Africa, Mont Blanc, Peugeot, Tower Bridge) but sometimes take *the* (e.g. the UK, the Alps, the Seine, the Tower of London), particularly for other [+Art] learners because these conventions are so different - the French would say '*la France*' and '*le Royaume Uni*', adding a definite article regardless of whether a country is singular, plural, and group of islands or a Kingdom. In some of the literature, the 'null' article refers to the zero article

found with singular nouns/proper but, in this paper, the  $\emptyset$  symbol is used to describe all free morphemes whether they occur with common or proper nouns.

### 2.3 Alternative linguistic frameworks to describe article use

The most successful attempts to form linguistic frameworks which combine both grammatical and pragmatic article use build on Bickerton's semantic space framework (1981). As shown in figure 1, this much used framework classifies noun phrases in terms of the discourse features of the noun contexts, namely whether the thing is construed by the user as a specific referent [ $\pm$  SR] and whether or not the thing is known [ $\pm$  HK] to the hearer. This framework has the advantage of explaining both generic and non-referential uses of articles in addition to definite and indefinite use. The main limitation as a pedagogical tool is that there is little form/meaning relationship, with the only predictive class being the 2<sup>nd</sup> [+SR, +HK] definite article, where only *the* is possible. For this reason alone, many commentators argue against using such a complex model (Berry, 1991) while others argue that it is completely inappropriate for lower level/intermediate level learner (Master 1990). A second limitation is that the original framework developed by Bickerton and later applied to a language acquisition study by Heubner (1983) is that it ignores uses of the article governed by conventional use.

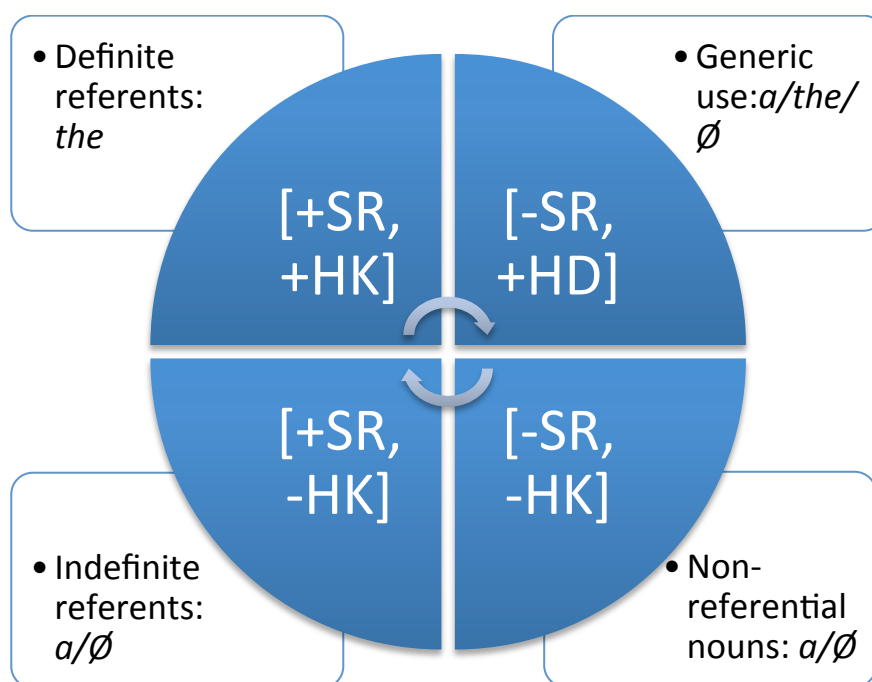


Figure 1: Bickerton's semantic use of article framework

### 2.4 Further research into Mandarin L1 learners using the Bickerton Semantic wheel

Bickerton's model has been applied to subsequent research of both a corpus-based and qualitative nature.

### 2.4.1 Corpus-based research

Diez-Bedmar and Papp (2008) used the Bickerton/Heubner framework for their corpus-based learner corpus study in which they compared Chinese, Spanish and English L1 speakers' use of the article system in English. Their Chinese learners, with 8 years of English study, had the most problems with definite article accuracy (32.6% non-target like use), followed by a 26.8% non-target like accuracy with indefinite *a/an* and 24% non-target like use of  $\emptyset$  articles. The 'Target-Like Use' measure, in which correct use is expressed as a ratio to both use and overuse, was first developed by Pica (1984: as cited in Diez-Bedmar and Papp, 2008).

### 2.4.2 Research into learner development

There is some evidence that accuracy of article use develops alongside general English proficiency (Master, 1995). However, this development is likely to be non-linear, as shown by Heubner's (1983) research which further developed Bickerton's framework and applied it to the case of a single [-Art] learner who displayed first the underuse of the definite article but then 'flooded' their language with its overuse at later stages before normalisation.

Finally, a few researchers have looked into the types of definite articles that more persistently cause English learners problems, even at advanced stages. According to Liu and Gleason (2002), problems with idiomatic use of the article – or what they called 'cultural' and 'conventional' uses – were more likely to persist for advanced learners than referential uses, as shown in figure (2).

