

The Asian EFL Journal
Professional Teaching Articles
February 2011
Volume 50



Senior Editors:
Paul Robertson and Roger Nunn



Published by the Asian EFL Journal Press

Asian EFL Journal Press
A Division of Time Taylor International Ltd
Time Taylor College
Daen dong
Busan, Korea

<http://www.asian-efl-journal.com>

©Asian EFL Journal Press 2011

This book is in copyright. Subject to statutory exception
no reproduction of any part may take place without
the written permission of the Asian EFL Journal Press.

No unauthorized photocopying

All rights reserved. No part of this book may be reproduced, stored
in a retrieval system or transmitted in any form or by any means,
electronic, mechanical, photocopying or otherwise, without the prior
written permission of the Asian EFL Journal.
editor@asian-efl-journal.com

Publisher: Dr. Paul Robertson
Chief Editor: Dr. Roger Nunn
Journal Production Editors: Dr. Aradhna Malik

ISSN 1738-1460



Table of Contents:

1. Peter Crompton.....	4-34
- <i>Article Errors in the English Writing of Advanced L1 Arabic Learners: The Role of Transfer</i>	
2. Michael John Alroe.....	35-71
<i>Error Correction of L2 Students' Texts – Theory, Evidence and Pedagogy</i>	



Article Errors in the English Writing of Advanced L1 Arabic Learners: The Role of Transfer

Peter Crompton

American University of Sharjah

Bio Data:

Peter Crompton has been teaching at AUS since 2006. He has previously taught in universities in England, China, Saudi Arabia, Brunei, and Lithuania. He holds a PhD in Linguistics from Lancaster University. He teaches English for academic purposes and applied linguistics and his research interests are written discourse analysis, corpus linguistics, learner corpora, and academic writing.

Abstract

The problems encountered by English learners whose mother tongue does not have an article system have been researched extensively. The problems encountered by learners whose mother tongue does have an article system, such as Arabic, have been studied less. This article describes an enquiry into article system errors in a corpus of English writing by tertiary-level L1 Arabic speakers. Frequencies of articles are compared with those in native English and non-native English speaker corpora. A detailed account is given of the commonest types of errors, classified according to the mis-use of each article. It is found that the commonest errors involve mis-use of the definite article for generic reference. The strong likelihood that these errors are caused by L1 transfer, rather than an interlanguage developmental order, is argued by a comparison of the forms of generic reference in English and Arabic. It is suggested that even for learners of English with mother-tongues which have an article system, such as Arabic, L1 transfer may be a problem and as such could be usefully addressed in language instruction.

Keywords: article error, Arabic learners, generic reference, transfer, English writing

Introduction

The paper describes research into the scale and nature of article system errors made by tertiary-level L1 Arabic learners of English in a corpus of written text. It

begins with reviews of research on the role of transfer in language acquisition in general and in article acquisition in particular. It then reviews the literature on errors in article use by L1 Arabic learners, reports some corpus-based research by the author which brings together both acquisition and use perspectives, discusses the findings, and suggests some implications for future research and pedagogy.

As the paper gives considerable attention to *generic reference*, that term requires clarification. In the literature not all NPs are described as referential nor all non-generic reference as specific reference (e.g. Huebner, 1983; Thomas, 1989). For simplicity's sake, however, this paper follows the practice of Quirk, Greenbaum, Leech and Svartvik (1985) in treating all NPs as referential and using the term *specific* as a synonym for *non-generic*.

Status of transfer studies

Corpus studies show that *the* is the commonest word in English and *a* in the top five (Master, 2002) and together *the* and *a* make up 8.5% of all English text (Berry, 1993).¹ Given both the ubiquity of the article system and the comparatively long history of ELT in many Arabic speaking countries, surprisingly little has been written on English article system errors made by L1 Arabic speakers. Overall, there is less literature on errors by learners categorized on the basis of their L1 than one might expect: this can be attributed to a temporary eclipse of contrastive studies and error analysis (James, 1998; Mohammed, 2000; Odlin, 1989). In brief, the dominant view in SLA research for many years was that second and first language acquisition were essentially similar. This view, based on studies into first language acquisition (Brown, 1973), child second language acquisition (Dulay & Burt, 1974) and adult second language acquisition (White, 1977), explicitly opposed the contrastive approach (Lado, 1957) of the 1960s. As summarized by Dulay, Burt and Krashen (1982, p. 198) the “great majority of grammatical errors found in the language output of L2 learners is similar to those made by L1 learners of the target language rather than to the structure of the L2 learner’s mother tongue.” Interestingly for our purposes, the article system was one of the morphemes analyzed in these studies. Dulay and Burt diagnosed only one of the 74 article system errors in their data as “interlingual”, i.e. originating from L1; White (1977), for example, found that of the 31 article system

errors in their data none was interlingual. The proper focus of such study were errors conceived of as “intralingual” or developmental errors, evidence of a process of creative construction in a learner’s interlanguage. In the light of this, the term “error” in itself was later conceived of as unhelpful and misleading, in that what teachers registered as errors in target language production were in fact system features of learner interlanguage: to describe such features as errors was to commit the “comparative fallacy” (Bley-Vroman, 1983) of describing any language as a deficient form of another.

However, as noted in James (1998), approaches to analysis of learner language which expected and looked for errors arising from L1 influence survived, albeit rebranded slightly, as studies of cross-linguistic or language *transfer* (Odlin, 1989). As James (1998) argues, the concept of transfer now enjoys “renewed acceptance as a crucial component in modern L2 learning theories.” In a recent review of the evidence on transfer, Ellis (2006b) goes further than most in arguing for the essential correctness of the position originally taken by Lado’s (1957) contrastive analysis hypothesis; Ellis marshals evidence from a meta-analysis of SLA studies on morpheme acquisition and “associative and connectionist theories of animal and human learning” to argue that “second language learning is...faced with maximal transfer and interference from L1” (2006b, p. 186). Indeed, all the literature on article system acquisition in the last twenty years recognizes L1 transfer as a variable at least worthy of investigating (e.g. Butler, 2002; Jarvis, 2002; Master, 1990, 1997; Thomas, 1989; Zdorenko & Paradis, 2008; Zegezac, 2004).

Studies of English article acquisition

In English the article system is a marker of definiteness. The extensive literature on definiteness is usefully summarized in Lyons’ (1999) cross-linguistic study. According to Lyons, definiteness is not a semantic concept but a grammaticalization of the semantic/pragmatic concept of identifiability: a marker of definiteness, such as a definite article, “directs the hearer to the referent of the noun phrase by signaling that he is in a position to identify it” (5-6). Like other grammatical categories there is not a simple one-to-one meaning-form correspondence. Definiteness is manifested in other forms, such as pronouns and proper nouns, and diachronically the prototypical

(identifiability) meaning can be both added to and subtracted from. As Lyons notes, “some languages will require generics to be definite while others do not” (1999, p. 278). Arabic, like French, is in the former class, while English is in the latter. In the course of the research described in this paper it becomes apparent that this particular divergence appears to have serious consequences for Arabic speakers learning English.

Unfortunately, much of the language acquisition literature on articles offers little of direct interest to those investigating L1 Arabic learners’ acquisition of English, for two main reasons: (i) the broad-grained nature of their specification of the L1 from which transfer is analysed; (ii) the broad-grained view nature of their specification of the functions of the article system.

Turning to broad-grained L1 formal specification first, in general, SLA research on English article acquisition has divided L1s into a simple binary typology, languages which have an article system [+article] and those which do not [-article]. Arabic is a [+article] language but unlike English in that there are definite articles but no indefinite articles (Schulz, 2004). In general, researchers have found that subjects with [-article] L1s, such as Japanese and Russian, take longer to acquire the system than those with [+article] L1s (Chaudron and Parker, 1990; Liu & Gleason, 2002; Master, 1997; Thomas, 1989). Using a corpus-based study of the morphemes in Dulay and Burt’s (1973) “morpheme order” studies in the Japanese component of the Longman Learners Corpus McEnery, Xiao, and Tono, (2006) found that articles were the most difficult to acquire, with even proficient learners not having achieved the acquisition rate of 90% specified in Dulay and Burt’s Bilingual Syntax Measure. This differs from the order predicted by Dulay and Burt (1973), prompting McEnery et al to speculate that this is because Japanese is a [-article] language. Ionin, Ko, and Wexler (2004) and Zdorenko and Paradis (2008) use a further binary distinction, based on a Universal Grammar principles and parameters model, according to which [+article] languages may be further categorized into languages in which the article system marks the semantic category of definiteness (e.g. English), and languages in which the system marks the category of specificity (e.g. Samoan). However, their research focuses on comparative error rates for learners from [-article] and [+article] L1s rather than types of errors made by learners from [+article] L1s. Two much-cited

longitudinal studies into morpheme acquisition which included articles (Huebner, 1983; Parrish, 1987) also had subjects with East Asian [-article] L1s, limiting generalizability of findings to Arabic – unless a universal order of acquisition is assumed.

Turning to broad-grained functional analysis of article use next, the instruments used to elicit data, whether obligatory occasion or discursive tasks, are based on narratives (Chaudron & Parker, 1990; Jarvis, 2002; Liu & Gleason, 2002; Thomas, 1989) in which referents are concrete rather than abstract and specific rather than generic. Ionin, Ko and Wexler (2004) and Zdorenko and Paradis (2008) reduce options further by setting up contexts in which \emptyset is never accurate. This has helped to focus on basic semantic features of the article system, such as previous mention, of interest in the early levels of language acquisition, but reduces the applicability of their findings on overall accuracy for later levels.

In general, for the kind of errors made by intermediate to advanced learners a more fine-grained analysis of the functions of the articles is required. Rather than merely recording correct and incorrect instances of article use, some analysis of the contexts in which errors occur is required. One such study is by Liu and Gleason (2002), who note that previous studies have neglected to investigate the fact that “certain uses of [non-generic] the might be more difficult than others” (5), and find evidence in their own data for an order of acquisition for these uses. Their data was mostly from subjects with East Asian [-article] L1s – so its applicability to learners from other L1 backgrounds is an open question. A further gap in the data used in the literature, noted by Thomas (1989), is that relating to use of the articles for generic referents.