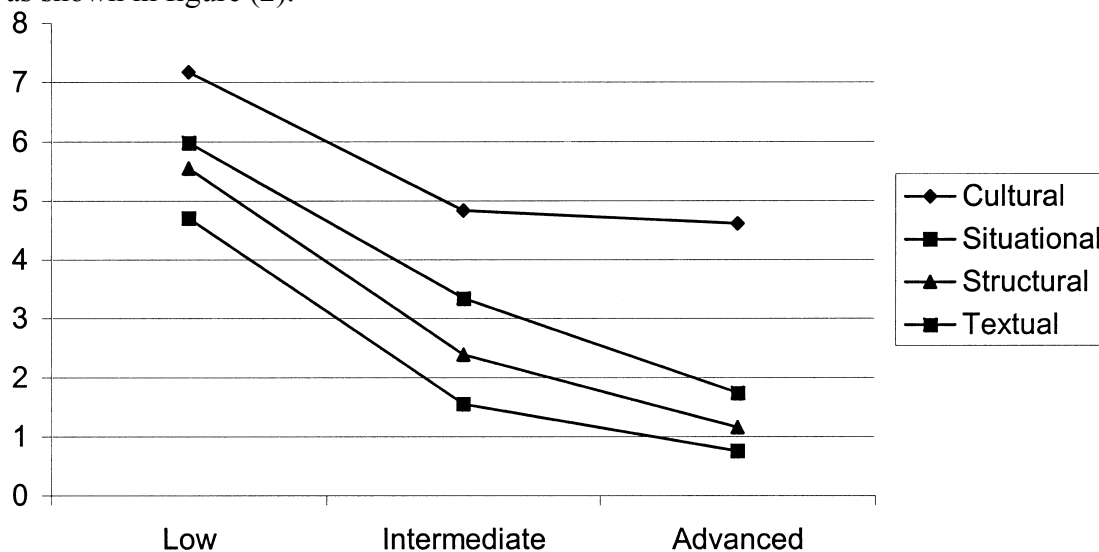


Figure 2: Comparison of means of missed the by category (Liu and Gleason, 2002: 13)

## 2.5 The effect of teaching the English article system

The emphasis given to the anaphoric use of the definite article is probably due to the perceived simplicity of teaching it to learners as a standalone use of the article system and the SLA evidence often presented to highlight the advantage of focussing grammar teaching on one defined area at a time. For example, research has shown (Bitchener and Knoch, 2009) that lower level learners make progress in their use of the definite article for anaphoric reference with focussed oral and written corrective feedback on this error. However, while their research shows that corrective feedback worked during the study, it may have had additional less desirable effects and may not have been sustained, whatever the ethical considerations of such laboratory style tests. In Berry's opinion such an overemphasis on

anaphoric reference leads to a misconception by learners. Indeed, many noun phrase referents will be mentioned the first time with a definite article/zero article (not *a/an*) and many referents mentioned first time with an indefinite can also be mentioned a second time with an indefinite.

As Berry states, teachers' efforts to simplify the article system for learners, however well meaning, can be problematic:

*the rule seems to end up in learners' awareness as 'the first time you mention a noun use a; the second time use the'. This offers a straightforward remedy for dealing with both articles, but it is quite wrong. (Berry, 1991:255)*

Peter Master can be seen as the strongest advocate of the formal grammatical teaching of articles during an era (1980s-1990s) when a focus on form was being challenged by fluency based methods and criticism of grammar teaching in general. His body of work generally suggests that teaching articles can have a measurable beneficial effect on learners' production, although he claims both lower levels and the most advanced levels benefit less than intermediate levels of proficiency. Master developed several holistic pedagogical choices of article systems and eventually (1990) took the decision to simplify the pedagogic model into a binary choice for learners: between *the* for identifying referents and *a/Ø* for classifying noun phrases. However, in his later work (1997), Master conceded that it was difficult to claim a sustained impact of teaching the English article system since attention upon the error alone may have caused a temporary improvement in article use accuracy. This problem of producing a methodology which somehow accounts for such variables is a key concern in the author's PhD.

## 2.6 Summary of previous research and its limitations

In regards to the nature of the article errors made by Mandarin speakers of English, Diez-Bedmar and Papp (2008)'s use of the Bickerton/Heubner framework has provided a more data-driven understanding than previous qualitative work of the general type of article errors which they can make in writing. Since they used essays in their corpus, rather than the closed-choice test methodology of previous researcher, their findings relate to authentic work and errors that learners really make in real life. Unlike Heubner's (1983) case-study of one learner, their more quantitative findings also on first sight allow for further replication. However, this framework did not focus on the many different types of definite article that can be taught over and above the definition of having [+] hearer knowledge and [+] specific reference. To have pedagogical insights, it was decided to build on the Bickerton/Heubner framework and add conventional grammatical classifications from Quirk and Crystal (1985). Another limitation of the Bickerton/Heubner framework was its failure to account for idiomatic use, which was again incorporated into this paper's framework.

In regards to the development of learners' article interlanguage the biggest gap in our knowledge relates to how and why learners go from intermediate to advanced proficiency in article use. Although many studies such as those by Liu and Gleason (2002), have compared advanced learners' use with intermediate levels, few studies have used quantitative methods to measure the development of the same learners over time. We therefore find quantitative but static studies juxtaposed with more qualitative data-rich studies with a longitudinal timeframe. It was therefore planned to take the essays for this study's corpus throughout a 3-month Preessional course in the hope that more insights would be gained into the developing proficiency of article use.

Finally, research into the effect of grammar teaching have found it notoriously difficult to control all necessary variables and the impact of greater attention and learner focus to a problem may be sufficient to explain the improvements that such studies have reported (Master, 1997). In addition to a control/experimental element, this study therefore also used a 2<sup>nd</sup> control group which had its students' attention focussed on article use with little explicit grammar teaching.

## 2.7 Research questions

The following five research questions resume the objectives of this study:

1. What is the 'hierarchy of difficulty' of English articles (*a/the/Ø*) among the 30 L1 Mandarin learners of English? ( will there be a replication of Diez-Bedmar and Papp's 2008 findings?)
2. Which definite articles are underused by the 30 L1 Mandarin learners of English?
3. Which definite articles are overused by the 30 L1 Mandarin learners of English?
4. To what extent will the learners' 'Target-like use' improve after explicit grammar teaching?
5. Will any such improvements be sustained and will a control group with their attention focussed upon article use make similar advances?

## 3. Participants

### 3.1 Four classes and five teachers used in the study

25 Mandarin participants (of Upper-Intermediate level) were taken from three classes doing a 15-week Summer Preessional Business English programme. Of these students, 15 were Chinese and 10 were Taiwanese, but all declared Mandarin to be their first language. 16 of these 25 Upper-Intermediate students had recently achieved a '5.5' band in IELTS tests while the remaining 9 students had achieved a writing IELTS band of '6.0'. The ages of the students ranged from 23 to 31, with an average English learning history of eight years and all participants had spent less than three months in the UK at the time of the study.

For comparative purposes, a further 5 students with an intermediate level of English and a slightly younger age (18 – 22) were taken from an undergraduate Preessional class at the same University. These five students all had lower B1 (intermediate) levels of writing ability (with IELTS band 5.00 scores). All students were Chinese (L1 Mandarin).

Finally, five teachers kindly agreed to write responses to the first essay. All teachers were L1 speakers of English with standard University English teaching certification (Masters level and Teaching Qualifications).

### 3.2 Ethical considerations

All students were given the chance to opt out of the study and they were given several days to discuss any concerns with their teacher in private before being asked to sign a consent form. As a group and as individuals, the study was perceived to be wholly beneficial for participants.

## 4. Methodology/Corpora

### 4.1 The corpora

Short (500-800 word) argumentative essays relating to business topics were taken from the students in week 1, 8 and 14 of the 15-week Preessional programme, achieving a total corpus size of 40,100 words. It had first been planned to take assessed essays (done in controlled conditions) during the course. However, it was not possible to use these assessments because an element of choice in the final assessments meant students would write different responses and it was thought essential to compare essays written on the same question. Therefore, the essays submitted were all done in homework (non-controlled) conditions.

### 4.2 The tagging framework

The Bickerton/Heubner framework, as used by Diez-Bedmar and Papp (2008), was adapted to include a further grammatical categorisation of definite article use. The data from a pilot study carried out in 2009 was analysed and the descriptions from Quirk and Crystal (1985) were chosen for this purpose. As can be seen in table 1 below, a further ‘type 5’ for idiomatic use and ‘type 6’ for the use of an alternative determiner were also added.

Table 1: Adaptation of Bickerton/Heubner framework

Type	Features	Form	Description
1	[-SR, +HK]	<i>a/an, Ø, the</i>	Generic reference
2	[+SR, +HK]	<i>the</i>	The definite article: 2.i Cataphoric reference and post-modification 2.ii Logical use with adjectives 2.iii Unique reference 2. iv Anaphoric (second mention) reference 2. v Immediate situation
3	[+SR, -HK]	<i>a/an, Ø</i>	Indefinite article
4	[-SR, -HK]	<i>a/an, Ø</i>	Non-referential nouns
5	Conventional use	<i>a/an, Ø, the</i>	Idiomatic use of articles
6	Alternative determiner	<i>All, any, some, numeric quantifier, possessive etc.,</i>	The use of an alternative determiner

### 4.3 The tagging process

A manual tagging process was necessary for this study because article use is a pragmatic in addition to grammatical feature of language. However, before the noun phrases were tagged, all noun phrases were extracted from the corpora automatically using a Java based programme developed by Dr. Oliver Mason at the University of Birmingham. This programme greatly added to the reliability of the manual tagging process. As is shown in figure 3, each sentence (#) was left intact to allow for contextualising the noun phrase during tagging. The researcher was then able to tag the noun phrases extracted below the sentence, as shown in figure 4.



# For example, *The United Kingdom and Poland both belong to EU; however, the consumers are totally different .*

*The\_DT United\_JJ Kingdom\_NN*

*Poland\_NP*

*EU\_NP*

*the\_DT consumers\_NNS*

Figure 3: example of the noun phrases extracted before manual tagging

The noun phrases in each essay were then tagged before being analysed using the Antconc concordancer.