Two somewhat more fine-grained discourse-oriented studies, Robertson (2000) and Jarvis (2002), looked for systematicity in the errors learners made and found it; some of this systematicity was attributable, as expected, to discourse factors and some to linguistic context, i.e. variation in the syntactic forms of the host NP. However, both researcher also found evidence for “unsystematic variation” (Robertson, 2000, p. 135), i.e. individuals didn’t always make the same article choices in the same contexts. Jarvis attributes this to individual learners simultaneously entertaining “multiple hypotheses about article use” (Jarvis, 2002, p. 416). Robertson posits a

“remapping principle” according to which learners have to effect an adjustment from (a) how semantic and pragmatic features are mapped onto syntactic and lexical resources in L1 to (b) how this mapping is done in L2: the variation is a reflex of the difficulty of this task. Robertsons’s subjects were L1 Chinese, and Jarvis’s L1 Finnish and L1 Swedish.

Studies of Arabic ESL learners’ article errors

Observations on likely errors are provided in collections by experienced ESL teachers (Kharma & Hajjaj, 1997; Smith, 2001; Thompson-Panos & Thomas-Ruzic, 1983). More formal investigations are described in Scott and Tucker (1974), Kharma (1981) and Bataineh (2005), and a study of translator errors appears in Farghal and al-Zou’bi (2004). Scott and Tucker (1974) found article errors among the top four types of error among high school graduate L1 Arabic learners: the predominant error was omission of *a*. Kharma (1981) looked at tertiary level students’ errors in a cloze test and found a mean error rate of over 25% for all articles (*a*, *the* and \emptyset), highest for *a* and lowest for *the*. However, in production data – a corpus of essays – Kharma (1981) reports half of all errors being in the use of *the*. Bataineh (2005) also looked at errors associated with *a* in a corpus of tertiary essays. Focusing only on the 319 syntactic/semantic errors in her data (i.e. leaving aside spelling and orthographic errors) and collapsing some of Bataineh’s error categories, 61% of errors were instances of under-use:- “ \emptyset -for-*a*” (46%), “*the*-for-*a*” (15%). Overuse counted for the remainder with the largest types of error being “*a*-for-*the*” (27%) and “*a*-for- \emptyset ” (8%).

These studies do not all classify errors in the same way or count the same things; Bataineh’s (2005) only looks at *a*, for example. Unfortunately Scott and Tucker (1974) and Bataineh (2005) do not tally accurate use instances so an overall success or failure rate for each article cannot be calculated. However, under-use of *a* in obligatory contexts is common in all three: for both the studies which considered all three articles, Scott and Tucker, and Kharma, this is the single most common type of article error. In Scott and Tucker’s (1974) and Bataineh’s (2005) data, “ \emptyset -for-*a*” is the more frequent error subtype; in Kharma’s data “*the*-for-*a*” is slightly more frequent.

Kharma's (1981) data is the most detailed about types of article use; he identifies 5 types of use for the and a and 3 for Ø. Of the ten most common mistakes in his cloze test six are categorized as instances of generic or general use (3 where the correct answer was Ø, and 3 where it was a), and 3 as mis-renderings of idiomatic phrases (e.g. *for an example*).

The guides for teachers, Kharma and Hajjaj (1997) and Smith (2001), point out similarities and differences between articles in Arabic and English. As for positive transfer, Smith (2001) notes that the definite article in Arabic, is used as in English to refer back to indefinite nouns in previous discourse and for unique references (*the floor*). As for negative transfer, both studies have similar lists of likely problems:

- Because of the absence of an indefinite marker in Arabic,² initial underuse of *a* (*This is book*) is to be expected and is likely to be followed by overuse (*These are a books*)
- Differing patterns of definiteness for the nouns in genitive constructions are likely to transfer (*Car the teacher*)
- In a range of idiomatic uses learners are likely to “reinstate” definite articles omitted in English (*I went to the bed*)
- Proper nouns in Arabic often contain the article (*He lived in the India*).

Kharma and Hajjaj (1997) add four other likely transfer problems:

- non-ellipsis of articles in compound noun phrases (*the salt and the pepper*)
- use of the definite article (obligatory in Arabic) in generic plural noun phrases (*The horses are useful animals*)
- use of the definite article for abstract nouns (*All men fear the death*)
- use of the definite article for mass nouns (*The milk is nutritious to the body*).

An interesting confirmation that generic uses of the article system might be one of the main stumbling blocks for advanced L1 Arabic English learners is Farghal and al-Zou'bi (2004), who compare three different English translations of the Koran and show evidence that none of the translators has fully acquired how English handles generic reference, using *the* instead of Ø with generic plural NPs (*The lions are ferocious animals*) and using *the* in generic singular NPs where English would require *a* (*The wolf ate him for A wolf ate him*).

In terms of the role of transfer in causing errors, Scott and Tucker (1974) attributed half their subjects' article errors to transfer, and Kharma (1981) attributes the majority of his subjects' errors to transfer but adds the compounding effect of "inadequate instruction". Bataineh (2005) considers transfer responsible only for "*Ø-for-a*" errors and even then only as a possibility, citing overgeneralization, training transfer, learning and communication strategies as other possible causes.

Teachability

All the literature attests to the difficulty of the task L2 learners face in acquiring the English article system. Another pedagogic consideration is error gravity: if learners can communicate satisfactorily without mastery, is it worth devoting much time to article instruction? Master (1997, 2002) notes that non-mastery of the system damages the rhetorical credibility of those using English for academic purposes and argues that for these learners pedagogic intervention is called for: he cites research suggesting that instruction can make a difference to proficiency (Master, 1994, 1995, 2002) and in Master (1997) makes specific suggestions about instruction for learners at the main stages of proficiency. Liu and Gleason (2002) also suggest an instructional order appropriate to the order of acquisition they posit for non generic uses of the.

On the attested difficulty of acquiring high frequency L2 morphemes such as articles, in spite of the vast exposure to them learners receive, Ellis (2006b) argues that such difficulty is evidence that L1-like, naturalistic acquisition does not take place. He argues that this is because the form-function mappings of L1 inhibit or "block" learners perceiving those of L2, i.e. they fail to notice differences, and his conclusion is that without instruction fossilization is inevitable.

Research questions

There have recently been a few corpus-based studies of article use in naturalistic L2 English. Ringbom (1998) noted that advanced learners of English with western European L1s all use the less frequently than native English speaker (NES) writers. Kachru (2003) looked at definite article use in acrolectal world English varieties and found little variation from NES norms. As we saw, McEnery, Xiao and Tono (2006)

found evidence suggesting that even proficient L1 Japanese learners had not acquired articles. The research described in this article is a corpus-based enquiry into article system errors in the writing of intermediate to advanced level L1 Arabic English learners. It tallies and categorizes the commonest errors in a naturalistic data corpus (i.e. a corpus of language not produced to elicit data on a specific linguistic item) and seeks to identify how much error can be attributed to transfer. It seeks to combine the delicacy of Kharma's (1981) categorization of types of error, Master's (1997) dual description of error frequencies in terms of form supplied and form required³, and Kharma and Hajjaj's (1997) categorization of transfer errors with a larger and hopefully more representative set of up to date naturalistic data than Kharma's (1981) to answer the following questions:

RQ1 How common are article system errors in university level L1 Arabic English learners?

RQ2 What are the commonest types of errors?

RQ3 To what extent are these errors likely to have been caused by transfer?

The article will conclude with a discussion of the research and pedagogic implications of the answers to these questions.

Method

Data

The corpus used was a subcorpus from a larger corpus of argumentative essays, modeled on the International Corpus of Learner English (ICLE) (Granger, Dagneaux & Meunier, 2002) written by first and second year students (aged 18-20) at the American University of Sharjah, an English medium university in the United Arab Emirates. The subcorpus, entitled the Arabic Learner English corpus (hereafter the ALE corpus), comprises 95 essays totaling 42,391 words (mean essay length 446 words) written by students giving Arabic as their L1. Only 21% are Emirati, the others coming from a cross-section of the many different Arab nationalities resident in the Emirates. The essays were word-processed and submitted as regular class assignments; they were often completed in the student's own time and not produced under test conditions. The English proficiency level of the students varied considerably; all had had to score a minimum of 500 in the TOEFL exam to enter the

university, a few may be bilingual with one Anglophone parent, and secondary school education may have been in English, Arabic, or a mixture.

Procedure

All tokens of *the* and *a* in the data were identified. The data was analyzed with the help of *Wordsmith Tools 4* (Scott, 2006). To identify \emptyset tokens and to speed analysis (by sorting KWIC concordances into recurrent patterns) the corpus was annotated with part of speech (POS) tags using the software application *Qtag* (Mason, 2008). As identifying tokens of \emptyset proved very time-consuming (see Mason and Uzar [2000] on the technical problems), it was decided to analyze a sample of these and extrapolate a likely overall frequency for the corpus. The sample of \emptyset tokens was obtained as follows: a random sample of 1000 of the 10,075 POS noun tokens was analysed. Instances (i) where noun tokens formed heads of NPs which included *the* and *a* and (ii) where noun tokens served as modifiers of other NP heads were removed. The remainder of the instances (373) were identified as \emptyset tokens.

Once the article tokens had been identified, the totals of tokens for each article type were converted to standardized frequencies and compared with those in other L1 and L2 English corpora. Next, each token was analyzed in context for correctness or incorrectness. This often required searching prior discourse for specific reference to the referent of the NP of which the article was part – directly or indirectly. For example, if the article token *the* was part of an NP *the student* did the prior discourse contain prior mention of *a student* or invoke a general situation (Quirk et al., 1985, p. 266) of which *the student* forms part? Conversely, if the article token *a* was part of an NP *a student* was this the first mention of this referent? If not, were these plausibly generic references? Once each “incorrect” token had been identified, an analytical decision was made as to which article would have been correct in its context. Usually this was unproblematic but occasionally it was concluded that no article at all fitted the context – the noun phrase needed a different determiner to be correct or no correction could be made at all. Examples of the former error are given in “Results” below. Tokens of the different error types for each article, e.g. “the-for-a”, were tallied. Finally, within each error type errors were further classified into typical

patterns. In the next section the findings are presented and compared with those discussed in the literature review.

Results and discussion

Detailed frequencies of article tokens in the ALE corpus – and, if used in error, their corrections – are given in Table 1. Figures 1 to 5 in Sections 4.1 and 4.2 are intended to represent visually the main tendencies revealed in Table 1.

Table 1: Frequencies of articles supplied in ALE by article type and correction type

form required in context	article supplied				
	The	a/an	Ø*	<i>total visible articles</i>	total all articles
Ø	288	27	3686	<i>315</i>	4001
a/an	61	995	50	<i>1056</i>	1106
The	1872	6	20	<i>1878</i>	1898
another determiner	43	8	0	<i>51</i>	51
Total	2264	1036	3756	3300	7056
Error	392	41	70	<i>433</i>	503
correct use	1872	995	3686	<i>2867</i>	6553

* projected

Article form frequencies

Comparison of frequency counts of linguistic features in corpora give a useful overall picture of ways in which variation might occur. As Hunston (2002) notes, several studies of learner corpora show that learner writing has features more typical of spoken than of written academic English, such as more first and second person pronouns and more emphatic particles. Figure 1 compares frequency data from the ALE corpus, originally and after correction, with data from four registers in the Longman Grammar Corpus (LGC) as shown in Biber et al. (1999), and the Louvain Corpus of Native English Essays (LOCNESS). Why the LOCNESS corpus should have higher frequencies overall and a greater proportion of *the* than the other NES corpora is not known. The uncorrected ALE corpus appears to be much closer to the

LGC academic register than either the spoken register or fiction, both in terms of overall frequency and proportion of definite to indefinite articles, but when corrected it resembles the LOCNESS corpus even less than before and the LGC academic writing register less than those of fiction and news.

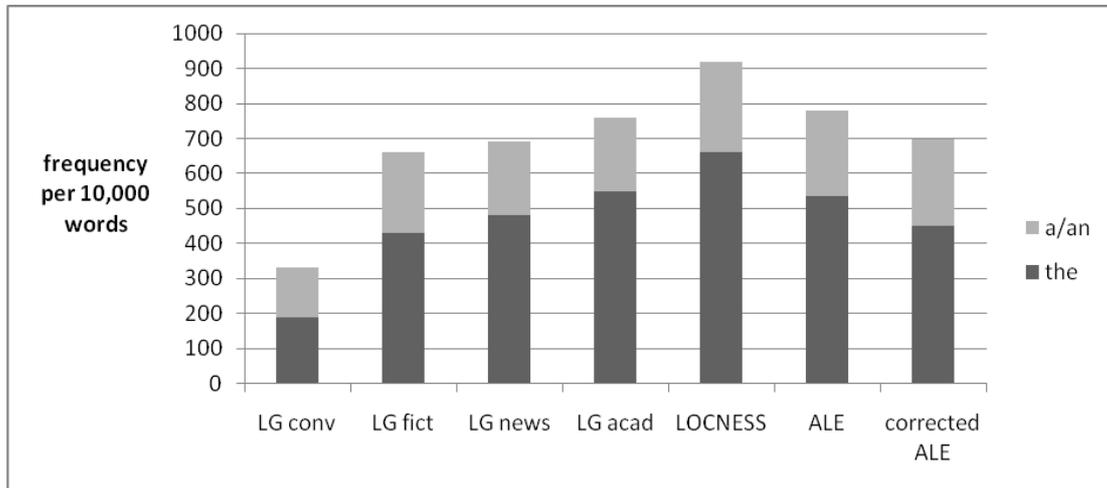


Figure 1: Frequencies of the visible articles in various NES corpora and ALE

Figure 2 compares the same LOCNESS and uncorrected ALE frequency data with data from two other learner English corpora from the ICLE corpus. Ringbom (1998) notes that NES writers use the more than any of the NNES writers in the ICLE corpus and the ALE corpus seems typical in this respect. The frequencies for the French corpus more nearly resemble those of the NES corpus. Interestingly, from a transfer perspective, although Arabic is a [+article] language the ALE corpus frequencies more closely resemble those of the Polish corpus, Polish being a [-article] language.

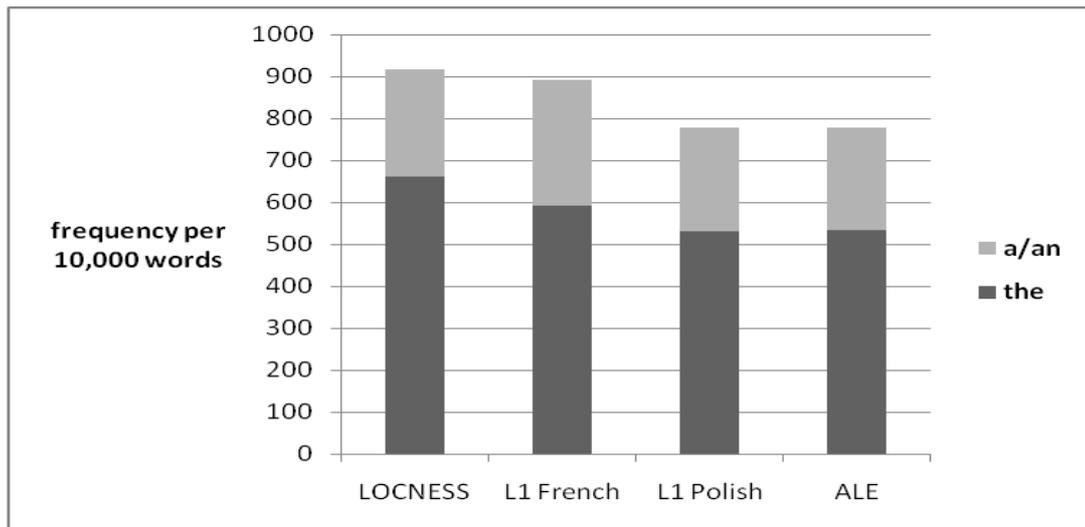


Figure 2: Frequencies of the visible articles in LOCNESS and other learner corpora

Frequencies of \emptyset in NES language are harder to obtain, presumably because of the difficulty in automatizing identification mentioned earlier, and they are not supplied either for the Longman corpus in Biber et al. (1999), or so far as I am aware, for the LOCNESS corpus. Master (1997) contains relative frequencies for the, a, and \emptyset counted in a corpus of 200,000 words from an NES corpus of EST articles. For the ALE corpus, extrapolating from the sample, we can estimate a figure of 888 \emptyset tokens per 10,000 word tokens of running text. Figure 3 compares Master's NES figures with the proportions in ALE, uncorrected and corrected. The uncorrected proportions would suggest that the L1 Arabic writers are underusing the and overusing \emptyset . However, the proportions in the corrected ALE corpus resemble the NES corpus even less. It may be the case that the variation here is at least partly due to register.

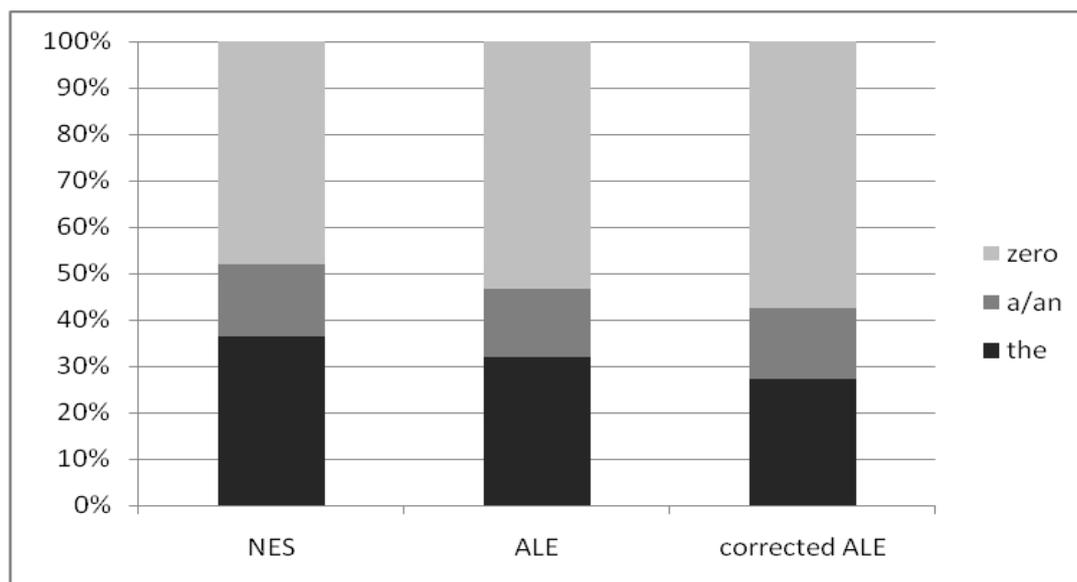


Figure 3: Proportions of articles in NES, ALE and the corrected ALE

In summary, one might have expected that with corrections, Non-Native English Speaker (NNES) text would have resembled what could be deemed target types of NES text more rather than less. These findings suggest that \emptyset is over-used generally by L1 Arabic writers, but that variation between article system use in NES and NNES text is not simply explicable in terms of linguistic proficiency: the reduced ratio of the to a suggests that the ALE corpus is conforming to Hunston's (2002) pattern of learner corpora resemblance to speech rather than writing. Use of articles in general and \emptyset in particular in NES language requires further research.

Proportions of correct and incorrect use

Turning from the overall frequencies of the articles, the issue of accuracy will now be addressed. Errors in article use can be conceptualized as a mismatch between a form and a context. There are two perspectives for viewing such mismatches – that of the article that has been supplied and that of the article which the context requires (or in some cases, the several article forms which the context permits). Figure 4 shows accuracy from the perspective of the article supplied. As can be seen, the errors are far from evenly spread between *the*, *a* and \emptyset . Do the error rates suggest that these L1 Arabic learners have acquired the English article system? Taking Dulay and Burt's (1974) accuracy rate (after Brown, 1973) of 90% or higher as a sign of

acquisition, if we include \emptyset as part of the system, the answer would be “yes”, the accuracy rate is 93%. If we leave out \emptyset and only consider the visible articles *the* and *a*, however, the accuracy rate drops to 87%. Given that this production is self-edited and largely written without time constraints, the issue seems borderline. Considered as subsystems, it would seem fair to conclude that \emptyset (98%) and *a* (96%) have been acquired but that *the* (83%) has not. That is to say only 83% of all *the* tokens in the corpus are used correctly. To check that *the* errors were not confined to a few low-proficiency authors the spread of the errors was checked: 90% of authors (86/95) mis-used *the* at least once, and the median number of mis-use errors per essay/author was three.