# For example, *The United Kingdom and Poland both belong to ^EU; however, the consumers are totally different .*

<5DA> *The\_DT United\_JJ Kingdom\_NN*

<5ZA> *Poland\_NP*

<5GADA> *EU\_NP*

<2DAv> *the\_DT consumers\_NNS*

Figure 4: after manual tagging of noun phrases

#### 4.4 Teaching intervention study

In week 1 the first essay was taken from all classes to build the ‘static’ part of the corpus-based study (the four student groups and the teacher group) before any teaching of the English article system had begun (12,000 words from students, 2,000 words from teachers). Then, between weeks 4 and 8 (of the 15-week programme), one class was given five extra 30 minute ‘article use’ workshops. One pure control group was given no extra attention to articles. A further ‘control +’ group was given one 30 minute article use workshop and all their article errors were underlined between weeks 4 and 8 of the study. The inclusion of this ‘control +’ group meant that the experimental group would be compared not only to learners who had no intervention but also to learners whose ‘attention’ had been focussed on the problem just as much as the experimental group.

## 5. Findings

The analysis of this study is ongoing. However, the preliminary findings are presented below.

### 5.1 The ‘hierarchy of difficulty’ of English articles (*a/the/Ø*) among the 30 L1 Mandarin learners of English

As shown in table 2, the learners had problems with both underuse and overuse of articles. As can be seen, the 30 Mandarin speaking learners had the least problems with Ø article and then

slightly more problems with *the* and marginally greater difficulty with *a/an* – but there was no significant difference in their ability with *a/an* and *the*.

Table 2: Underuse and overuse of all articles (total corpus)

	<b>a/an</b>	<b>The</b>	<b>Ø</b>
Obligatory use	1031	2331	5089
Correct/obligatory	847	2110	4900
Supplied in non-obligatory context	18	216	373
Target Like Use	0.81	0.83	0.9

The above results differ from Diez-Bedmar and Papp's (2008) results, as compared in table 3 below. In their study, the Chinese learners had the most difficulties with *the* followed by *a/an* and then  $\emptyset$ . This paper's findings suggest a different hierarchy of difficulty - that the learners had slightly more 'target like' use of *the* – of a more or less equal target like use as *A/An*. Moreover, the analysis using Pica's 'Target Like Use' measure showed higher than average accuracy rates in all articles than those in Diez-Bedmar and Papp's (2008) study, despite the inclusion of 5 intermediate level learners' data.

Table 3: Target Like Use compared to Diez-Bedmar and Papp (2008)

	a/an	the	Ø
<b>Corpus of 30 Mandarin speaking learner essays</b> All essays and all students: N=30, 40,100 words Hierarchy: A/The> Ø	0.81	0.83	0.9
<b>Diez-Bedmar and Papp (2008)</b> N=74 L1 Upper Int-advanced 39,663 words 500 word essays Hierarchy: The >a > Ø	0.73	0.67	0.76

## 5.2 The definite articles underused by the 30 L1 Mandarin learners of English.

As table 2 shows, the definite article was omitted in 221 cases in which it was judged that an L1 speaker would use *the*. To illustrate the contexts [Type 1, 2 and 5] in which the Mandarin learners most often omitted the obligatory use of *the*, the omission rate for each type of definite article in essay 1 alone (before any teacher intervention) is illustrated in figure 5.

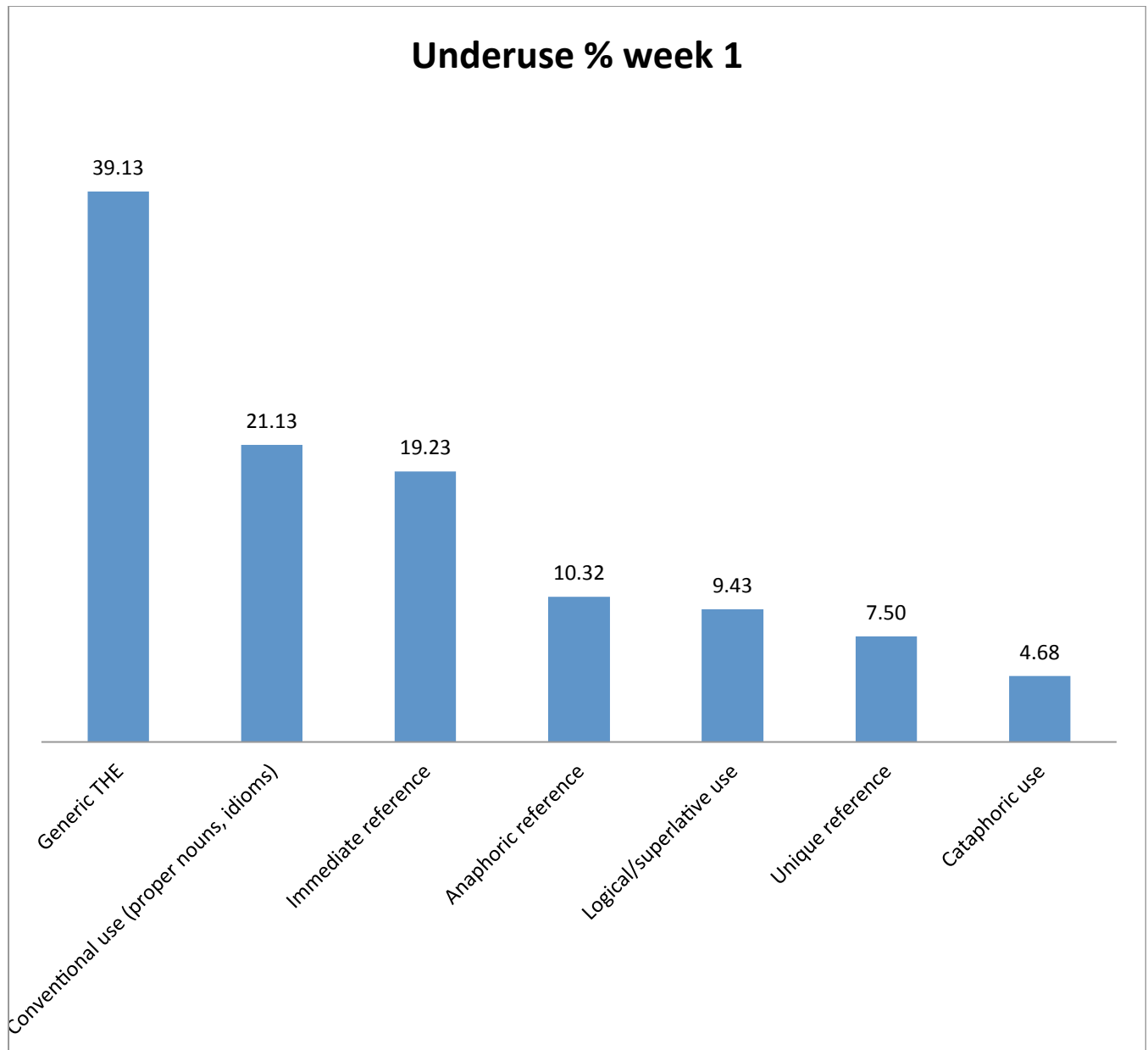


Figure 5: underuse of the definite article

As can be seen from the above chart, the learners omitted around 39% of obligatory generic definite articles and 21% of definite articles needed by convention (with idiomatic phrases or proper nouns). As is clear from this chart, Type 2 contexts for definite articles [+ Hearer Knowledge and + Specific reference] caused less of a problem for learners than Type 1 (generic contexts) and Type 5 (idiomatic contexts).

However, although learners obviously struggled with definite articles in generic contexts, their use in academic English is relatively rare, as shown in table 4 below (just under 3% for both students and teachers). It is interesting to compare the accuracy rate with each type of definite article with the frequency of use among both the learners and their teachers.

Table 4: Underuse of definite articles compared to frequency of use (first essay only)

Use of THE in 1 <sup>st</sup> essay	Student use	Teacher use	Missed when obligatory context
Generic THE	2.99%	2.98%	39.13%
Conventional use (proper nouns, idioms)	10.99%	18.63%	21.13%
Immediate reference	10.75%	19.61%	19.23%
Anaphoric reference	26.5%	2.94%	10.32%
Logical/superlative use	11.13%	5.88%	9.43%
Unique reference	7.48%	7.84%	7.5%
Cataphoric use	30.13%	42.16%	4.68%

Interestingly, the definite article that Chinese learners had the most difficulty with in the first essay was in generic contexts which were the least frequent use of definite articles among both learners and teachers. Meanwhile, the learners were most accurate – that is to say that missed the least obligatory contexts for *the* during cataphoric references – for example in ‘of phrases’ which were the most frequent reasons for using a definite article in both the learner and teacher corpora. This could in theory support the hypothesis that the students learn best what they use most, but obviously the anaphoric and conventional use inaccuracy complicate this assumption.

### 5.3 The type 2 definite articles overused by the 30 L1 Mandarin learners of English.

In all, of the 2326 Type 2 definite articles used in the total 40,100 word corpus, 216 (just over 10%) were judged to be examples of overuse. As table 5 shows, of these 216 examples of overuse, in 3% of cases it was judged that an indefinite article (Type 3) was necessary while some 5% were in Type 6 idiomatic contexts (*‘in the Africa...over the time...in the other words..’*). By far the largest overuse of the definite article (92% of overuse) was in Type 4 contexts in which it was judged that an L1 English speaker would not have used a referential noun phrase.