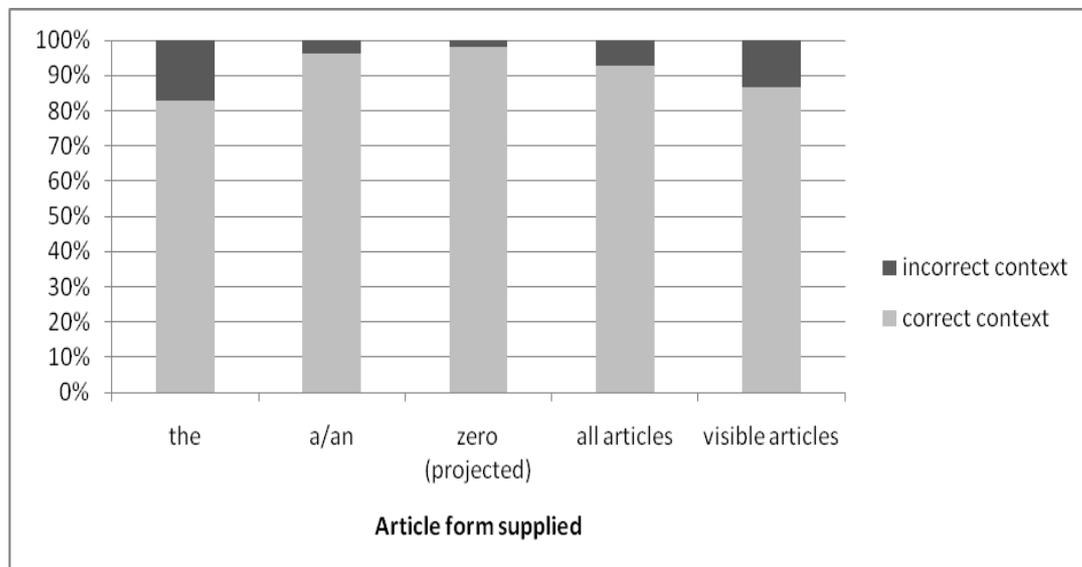


Figure 4: Proportions of article forms in ALE used correctly

An alternative way of considering the data, more comparable with the suppliance in obligatory context (SOC) elicitation tasks used in some research (e.g. Kharma, 1981; Master, 1997), is to analyze errors in terms of the article required in context. A slight disadvantage of this perspective is that it ignores overproduction of articles in toto: in this corpus, there were 51 instances of articles being used where the syntax of the NP required something else, for example a possessive or demonstrative. It is worth noting here that the existence of such instances highlights a limitation of some SOC elicitation techniques, i.e. that they will not show up features of interlanguage lacking a cognate in the target L2.

Figure 5 represents the error data from the perspective of context and suggests that in relative terms the article on which learners have the weakest hold is a (90%), followed by \emptyset (92%), and then the (99%). That is to say, in only 90% of the contexts in which a was required was it supplied. In terms of the 90% accuracy rate, learners can be said to have acquired a. Checking to see the spread of error in the corpus can only be done for “the-for-a” errors as the “ \emptyset -for-a” figures are projected. The “the-for-a” figures show that this error appears in the texts of only 38% (36/95) authors. The more detailed analysis of this error-type below reveals that the error involves one particular use of a, for generic reference. Without a search of and frequency count of correct instances of this particular use, it is impossible to judge whether the relatively low spread of this error is due to either (i) mastery of this particular use by the majority of authors or (ii) a low number of attempts at this use.

To summarize, the answer to RQ1 is complex, then, but appears to be that errors with these are common enough to cast doubt on these learners’ mastery of the form.

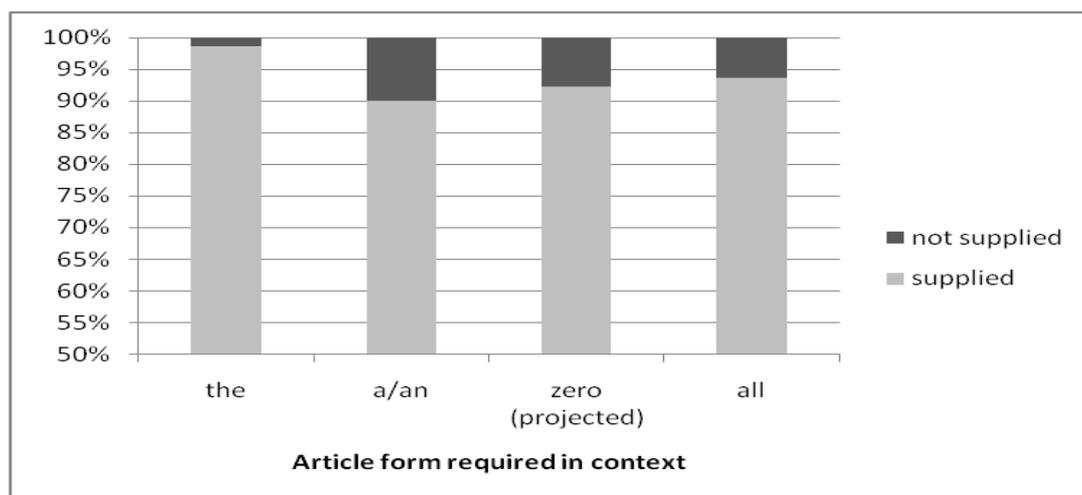


Figure 5: Proportions of article contexts in ALE for which the correct article form was supplied

Patterns of incorrect use

In this section the main patterns of error are reviewed in some detail and the plausibility of L1 transfer being the cause is discussed. Table 2 answers RQ2, listing errors from the perspective of the item supplied, ordering errors from most to least common.

Table 2: Frequencies of types of article error

Type of error	N	%
the-for-Ø	288	57
the-for-a	61	12
Ø-for-a*	50	10
the-for-other determiner	43	9
a-for-Ø	27	5
Ø-for-the*	20	4
a-for-other determiner	8	2
a-for-the	6	1
Total	503	100

* projected

“the-for-Ø” errors.

As we have seen *the* is the most frequently mis-used article and its most frequent mis-use is in Ø contexts. These contexts can be further subdivided into the error types shown in Table 3.

Table 3: Frequencies of subtypes of ‘the-for-Ø’ errors

Type of construction	N	%
generic non count nouns	128	44
generic plural count nouns	109	38
generic singular count nouns	34	12
Idioms	17	6
Total	288	100

The following generic non-count nouns, usually abstract, all occurred more than once: business, company profit, culture, growth, human nature, imagination, lack, money, nature, environmentalism, ozone depletion, poverty, society, real life, television, war, water, work. It should be noted that all of these could appear non-generically with the in a range of contexts, such as previous mention, or invocation of a schematic situation referred to or implicit in the discourse context. However, each analysis of incorrectness was made on the basis of careful inspection of the previous

discourse. The society (21), the money (15), and the real life (9) were the top three.

Here are some characteristic examples:

- [1] ...These things destroy *the society*.⁴
- [2] Some of us consider *the money* as the force which controls our lives, while others...
- [3] King Fahd University graduates are knowledgeable and ready to join *the real life* from the first day in their business.

The occurrence of the money is perhaps some small evidence in favour of Ellis's (2006b) "blocking of perception" account of L1 influence, as Ø money actually featured in the prompt used to elicit many of the texts in which it appears ("*Money is the root of all evil*". *Discuss.*) The real life is an interesting case where the prompt contained the semantically related expression *the real world*: "*Most university degrees...do not prepare students for the real world.*" *Discuss.*

Examples of idioms involving generic non-count nouns are the following:

- [4] At the moment a problem arises, the first thing that *comes to the mind* is how to deal with it.
- [5] ...students can memorize some materials and *put it on the paper* the next day.

The next largest error category was generic plurals:

- [6] Others say that *the degrees* are mainly theoretical...
- [7] This invention allowed people to value *the goods and services*.
- [8] ...they are usually good with thinking about the whole including *the individuals*.

Related to this is the next error category – singular nouns which are intended generically but do not satisfy the normal conditions for using *the*, for example:

- [9] *The human* lives in the earth and his existence depend on nature's ...[*Humans live on earth*]
- [10] ...in some way it seems as though *the society* is "baby-sitting" criminals [Ø *society*]
- [11] The root of all evil is *the man's* desire for status and greed... [*man's* or *mens' desire*]

Examples of errors which show learners to be unaware of idiomatic NES non-use often involve nouns representing institutions – *jail, prison, university, hospital*, for example:

[12] *In the university*, students do not gain the experience...

“*the-for-a*” errors.

The vast majority of these errors (55 out of 61) were where the writer intended a generic expression. Another way of correcting these would be to pluralize them and use \emptyset , but because this involved two corrections rather than one they were analysed as singular:

[13] I believe that the personality of *the person* is what stands out rather than just ... [*a person, people*]

[14] The only thing a university degree does is provide *the student* with the basics required for him to understand ... [*a student, students*]

[15] For example, *the water bottle* needs years to decompose. [*a water bottle, water bottles*]

“ *\emptyset -for-a*” errors.

There is only a small sample of these. All involve a generic, classificatory meaning, for example:

[16] Money is [\emptyset] way to ease a person’s life.

[17] For example, you are [\emptyset] person that do not have enough money...

[18] In conclusion, [\emptyset] theoretical education is very poor...

The head nouns in both the first two examples are countable and would require an article in any context, generic or non-generic, but the last is more complex in that *education* can be a non-count noun and only requires *a* here to carry the generic meaning required in this context – a discussion about types of education.

“ *\emptyset -for-the*” errors.

The sample of these errors is even smaller. One involves an idiom:

[19] While in [\emptyset] contrary, the sentimental environmentalism are spending ...

“*a-for-Ø*” errors.

The sub-types of error in this category are shown in Table 4.

Table 4: Frequencies of subtypes of ‘a-for-Ø’ errors

Type of construction	N	%
generic non-count nouns	16	59
plural count nouns	6	22
Idiom	3	11
Other	2	7
Total	27	100

The majority of errors involved generic non-count nouns, for example:

- [20] Since money is the most accurate measurement of *a status* for individuals and citizens in a country.
- [21] Every one in that family lives in *a harmony*...

The next largest category involved plural countable nouns, for example:

- [22] People have different point of views regarding if *a university degrees* make individuals more capable or are they just worthless certificates.

This surprising error type is listed by Smith (2001) as to be expected and Bataineh’s (2005) data contained 14 such errors. Such errors are presumably the trace of an earlier stage of learning.

4.3.6 *Other determiner required.* Of constructions which require a determiner other than an article, the largest type (22/51) are those requiring a possessive, for example:

- [23] ... and as he's the son of *the father* at that time, crime is therefore an act of instinct [*his father*]
- [24] These groups, jump from the A and B, and started their life from C, and jumping will break *the leg* even after a long time... [*their legs*]
- [25] So they use money to remove what ever that can face *the dream* by using money. [*their dream*]

There are 6 cases of *most of the* where *most* alone is required, e.g.

- [26] This is why *most of the people* fight to increase their fortune.

These can be considered as modified generics, [26] being for example a hedged form of *This is why people fight to increase to their fortune*.

Transfer

Having surveyed the commonest types of errors in some detail, what can we conclude about RQ3, the role of L1 transfer? The English article system is famously complex. Analytical taxonomies of article use in the literature typically list around twenty contexts for the articles *a*, *the* and *zero*, of which only three are generic (Berry 1993; Thomas 1989). Liu and Gleason argue that ‘the non-generic use of *the* is much more complex and hence more problematic for ESL students than the generic use’ (2002, p. 6). First, let us note the surprising fact that none of the errors involved countable specific referents: most involved either uncountable or countable generic referents. This suggests, as predicted by Smith (2001), positive transfer of use of the definite article for specific reference.

It is, secondly, striking that the largest single category of errors is incorrectly formed generic expressions: looking at the “visible” articles alone (as the small sample size for \emptyset errors makes the figures unreliable) the proportion of errors which involve an apparent intended generic meaning is 79% (342/433, see Table 5 for a detailed breakdown). Of these errors only 16 did not involve the incorrect use of *the*. In other words, 75% of all the errors in use of *a* and *the* consisted of learners using *the* where it was not required in order to form a generic expression.

Table 5: Frequencies of errors with ‘visible’ articles in generic reference by NP head type

Supplied	required	NP head	N
The	\emptyset	generic non count noun	128
The	\emptyset	generic plural count noun	109
The	\emptyset	generic singular count noun*	34
The	A	generic singular count noun	55
A	\emptyset	generic non count noun	16
Total			342

* pluralization also required

Are these errors likely to be due to L1 transfer? As we noted earlier Kharma and Hajjaj (1997) predicted errors with generic noun phrases and with abstract and mass nouns – it happens that in these essays abstract and mass nouns invariably appeared as generics, so these error types are often conflated.

In Arabic, in which there is no indefinite article, indefiniteness is marked by the absence of the definite article (Schulz, 2004).⁵ The only forms in Arabic permissible for generic NPs, singular, plural, or non-count, require the definite article (Farghal & al-Zou'bi, 2004). In addition to the fact that there is a number distinction in the English indefinite article and therefore two forms (*a* and \emptyset), there are two factors likely to make acquisition of English articles in generic expressions difficult for L1 Arabic speakers. Firstly, recalling Lyon's (1999) categorization of generics as being identifiable, the Arabic article system does not deviate from the prototypical meaning-to-form mapping of identifiability-definiteness as the English system does. Secondly, the English system does not deviate consistently, but sometimes admits generics as definite. Table 6 compares some feature of the two systems.

Table 6: Comparison of definiteness marking according to number and countability in English and Arabic

head of generic NP	pattern	definiteness marking in article	English	Arabic
plural count noun	i	indef.	✓	✗
	ii	def.	✓ *	✓
singular count noun	iii	indef.	✓	✗
	iv	def.	✓ *	✓
non count noun	v	indef.	✓	✗
	vi	def.	✗	✓

*= sometimes

There are in theory 6 patterns possible. Berry (1993), based on data from the COBUILD corpus, states that the commonest generic reference patterns in English are (i) and (v). If the less common (ii) and (iv) patterns never appeared in English there would be complete formal definite/indefinite asymmetry and it might arguably be cognitively easier for Arabic speakers to notice the differences between English and Arabic. However, there are special cases in which the Arabic patterns are permissible in English (Berry, 1993):

- (ii) singular count noun + definite article
- regular participants and roles in a situation – *the teacher* in education
 - animal species – *the red squirrel*
 - body parts as referred to by doctors – *the head*
 - inventions – *the computer*
 - rooms – the kitchen.
- (iv) plural count noun + definite article
- nationality words – *the Americans*
 - nominalized adjectives – *the poor*.

One could predict that even without “blocking” effects from L1 acquisition (Ellis, 2006b) acquiring the semantic cues for these special cases will require a lot of exposure.

Table 7 breaks down the errors with generic expressions into the six patterns. It shows that the vast majority of the errors follow patterns which would be correct in Arabic (ii, iv and vi). The small remainder come from making a wrong choice about the form of the indefinite article (v).

Table 7: Errors with generic reference by type, ranked from most to least frequent

Pattern	nature of error	Correction	N
vi	the + non count noun	def. for indef. the to Ø	128
ii	the + plural count noun	def. for indef. the to Ø	109
iv	the + singular count noun	def. for indef. the to a	(55)
		the to Ø + pluralization	(34) 89
v	a + non count noun	wrong indef. Ø to a	16
Total			342

Until L1 Arabic learners notice the differences between the systems in English and Arabic, a strategy for count nouns of sticking to L1 patterns (ii and iv), i.e. routinely using the definite article in generic expressions, will produce correct English some of the time. In other cases, interlocutors (including English teachers) may not recognize the error as a formal one, and may incorrectly assume that the speaker has made a pragmatic error, i.e. is using the definite article infelicitously to make a specific reference - for example to a referent she imagines to be familiar to the

hearer - rather than a generic reference. Such misinterpretations may in turn reduce the chance of the learner's attention being drawn to the error.

In terms of causality, generic article errors are arguably both interlingual and intralingual: for example, the learner who writes "*the degrees* are mainly theoretical" has both transferred an L1 form and possibly overgeneralized that this is a permissible form in L2 from the evidence of L1 input, both generic, such as "*the Americans*" above, and non-generic, such as "*the best degrees*". However, in terms of the original interlingual-intralingual distinction these errors are not obviously intralingual because they do not take the form of constructions which would be impossible in L1.

Let us summarize by reviewing the evidence needed to answer RQ3, which concerns the extent to which article system errors in this corpus appear likely to have been caused by transfer. The situation revealed by the data is that learners appear to have mastered all the articles and usually supply *a* and *the* in the contexts in which they are required. Unfortunately, however, they also tend to use *the* in contexts where it is not required. This over-use of *the* is not scattered equally throughout all possible article contexts, however, but is, in the main, as we have seen, confined to a small minority of contexts, in particular those involving a generic meaning. Although, generic expressions are much less frequent than specific expressions overall, three quarters of all errors made with *the* and *a* occur in what are intended as generic expressions. Examination of the L1 shows that there are only three permissible forms of generic expression, each of which requires use of the L1 cognate of *the*. In short, it would seem reasonable to conclude there is a strong likelihood that transfer from L1 is a major factor in causing most of the article errors found in the corpus. In terms of form-function mapping one would say that Arabic learners have learnt from L1 to map the function 'marker of generic expression' onto the form 'definite article'.

Conclusion

There is not space here to explore the role of transfer in some of the much lower frequency types of error not related to generic reference noted in the corpus. However, in terms of the second research question about the commonest errors, those with generic reference are clearly predominant. The argumentative genre of the texts

which comprise the data for this research probably contain a particularly high rate of generic reference; a corpus of narrative texts would presumably have a higher level of specific reference, anaphoric reference in particular, and could thus well have had lower overall error rates in article use. However, texts like those in ALE and LOCNESS are not untypical of assignments in academic writing classes, and it can be expected that much disciplinary academic writing will require students to make general statements about classes of things and thus use generic expressions.

Because the analysis in this research has been confined to errors, we do not know how far the incorrect article use described here exists among individual authors in free variation with correct article use (as found in Jarvis, 2002; Robertson, 2000). My assumption is that even if there were also a high level of correct article use for generic reference, the errors discussed here are evidence of unsystematic variation, and therefore evidence that acquisition has not taken place. Overall then, to answer the research questions on the evidence so far discussed, it seems fair to conclude (RQ1) that despite an impressive accuracy rate in article use for specific reference a majority of the L1 Arabic advanced learners of English represented in the corpus have not fully mastered the English article system, (RQ2) that most of the errors they make with the visible articles are in the expression of generic reference, most of these errors involving over-use of *the*, and (RQ3) that transfer is a likely factor in such errors. These findings suggest that, despite the considerably increased exposure to English brought about by social, economic, and technological changes in the last quarter century in the Arabian Gulf region, the situation has changed little from when Kharma (1981) tested article acquisition among tertiary level L1 Arabic learners of English.

The broader implications of these findings seem to be as follows. Learners from languages without article systems apparently have more problems (a) overall and (b) earlier in acquiring the English article system (Master, 1997); however, learners from languages with article systems, such as Arabic, may also face significant problems and these problems may survive until relatively advanced stages of learning. Furthermore, the problems faced may be related to specific L1 system features, and not, as some earlier SLA theory suggested, features of a common developmental interlanguage system. In general, in confirming the persistence of such errors, in the

face of high frequency exposure to L2 input, the findings also tend to support Ellis's (2006b) argument for maximal transfer and interference from L1 in L2 learning.

Overall, research into learner problems with English articles in general and with generic reference in particular would benefit from more corpus-based research into frequencies of the major reference patterns in native-speaker English, similar to that shown in Biber et al. (1999), but extending to indefinite NPs, including the numerically dominant \emptyset NPs.⁶ The findings of the current study suggest that research into generic reference in the English of learners from L1s with and without articles would be worthwhile, possibly revealing negative transfer effects worthy of pedagogic intervention. In particular, across learners from [+article] L1s it would be interesting to compare similar data from two further subcategories, ELS learners with L1s in which, as in Arabic, generics are marked by definiteness, and those with L1s in which, as in English, generics are marked by indefiniteness, subcategories, which we might designate as [+generic definiteness] or [-generic definiteness]. The research described here has concentrated on errors: to fully gauge learners' accuracy rates at the different types of article use it would be useful to analyze also frequencies of successful article use in generic expressions in L2 corpora. Research on article use by L1 Arabic speakers at earlier levels of language learning would also help to confirm whether the hypothesis of early positive transfer of specific reference is supported.