Table 5: overuse of definite articles (n=216)

Context	
Type 4 (non-referential )	92%
Type 3 (indefinite)	3%
Type 5 (idiomatic)	5%

### 5.3.1 Examples of non-referential (Type 4) noun phrases given a definite article

The examples below are a representative sample of the 179 noun phrases judged to be Type 4 (without specific reference or assumed hearer knowledge) and hence not requiring a definite article. This paper can only speculate on the reasons for the inaccuracies, having not interviewed the participants about their choices. However, with contact with the participants and data the researcher came to the conclusion that examples 2, 4 and 5 originated from a misunderstanding of anaphoric (first and second mention of nouns) in academic English writing. It is possible that learners may have sometimes been mistaken by the head noun in compound nouns, for example in sentences 2 and 3 (*brand loyalty* and *sun damage*). In examples 3 and 4 there are many nouns which can be used in both countable and non-countable (abstract) ways (fear, damage). In some examples (1 and 5) it might arguably be speculated that the learners are misunderstanding generic references for ‘classes’ of words and assuming that definitions use a definite article.

1. ....*the role of the human resources is not only in the manufacture but also important in the management* [M.024294-2-Wk8]
2. ....*consumers’ satisfaction with their own brand is an important driver of the brand loyalty* [M.1121930-4-Wk14]
3. ....*they also demand other functions from their shampoo such as fresh smell and.....prevention of the sun damage* [1035310-2-Wk8]
4. ....*that is the reason why the fear works. All in all I believe that fear alone can encourage people to achieve...* [M.0955237-4-essay1]
5. ....*might think about the segmenting which helps us to identify the customers....* [M.0955237-4-LE]

### 5.3.2 Further examples of the overuse of anaphoric reference

Below are shown two further examples in which the learners have used a definite article in later mentions of a noun phrase. For instance, in an essay about customer loyalty (sentence 6) there was a tendency by learners to first use term without an article and then add one as if were needed for a second mention. In another essay (sentence 7) some learners repeated one of the terms of the question (segmentation) with a definite article after using it originally without.

6. ....*companies want them to promote the loyalty*. [M\_1035310-2\_wk14]
7. ....*through the segmentation the needs of customers can be matched...*[M\_1056232-4-wk8]

As always, such speculation about the learners’ motivations for overuse are very risky without further evidence. Especially so given that in some examples the learners may have been meaning to use ‘immediate reference’. As can be seen in examples 8 and 9, the researcher was often faced with grey areas in which the definite article could arguably be being used as a conversational ‘immediate reference’ device. These examples were particularly difficult to tag since on different days a lower tolerance of genre sensitivity is inevitable. That is to say, six days a week the researcher might decide that writing about ‘the retailers’ (as if they were in immediate sight) was overly informal, but one day of the week it might appear acceptable. Of course, in this genre of academic writing such a conversational style would normally be judged to be ‘incorrect’. However, after consideration it was decided to tag noun phrases as incorrect only when they were clear cut cases that were grammatically impossible.

8. *Motivation is not the only way to incentivise people in practice, since the managers just search for the best ways in every situation [M\_0955237-4-Wk1]*
9. *Therefore, the retailers can use this information to offer....[M\_035310-2-Wk14]*

This choice to allow very marked uses of anaphoric (or immediate) reference may naturally account for the differences reported in these learners Target Like Use of articles and the Diez-Dedmar and Papp's (2008) study. It was not possible to attain any examples of their tagged data for comparisons so they may simply have been more exacting in their judgement.

What is interesting, moreover, is the marked use of this anaphoric referential function (as inferred by the researcher) among Mandarin speaking learners when compared to their teachers writing the same essay. As shown in table 6, over a quarter of contexts in which learners used a definite article were for anaphoric reference while teachers used definite articles for this function in less than 3% of contexts.

Table 6: Learners and teachers' use of anaphoric reference compared

	Students (N=30) 12,500 words	Teachers (N=5) 2,000 words
Use of THE for anaphoric reference	26.5%	2.94%

#### 5.4 Inter-reliability tagging of research findings

Naturally, the above inferences are problematic without independent verification. Unfortunately, the checking of the above findings by a second researcher has not been completed due to time constraints. It will clearly be necessary to check the reliability of a single researcher's judgements before any publication of this study.

#### 5.5 The extent to which the learners' 'Target Like Use' improved after explicit grammar teaching

As planned, three groups were taken further through an 'experimental' part of the study. At the end of the 'intervention' in week 8, in which one group was explicitly taught article use rules and a further 'control +' group had all their article errors underlined, the 2<sup>nd</sup> essay was compared with the 1<sup>st</sup> essay's article use. Furthermore, in order to investigate whether any learning was sustained, a third essay was taken in week 14 (6 weeks after the extra attention upon articles had been stopped). Although it was not able to collect the third essays from one of the original control groups (there is a gap in the data as shown in table 7) this was less of a problem given the findings reported.

As can be seen, the improvement in the 'control +' group which was focussed upon article production errors was almost identical to the progress shown in the experimental group which was most actively and explicitly taught about the English article system. Moreover, the 'learning' in both groups was not sustained – their Target Like Use of the article system

falling back to similar levels as week 1 six weeks after their attention was focussed upon it.