The pedagogic implications seem to be as follows. Problems related to the impact of the count/non-count distinction on definiteness marking would presumably be reduced if vocabulary teaching in the earlier stages of learning emphasized this distinction more, as suggested by Master (1997). The commonest form of generic expression for countable nouns in English is that with \emptyset and plurals. For learners with [-article] L1s this use of \emptyset will transfer quite naturally. For L1 Arabic learners, and presumably for learners with other [+generic definiteness] L1s, such as French, negative transfer can be expected. At the intermediate level, Master focuses on the discourse dependence of many learner problems but in the case of L1 Arabic learners we have seen that this is not a primary cause of errors. However, Master's suggestions for helping advanced [-article] L1 learners overcome "Ø-for-the" errors may also apply to helping [+article] L1 learners overcome "the-for-Ø" and "the-for-a" errors. He recommends focusing on articles in both reading and writing

development, “especially the contrast between minimal pairs of lexical phrases with \emptyset and *the*” and the need for repetition of the process “with many different lexical items since it can no longer be assumed that learners at this stage will generalize from specific instances” (1997, p. 228). Master assumes that conscious learning has little effect on acquisition but if recent theorizing about the role of consciousness upon implicit cognition (Ellis, 2006a; 2006b; Ellis & Larsen-Freeman, 2006) is correct it would also be worth spending class time explicitly pointing out the various formal possibilities and impossibilities for generic reference in English. Finally, and perhaps most importantly, in order for maximal opportunity to be taken of promoting noticing, it would seem worth making known to those engaged in instructing L1 Arabic ESL learners the likely L1 origins of some persistent errors in their students’ production.

Notes

1. In this paper *a* will be taken to stand for both *a* and *an*. I have avoided using the expression “the indefinite article” to describe *a/an* as this practice is inconsistent with recognizing \emptyset as an indefinite article
2. In some varieties of Arabic, indefiniteness is marked by a form of affixation, nunation. However, while nunation is a feature of written acrolectal Arabic, in the modern spoken dialects it is described as having “all but disappeared” (Holes, 2004, p. 174) and being “rare” (Abu-Chacra, 2007, p. 18).
3. Master (1997) makes this distinction in terms of accuracy and overuse but I have avoided these particular terms as being susceptible to misinterpretation outside the context of his paper.
4. Errors in quotations from the ALE corpus are not corrected – although suggested alternatives are sometimes presented in square brackets - and all italicization and ellipsis is mine.
5. See note 2 above.
6. Biber et al. (1999, p. 266) gives figures for generic *the* but not for *a* or \emptyset .

References

- Abu-Chacra, F. (2007). *Arabic: An essential grammar*. New York: Routledge.
- Bataineh, R. F. (2005). Jordanian undergraduate EFL students' errors in the use of the indefinite article. Retrieved June 11, 2008 from Asian EFL Journal: http://www.asian-efl-journal.com/march_05_rfb.php
- Berry, R. (1993). *Articles*. London: HarperCollins.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, Essex: Longman.
- Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33 , 1-17.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Butler, Y. G. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition* 24, 451-480.
- Chaudron, C. & Parker, K. (1990). Discourse markedness and structural markedness. *Studies in Second Language Acquisition: The Acquisition of English Noun Phrases* 12(1), 43-64.
- Dulay, H. & Burt, M. (1974). Errors and strategies in child language acquisition. *TESOL Quarterly* 8, 129-36.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York: Oxford University Press.
- Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics* 27(1), 1-24.
- Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, saliance, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2), 164-194.
- Ellis, N. C. & Larsen-Freeman, D. (2006). 'Language emergence: Implications for applied linguistics - Introduction to the special issue. *Applied Linguistics* 27(4), 558-589.
- Farghal, M. & al-Zou'bi, I. (2004). The Qur'anic definite article in translation: The case of citing animals. In R. Elhamy, H. Mansour, & J.R. Aydelott (Eds.),

- Contrastive rhetoric: Issues, insights and pedagogy* (pp. 61-68). Cairo: American University of Cairo Press.
- Granger, S., Dagneaux, E. & Meunier, F. (2002). *International corpus of learner English*. Louvain: Presses Universitaires de Louvain.
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties (2nd edn.)*. Washington: Georgetown University Press.
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor: Karoma.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition* 12(1), 3-69.
- James, K. (1998). *Errors in language learning and use: Exploring error analysis*. London: Longman.
- Jarvis, S. (2002). Topic continuity in L2 English article use. *Studies in Second Language Acquisition* 24, 387-418.
- Kachru, Y. (2003). On definite reference in world Englishes. *World Englishes* 22(4), 497-510.
- Kharma, N. (1981). Analysis of the errors committed by Arab university students in the use of English definite-indefinite articles. *International Review of Applied Linguistics* 19(4), 331-345.
- Kharma, N. & Hajjaj, A. (1997). *Errors in English among Arabic speakers: Analysis and remedy*. Beirut: York Press Librairie du Liban.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: Michigan University Press.
- Liu, D. & Gleason, J.L. (2002). Acquisition of the article "the" by nonnative speakers of English. *Studies in Second Language Acquisition* 24, 1-26.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Mason, O. (n.d.). *Qtag*. Retrieved June 1, 2008 from <http://www.english.bham.ac.uk/staff/omason/software/qtag.html>
- Mason, O. & Uzar, R. (2000). NLP meets TEFL: tracing the zero. In B. Lewandowska-Tomaszczyk & J. Melia (Eds.), *TALC '99 Practical Applications in*

- Language Corpora: papers from the international conference at the University of Lodz, 15-18 April 1999* (pp. 105-116). Frankfurt: Peter Lang.
- Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly* 24(3), 461-478.
- Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives in pedagogical grammar* (pp. 229-252). Cambridge: Cambridge University Press.
- Master, P. (1995). Consciousness raising and article pedagogy. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language* (pp. 183-204). Norwood, NJ: Ablex.
- Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System* 2(2), 215-232.
- Master, P. (2002). Information structure and English article pedagogy. *System* 30, 331-348.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Mohammed, A. M. (2000). Modern standard Arabic vs. non-standard Arabic: Where do Arab students of EFL transfer from? *Language, Culture and Curriculum* 13(2), 126-136.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning* 37, 361-383.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In Granger, S. (Ed.), *Learner English on computer* (pp. 41-52). London: Longman.
- Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research* 16(2), 135-172.
- Schulz, E. (2004). *A student grammar of modern standard Arabic*. Cambridge: Cambridge University Press.

- Scott, M. (2006). *Wordsmith tools 4*. Oxford: Oxford University Press.
- Scott, M. S. & Tucker, G. R. (1974). Error analysis and English language strategies of Arabic students. *Language Learning* 24(1), 69-97.
- Smith, B. (2001). Arabic speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems (2nd Ed.)* (pp. 195-213). Cambridge: Cambridge University Press.
- Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics* 10(3), 335-355.
- Thompson-Panos, K. & Thomas-Ruzic, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly* 17(4), 609-623.
- White, L. (1977). Error analysis and error correction in adult learners of English as a second language. *Working Papers in Bilingualism* 13, 42-58.
- Zdorenko, T. & Paradis, J. (2008). The acquisition of articles in child second language English: fluctuation, transfer or both? *Second Language Research* 24(2), 227-250.
- Zegerac, V. (2004). Relevance Theory and "the" in second language acquisition. *Second Language Learning* 20(3), 193-211.



Error Correction of L2 Students' Texts – Theory, Evidence and Pedagogy

Michael John Alroe

Chulalongkorn University Language Institute, Thailand

Bio Data:

Michael Alroe (malroe2003@yahoo.com.au) studied English Literature at Macquarie University then taught over for 20 years in New South Wales high schools. Recently he has been teaching EAP at Chulalongkorn University in Thailand.

Abstract

In the last ten years there has been a revival of support for error correction of second language students' written texts underpinned by disillusionment with the notion that comprehensible input alone is sufficient for acquisition of all aspects of a second language. Empirical studies in comprehensive and focused correction have aided this revival and demonstrated that error correction can produce significant benefits. These studies also indicate that comprehensive error correction of written work can be done economically by simple underlining, and is effective if students write a substantial amount and correct their errors. Further, correction can be integrated with content comment. The studies also show that focused correction may be limited in scope but powerful in effect.

Keywords: comprehensive error correction; focused error correction; direct, indirect, coded and marginal correction; content feedback; metalinguistic feedback; clarification requests.

Background

For most of the twentieth century error correction of second language students' texts was assumed to be an important and beneficial practice. That assumption is probably still held today by most students and many teachers, even more so in Asian countries where English is taught as a foreign language (Ho, 2008; Lee, 2005). But among theorists of Speech and Language Acquisition (SLA) the value of error correction began to be questioned in the 1980's. Earlier researchers such as Cohen

and Robbins (1976) and Hendrickson (1979), who investigated the practice in the decade before, reported results that could be read as negative towards error correction but opined that the problem was in the implementation not in the practice per se. Then in the 1980's error correction was reassessed. In L1 education, grammar instruction had already been undermined in many English speaking countries and was regarded as having little value (Hudson, 2001, p.1). As a major purpose of grammar instruction was to give students a standard against which they could calibrate the language they produced, its demise had an impact on error correction in L1 and later L2 education. Also, according to Hudson, the linguistic theory of Universal Grammar, as most famously embodied in the work of Noam Chomsky, made grammar and correction seem irrelevant because "grammar competence develops 'naturally' according to an innate programme" and this was taken to imply that correction was unnecessary (Hudson, 2001, p.1). These views were imported into SLA theory in the 1980's largely through the work of Stephen Krashen who argued that competence (in Chomsky's sense) in a second language was acquired implicitly and only by means of comprehensible input (Krashen, 1982, p. 10 et seq.). Grammar instruction and consequently correction could only be useful in learning about language but had nothing to do with its acquisition.

Krashen's theories on SLA were, and still are, influential and were widely disseminated beyond the eastern seaboard of the United States where he taught. In Bangkok for instance, a large element of one program for teaching Thai to foreigners is conducted strictly according to Krashen's principle of "comprehensible input" with students going through an extended silent period while they wait to begin to talk spontaneously and "naturally" in Thai (A.U.A., n.d.). In the United States, his principles informed the training of California's ESL teachers who were "taught to view errors as a necessary part of the developmental process of learning a second language" (Scarcella, 1996, p. 138). By the mid-1990's, however, Krashen's theory faced serious challenges both in the fields of pedagogy and research. In French immersion schools in Canada native English speakers had been receiving vast amounts of input for more than 10 years but, though they developed native speaker-like competence with respect to comprehension, their productive skills were found to be deficient compared to native French speakers (Swain & Lapkin, 1995, p. 372).

Krashen's comprehensible input hypothesis implied that after years of immersion in a French linguistic environment where they are exposed to massive amounts of high quality input, students should be able both to speak and write the language at a near native level. Krashen reconsidered the matter and suggested what was required was yet more input in the form of films, tapes and pleasure reading to remedy what he still saw as a deficiency of input (1998, p.10). Swain (2005), who reported the problem, dissented and suggested it was a deficiency of output, not input, that was the cause. Further, she advised that output errors, i.e. divergences from L2 target forms, should be corrected. In California, Krashen's colleague Scarcella reassessed the impact of his work and concluded that it had been harmful. She observed that ESL students in California had received vast amounts of input but "this did not lead to the development of basic features of English" (Scarcella, 1996, p. 136). She added that the lack of correction was a factor in students' inability to process the academic English they would encounter in tertiary study. Similarly Dana Ferris, another California based instructor, said the fact that she and other teachers were confronted with students who plainly needed error correction when the dominant theory of the time was against it, forced her to reassess, research and then publish a defense of correction (Ferris, 2002, p. x).

At about the same time as key theorists in ESL and French as a second language contexts were critically reviewing Krashen's application of L1 theories to L2 instruction, Professor Truscott, an instructor of English in Taiwan, published the first paper in a series that would defend Krashen's ideas on L2 instruction. He targeted, in particular, the practice of error correction in written texts (Truscott, 1996). This paper was followed by others on error correction (Truscott, 1999a, 2004, 2007, and 2009), another critiquing oral correction (Truscott, 1999b), one paper rejecting the Noticing Hypothesis (Truscott, 1998) and one specifying the errors teachers ought to correct if they ignored his prohibition and persisted with error correction (Truscott, 2001). Truscott's papers, crafted with impressive forensic skill, put a powerful counter argument to those who urged a move back to more explicit instruction including error correction. What is surprising about this is that these papers originated in an English as a Foreign Language (EFL) context. While Krashen's assertion that L2 acquisition proceeded only with provision of comprehensible input might, in theory, be

applicable to young learners in an English as a Second Language (ESL) context if Primary Linguistic Data (PLD) is available in sufficient quantity, in an EFL context like Taiwan, where students speak Mandarin or Taiwanese at home and study English for only three or four hours a week, the assertion that they could acquire English in the same manner as they did their L1 was a bold one. This was particularly so when ESL instructors in California and FSL instructors in Canada, whose students did have access to thousands of hours of comprehensible input, should find such exposure insufficient to acquire target levels in the L2 even after years of study, and were calling for a return to error correction and other explicit methods of teaching.

In the period since Truscott wrote his 1996 paper, ideas and research in error correction have been the subject of much argument. Below I will review this debate, the empirical evidence which has been used to bolster the arguments for and against error correction, and the implications for pedagogy in L2 writing instruction.

Three Theoretical Frameworks

In this section I will consider the theoretical framework from which error correction should be viewed. If error correction is effective, it should be able to be located within a theoretical context. Having such a framework should help us to make and test hypotheses and develop our understanding of it as a pedagogical practice. Moreover, as with Krashen's hypotheses, theory may well lead where practice will later follow.

Krashen and Truscott – Implicit Learning

Error correction for many of those who favor it has largely been a matter of common sense and pragmatism. Curiously Krashen, whose theories helped diminish the importance of error correction, did provide it with a role, albeit a reduced one, in the area of L2 writing. In more recent years he seems to have hardened his stance towards correction but in *Second Language Acquisition and Second Language Learning* (1981) he implied a place for it in discussing the Monitor Hypothesis.

In Krashen's theory, explicit L2 learning never results in acquisition as there is no interface between the two (Krashen, 1982, p.11). However, if there is a discrepancy between the output of acquired language and metalinguistic rules held in memory, the

student may notice his L2 is not on target and self correct. During natural speech in real time, the monitor has no time to operate. However, when writing, the learner has time to reflect and adjust the output. This self-correction is based on explicitly acquired knowledge and there is no reason why the knowledge gained from correction given by a teacher should not be included in this learning process. Krashen argued that the “optimal” learner is one “who uses learning as a real supplement to acquisition” (1982, p.20). He illustrated this with the case of ‘P’, a native speaker of Chinese whose acquired English, as evidenced in speech, was marked by numerous errors, yet who was able by “conscious knowledge of the rules” to write “virtually error-free English...” (Krashen, 1981, p.13).

Researchers who support error correction have not employed Krashen’s paradigm as justification. On the other hand, researchers who argue against error correction and believe that content feedback is the appropriate response to L2 student writing seem to do so from within Krashen’s framework. They argue that instructors should focus not on the form of the student output, but dialogue with students on the meaning of their texts. Presumably, a student who reads the teacher’s comment will consciously process the content while implicitly processing the formal aspects of the instructor’s language by means of their innate language acquisition device. Thus the teacher’s comment becomes potentially comprehensible input which assists acquisition. This idea finds expression, for example, in Kepner’s study on feedback given to American students learning Spanish in which the instructor’s comment was thought to provide a “model” from which the students could learn (Kepner, 1991, p. 310). Often, this notion is not formally acknowledged but it seems to underlie studies like those of Sheppard (1992) and Semke (1984) which compare the effects of error correction with those of content comments with the researchers endorsing the latter. Nonetheless, there is a serious weakness in these studies that derives from the fact that the researchers did not have a precise theoretical conception of how content comment was supposed to function. Also, their definition of content comment is not clear as it appears to include clarification requests (Sheppard) and recasts (Semke) which lead students to focus on their own output and which can be regarded therefore, as indirect forms of error correction. This issue will be considered further in discussion of Ashwell (2000) and Sheppard (1992).

When Truscott entered the debate in 1996 with his paper criticizing the practice of grammar correction, he used theoretical arguments that derived from Krashen and evidence from the available research. Since the publication of that paper, those in favor of correction have struggled to generate research to counter its conclusions and after each of their attempts Truscott has responded with critiques claiming their work fails to demonstrate that error correction has any benefit. In his original paper he argued, as did Krashen, that the process of acquisition follows a staged, natural order and is not amenable to intervention by a teacher through corrective input.

Truscott also adopts Krashen's distinction between learning language (by which he intends conscious and explicit) and acquisition of language and sharpens the contrast by renaming the former "pseudolearning" and describing it as "superficial and possibly transient" (Truscott, 1996, p. 345). He argues that grammar correction belongs in the "pseudolearning" category. More skeptical, albeit vaguer here than Krashen, Truscott appears to reject the Monitor Hypothesis as he asserts that learners will follow their intuition, not "pseudolearning" in self-editing. In taking this line he also rejects the arguments of DeKeyser (2003) who insists explicit knowledge can become automatized and unconscious to the extent that it is indistinguishable from acquired knowledge (p. 329). In a later paper on oral correction Truscott goes still further and speculates that explicit knowledge, if it becomes automatic, may form a cognitive block preventing the possibility of real language acquisition (Truscott, 1999b, p. 125).

Truscott's presumed rejection of the Monitor Hypothesis may derive from a conceptual oversimplification of the process of writing. Krashen's original notion of an interplay between explicit and implicit knowledge seems to better capture the way L2 learners approach writing. For in doing so, they must negotiate a complex interaction between implicit and explicit lexical knowledge, knowledge of grammar rules and assumptions about what language form and register is suitable for a given context. Truscott's view that the dominance of the learners' implicitly acquired L2 grammar cannot be influenced by conscious explicit language knowledge thus seems a little simplistic while Krashen's earlier notion of a complex interaction between different kinds of knowledge determining written output comes closer to reality. Moreover the support that Truscott provides for his contention that students rely on

their own intuitions ignoring teachers' correction is weakly based on an appeal to teachers' common experience and a highly selective reading of Cohen and Robbins' 1976 report on interviews of three ESL students (p. 248). It is true that in the 1976 study Hung, one of the students, said that in conversation he 'don't care the grammar' (p. 53). But he approached writing differently. He said that he experimented in order to see which forms were correct and then "directed his attention to those forms which had not been marked by the teacher" (Ibid.). Another student, Ue-Lin, reported that she reviewed her corrections and "sometimes kept a record of her errors" (p.55). The third student, Eva, reported that if she took time and went over her writing she was able to write English that was closer to standard (p. 58). None of them appear to match Truscott's characterization of how L2 students approach writing as they all very deliberately made use of conscious, not intuitive, self critical techniques to approximate target forms of the language by responding to corrective input from the teacher.

Thus, Krashen's theory as modified by Truscott is clearly inimical to error correction. But as shown above, Truscott's restatement of the theory is weakly based and misrepresents the L2 learner's process of writing. Yet, Krashen himself appears to endorse Truscott by citing his writings favorably (Krashen, 2004). Nonetheless, I contend that the earlier form of Krashen's theory (The Monitor) can still give limited support to error correction of L2 written texts.

The Output Hypothesis

Merrill Swain, as a result of her observation of native English speaking school children in French immersion programs in Canada, became convinced that output has a crucial role to play in second language acquisition. She did not reject the input hypothesis but argued it was not sufficient on its own. Thus she says: "...we wish to make the case that sometimes, under some conditions, output facilitates second language learning in ways that are different from, or enhance those of input" (Swain & Lapkin, 1995, p. 371).

As noted earlier, what led Swain to reconsider the role of output was the realization that years of comprehensible input appeared not to be sufficient to complete the process of acquisition (Swain, 1985, p. 246). She quotes one learner

who exemplifies the problem: “I understand everything anyone says to me, and I can hear in my head how I should sound when I talk, but it never comes out that way” (Swain, 1985, p. 248).

She hypothesizes that output has a noticing or triggering function, a hypothesis testing function and a metalinguistic function (Swain, 2005, p. 474). Error correction should be most operative to the hypothesis testing function for she suggests that learners experiment as they produce language, seeking confirmation for their hypotheses. She goes on to argue that if the learner’s hypothesis is wrong in a particular instance then negative feedback in the form of explicit correction, confirmation checks and clarification checks can make the discrepancy with target forms clear (Swain, 1985, p. 246). Here she is referring specifically to oral correction but it is reasonable to assume the same principles would apply to correction of written output.

In fact, written correction, though it may lack immediacy, has its own advantages. In large classes, individual students are not likely to be able to obtain much oral error corrective input, if any, from the teacher. But a student’s written output can be comprehensively corrected on a regular basis. Further, oral corrections can be lost in the immediacy of real time communicative activities while with correction of written work, the student has time to consider and reflect on the feedback received thus modifying subsequent output to more closely approximate native L2 forms. We can see an instance of this process in the three student drafts presented in Appendix 1 to Ashwell’s study of Japanese learners of L2 English (Ashwell, 2000, p. 248ff). Below, examination of the first sentences of the three drafts shows that the student is testing various alternatives in the process of redrafting, moving ever closer to target L2 forms (The corrections are Ashwell’s).