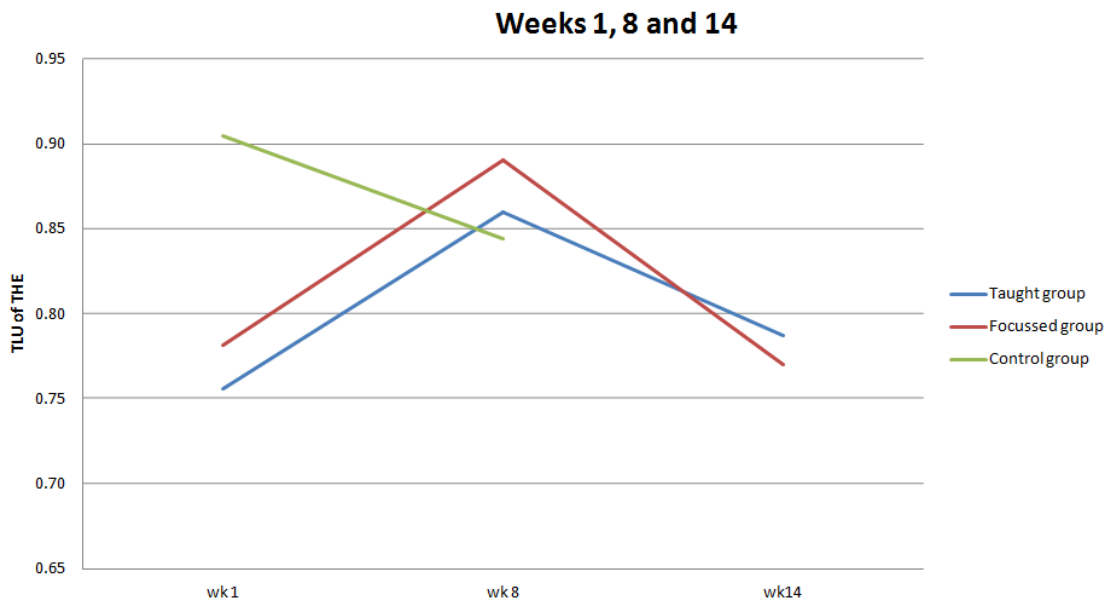


Figure 6: The three groups TLU compared longitudinally

Table 7: The TLU of three groups compared over three months

	Week 1	Week 8	Week 14
Experimental group	0.76	0.86	0.79
Control +	0.78	0.89	0.77
Control	0.91	0.84	n/a

### 5.6 What other factors may affect the learners' accuracy with English articles?

Supporting the findings by previous researchers (Master, 1997), the factor most affecting the learners' accuracy of use seemed to be their general writing ability. There was a significant (positive) relationship of .47 between the entry test writing grade and Target Like Use of THE in the 1<sup>st</sup> essay among the 30 participants (Correlation is significant at the 0.01 level, 2-tailed).

## 6. Discussion

This study attempted to replicate and develop further the method provided by Diez-Bedmar and Papp in their corpus-based study (2008). It is conceded that comparisons with this previous research will be of limited value until random verifications of the tagging reliability can be completed. With this note of caution, it is pointed out that although the findings replicated their observation that Ø article proved the least difficult for Mandarin speakers at



Upper-Intermediate level, there were some differences both in the relative difficulty the learners had with *a/the* and the higher overall accuracy displayed by the learners in the current study. As discussed, without knowing how previous researchers have tagged very marked uses of English articles (as either ‘wrong’ or ‘marked but correct’) such comparisons are very difficult to make. In sum, although it is often assumed that such quantitative learner corpus studies can be better replicated and hence more quasi-scientific, this is almost impossible without access to the tagged data itself. This researcher will certainly be happy to share samples of his tagged data to any future researchers in the hope that the current trend towards ‘open source software’ can be copied by future corpus linguistics.

Moreover, learner corpus researchers need to address the issue of how to quantify the ‘marked’ use of language features in addition to its accurate/inaccurate use since one stronger conclusion of this study is that Mandarin L1 uses overuse of the definite article is as often ‘marked but possible’ as it is ‘impossible and inaccurate’. Approximately a quarter of learners’ definite articles appeared to have an anaphoric referential function, compared to L1 teachers use at 3%, so it would seem that anaphoric reference among Mandarin learners of English is marked, whether or not it is ‘correct’.

The learners’ incorrect use of definite articles in non-referential Type 4 contexts also imply that these learners need further support in discussing topics and going from specific examples to more generalised statements about an issue if they are to stop the repetitive and marked/inaccurate use of the definite article, which will be of little surprise to most teachers of most language groups. This overuse of anaphoric references could naturally lead to several opposing conclusions, given that commentators such as Berry (1990) have already noted an overemphasis on their basic features in beginners’ grammar books. On the one hand, it might be suggested that these learners need greater instruction in the more advanced nature of anaphoric reference and the use of synonyms and abstract ideas that avoid continual repetition of a noun phrase (which need not always be given a definite article in second mention). Alternatively, it could be suggested that the earlier teaching of first mention/second mention anaphoric reference is itself the root cause of the problem – while such instruction may only achieve temporary results anyway.

In terms of underuse, it has been shown that Mandarin L1 learners of English often omit the definite article in generic and idiomatic use contexts. While generic use is needed least frequently (at least in the narrow definition applied in this study), the problems they have with proper nouns (e.g. *the UK, the Africa*) and idiomatic phrases (*in time, on one hand etc.*) suggest that this would be a fruitful subject for future research into English article pedagogy. It has already been shown that such errors remain fossilised in advanced learners’ interlanguage even at advanced levels (Liu and Gleason, 2002), so it could be speculated that this function may be a priority along with anaphoric reference.

Finally, through adding an intervention to two classes from which essays were collected, this study investigated some promising avenues for this PhD research which will soon focus on the effect of different pedagogical methods upon article underuse and overuse. The findings lead to the conclusion that the ‘learning’ in the class explicitly taught article use was neither sustained nor greater than the progress made by a class simply focussed on the error in question. This conclusion contradicts Master’s earliest hypothesis that article use can be explicitly taught. Naturally, it could be argued that the five 30-minute grammar sessions were ineffective or insufficient. However, what it definitely shows is how dangerous it is to make conclusions from experimental findings without data collection long after an intervention and the use of control groups for comparison.

## 7. Future aims

Efforts will now be made to check the tagging already completed by the use of a second researcher taking randomised samples of the 40,100 tagged corpus and 2,000 teacher corpus. Once validated, there are also further statistical measures to be applied before any publication is proposed. Future research methods need to address several issues raised in this study. For example, in future, the use of essays written in controlled conditions would certainly be preferable to the assignments done for homework with all the varying amounts of help and technological tools which may have been applied. The addition of 'marked' language in addition to 'incorrect use' should be introduced into future tagging systems. Although the full impact of grammar teaching will never be proved or disproved, such efforts to quantify the effect of simple attention upon an error should form part of any experimental research study investigating article pedagogy. In time, it is hoped that this PhD will inform several L1 groups of English learners (such as Mandarin learners of English) about both the type of English article they will find difficult and how they can achieve greater Target Like Use.

## References

- Berry, R. (1991) Re-articulating the article. *ELT Journal*, 45: (3): 252-259.
- Biber, D. (1999) *Longman grammar of spoken and written English*. Harlow: Longman.
- Biber, D., Leech, G.N. and Conrad, S. (2002) *Longman student grammar of spoken and written English*. Harlow: Longman.
- Bickerton, D. (1981) *Roots of language*. Repr. 1985. Ann Arbor, Mich.: Karoma.
- Bitchener, J. and Knoch, U. (2009) The Contribution of written corrective feedback to Language Development: a ten month investigation. *Applied Linguistics*, 31: 193-214.
- Bybee, J. (1998) "The Evolution of Grammar". Darwinian perspectives on the origins of language. Philadelphia.
- Diez-Bedmar, M.B. and Papp, S. (2008) "The use of the English Article System by Chinese and Spanish learners". In Gilquin, G.; Papp, S. & Diez-Bedmar, M.B. (Eds.) *Linking up Contrastive and Learner Corpus Research*. Amsterdam, Atlanta, Rodopi 147-175.
- Hawkins, J.A. (1984) A note on referent identifiability and co-presence. *Journal of Pragmatics*, 8: (5-6): 649-659.
- Hewson, J. (1972) *Article and Noun in English*. The Hague: Mouton.
- Huebner, T. (1983) *Longitudinal analysis of the acquisition of English*. Ann Arbor: Karoma Pub.
- Huong, N.T. (2005) PhD Thesis: Vietnamese learners mastering English articles. . University of Groningen.
- Kharm, N. (1981) Analysis of the errors committed by Arab university students in the use of the English definite/indefinite articles *IRAL-International Review of Applied Linguistics*, 19: (4): 333-345.

- Lacey, A. (1977) Rules in the Teaching of the English Articles. *ELT Journal*, XXXII: 33-37.
- Langendoen, D.T. (1970) *Essentials of English grammar*. New York: Holt, Rinehart and Winston.
- Liu, D. and Gleason, J. (2002) Acquisition of the Article by nonnative speakers of English. *Studies in Second Language Acquisition*, 24: (1): 1-26.
- Master, P. (1987) Generic the in Scientific American English for Specific Purposes, 6: (3): 165-186.
- Master, P. (1990) Teaching the English Articles as a binary system. *Tesol Quarterly*, 24: (3): 461-478.
- Master, P. (1995) "Consciousness raising and article pedagogy ". In Belcher, D. & Braine, G. (Eds.) *Academic Writing in a Second Language: Essays on Research and Pedagogy*. Norwood, N.J., Ablex Publishing Co. 183-204.
- Master, P. (1997) The English Article System: Acquisition, Function and Pedagogy. *System*, 25: (2): 215-232.
- Quirk, R. and Crystal, D. (1985) *A comprehensive grammar of the English language*. London: Longman.
- Sinclair, J. (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Whitman, R. (1974) Teaching the Article in English. *Tesol Quarterly*, 8: (3): 253-262.
- Yamada, J. and Matsuura, N. (1982) The use of the English article among Japanese Students. *RELC Journal*, 13: (1): 50-63.
- Yoo, I.W. (2009) The English Definite Article: What ESL/EFL grammars say and what corpus findings show. *Journal of English for Academic Purposes*, 8: 267-278.