1. The typical father in Japan is very busy and they are swamped with their work. And they don’t help with house hold^.
2. The typical father in Japan is very busy and they are swamped with their work. and they don’t help with ^households.
3. The typical father in Japan is very busy and he is swamped with his work and he doesn’t help with a houseworks.¹

Thus Swain's hypothesis restores a place for error correction without fundamentally rejecting the primacy of input in L2 acquisition.

Explicit Language Learning

While Swain's approach remains closely allied to Krashen's, providing an extension to his input hypothesis, DeKeyser's observations on the processes involved in adult second language acquisition are radically different. He argues that access to the cognitive processes that allow children to acquire a second language implicitly is lost sometime during the teenage years (cf. The Critical Period Hypothesis) and in order to achieve a high level of competence in a second language at or after this period, adults make use of conscious analytical skills with explicit focus on form (DeKeyser, 2000, p. 520). Among the research elicited to support this argument are the well-known study by Johnson and Newport (1989) and his own replication of that study (DeKeyser, 2000). He reproduced the earlier study using a uniform ethnic group - Hungarians - whose non Indo-European language is typologically distant from English while modifying other aspects of Johnson and Newport to address a variety of criticisms of their study. These included the length of the questionnaire used to evaluate L2 learners' grammaticality judgments and length of immersion in the L2 environment which was doubled from the original study to a minimum of 10 years. Thus improved, DeKeyser's study nonetheless arrives at an even more definitive result than Johnson and Newport's. These studies confirm the validity of the Critical Period Hypothesis by demonstrating that long term immigrants to the United States who had arrived before the age of 16 acquired English to a near native level of competence by means of non-explicit, non-conscious learning mechanisms while those immigrants in the sample who had arrived after this age consistently failed to reach native speaker levels of competence as measured through grammaticality judgments. Only 3 of the 42 adult subjects in DeKeyser (2000) managed to approach near native levels of competence and these cases were shown to have a high aptitude for language as evidenced by highly conscious analytical language learning skills. DeKeyser argued that the younger group was able to access implicit mechanisms, i.e. innate universal abilities that are available to pre-critical period learners to acquire the second language but that the adults had lost much of

this natural ability. Only the relative few who were skilled in using analytical and explicit learning techniques were able to nearly match the younger group in achievement. His conclusion is that explicit instruction to adult L2 learners is fundamentally flawed if it denies the use of the only mechanism available to them to master the basic structures of a second language (DeKeyser, 2000, p. 520).

This does not mean that for DeKeyser, comprehensible input is not at all appropriate for older learners. A large amount of high quality input is still required for post critical period L2 learners. However, it cannot be claimed that this is sufficient to achieve a high level of L2 competence in adult learners. Implicit mechanisms are still perhaps the only means of acquiring some structures where rules are difficult to formulate. If DeKeyser is right, then explicit instruction is entirely appropriate and necessary for older learners. This would presumably not only permit error correction in the course of L2 instruction, but it would endorse correction with metalinguistic explanation.² Such an approach is exemplified in Sheen's (2007) research into acquisition of one aspect of article use where metalinguistic comment was seen to enhance the performance of one of the experimental groups in a study of error correction to be discussed in detail in a later section of this paper.

Putting aside Truscott's version of Krashen's theories, the three theoretical frameworks outlined above are, in essence, not contradictory. In fact, a learner operating with Krashen's Monitor Hypothesis in mind might not appear to be very different from one using the analytical skills recommended by DeKeyser. The modifications to a theory of implicit learning made by Swain and later DeKeyser, can be seen as responses to evidence of the inadequacy of immersion programs to achieve the results of young learners fully immersed in an L2 linguistic environment and the fact that adults do not have the same access to the language acquisition device as children. While DeKeyser's conclusion is unequivocal, Swain is not as definitive about the benefits of explicit learning mechanisms. Her use of recasts and clarification requests without metalinguistic comment implies that she still supports implicit learning techniques. All three permit, or in the case of the latter two, endorse error correction but see its role differently and require different means of application. Part of the variation in the application of explicit instruction would be determined by the learners' age and access to PLD and conscious learning mechanisms. It is not

difficult to see the first two theories as appropriate to an ESL situation with young learners. However, as already suggested in the introduction, it is very difficult to argue that Krashen's or even Swain's recommendations would be appropriate for older L2 students, particularly those in an EFL context.

The Studies

A comparison

Theoretical support for error correction does not prove it is effective in practice. Such proof would require empirical studies which make comparisons between experimental and control groups. To illustrate the different ways this can be done we will consider two studies that typify those done before 2001: Sheppard (1992) which focused on measurable change in students' ability to write accurately and Ashwell (2000) which examined modifications in one essay as students worked through its three drafts.

Sheppard (1992) had two comparable groups of intermediate students who received 35 hours of instruction which included grammar and literature study. Both groups wrote seven compositions and received either coded error correction or requests for clarification on each one. This was followed by a conference with the teacher where the feedback received on their essays could be further discussed with students. The essays were subsequently rewritten. First drafts of essays 1 and 7 were compared to measure overall progress. Sheppard was thus able to gauge students' improvement over 10 weeks depending on what kind of feedback they received. In the end, with respect to one variable, so-called "sentence boundaries", Group B, which had received clarification requests, performed significantly better than Group A which had received coded correction. This, it was argued, showed the superiority of holistic feedback over attention to form on L2 students' writing assignments.

Ashwell (2000) examined 50 Japanese students who had practice in sentence combining and instruction on "features of connected text". They also wrote diaries and completed four 500-word assignments. On the third assignment Ashwell divided the students into four groups giving them different kinds of feedback on two drafts of their essays. Two groups received first content and then form feedback or vice versa while one group received both kinds of feedback on both drafts. The control group

redrafted without feedback. He found content comment failed to improve the quality of the content as the experimental groups which had received content comment did no better than the control which received none. However, subsequent drafts of the experimental groups who had received error correction were more accurate and this trend was maintained through to the final draft. Meanwhile, the accuracy of the control did not improve though they did redraft. Thus error correction (underlining in this study) improved form accuracy but content quality ‘was much less affected by feedback’ (Ashwell, 2000, p. 243). In an apparent reversal of Sheppard’s result Ashwell found comprehensive error correction superior for improving subsequent drafts of the same text.

Table 1: Comparison of Sheppard(1992) and Ashwell (2000)

Study	Sheppard 1992		Ashwell 2000	
Context	ESL New York N =26 Duration: 7 essays over 10 weeks		EFL Japan N=50 Duration: 3 drafts of 1 essay over 3 weeks	
Treatment	Group A	Group B	3 Experimental Groups	Control 2 redrafts without receiving any feedback
	Error correction + Conference + Redraft	Content comment + Conference + Redraft	1. Content comment then form correction on 2 different drafts	
			2. Form correction then content comment on 2 different drafts	
			3. A mix of form and content feedback on both drafts	

It should be noted that there is a definitional problem in both studies in differentiating the two kinds of feedback that were provided to the students. In his paper, Sheppard (1992) is inconsistent in how he describes group (B) which did not receive coded correction. In the introduction, he writes that his study contrasted “two forms of corrective feedback” (p. 104). But in the discussion he contradicts himself saying that one group ‘didn’t get corrective feedback’ (p. 108). Then in the abstract he writes that one group received feedback on “content”. It is this last term that Truscott seizes on, arguing that as a content group they received “feedback only on meaning” (Truscott, 2000, p. 259). Because Group B outperformed Group A with respect to accuracy in sentence boundaries, Truscott thus seeks to use the study to argue that teachers should abandon form correction and respond only to the content of

a text. But it is simplistic to conclude that because Sheppard writes comments on students' texts that include the word "meaning" that they therefore address meaning or content. To illustrate, if I, a monolingual Anglophone, am spoken to in French, I might reply: "I don't know what you mean. Could you say it in English?" This response does not address the content of his statement but his manner of expression of that content. This is what we find when we examine what Sheppard has done in this study. He does not respond to what the student means but the fact that the student fails to express that meaning effectively. Thus he says the students in Group B received "requests for clarification" such as: "Could you say it in other words?" (Sheppard, 1992, p. 105). To compare terminology, before Sheppard 1992, "comprehension checks and clarification checks" and "explicit correction" had been categorized by Swain and others as forms of negative feedback that told learners to revise their output because it had not been understood (Swain 1985, p. 245). The implication is that clarification and comprehension checks are less explicit forms of correction. Sheppard's description, moreover, echoes that of Swain when he describes the effect of his comments which "force" the writers "back to the initial stages of composing..." where "they are shaping and restructuring their meaning" (Sheppard 1992, p. 108). What Swain describes as "pushing" and Sheppard as "forcing" is intended to help the learner notice a problem in the form of language output so that she or he will modify it in such a way that the meaning will be more successfully communicated.

If Sheppard had provided genuine content comments to students he should have addressed the meaning not simply the failure to communicate it effectively. He could have responded with commentary addressing their ideas, for instance by asking for illustrative examples, by suggesting rearrangement of ideas or by expressing his emotional reaction to what the student had written. We find nothing at this level in Sheppard (1992). This is not a criticism of Sheppard's correction technique in itself; it is merely an attempt to arrive at some definitional clarity. Clarification requests can indeed be pedagogically astute responses to student writing. Such requests are non-judgmental for they do not suggest the student has erred. They can provide motivation for the learner to clarify expression, to improve communicative efficacy. They also imply the instructor has some interest in what the student writes. But they

are not content comment in themselves. I would suggest therefore, that we can reasonably conclude that Sheppard's reference to two forms of "corrective feedback" in the introduction to this study was an accurate way to describe his treatment. If a text belonged to a Group A student, a language problem was marked with a code and if it belonged to a Group B student, it was annotated with a request for clarification. Thus, in this way the study compared coded correction with indirect correction.

So Sheppard's (1992) work does not provide support for Truscott's argument against the practice of error correction, and this becomes clearer when we consider his 1992 study in the light of his earlier work (Dicker and Sheppard, 1985). In the 1985 study the experimental group wrote drafts in a Modified Garrison Sequence where error correction is reserved for the last draft. The control received traditional comprehensive feedback on so-called "surface" errors. Thus, both received error correction. In the 1992 study I argue, we have a direct correction group and an indirect correction group. The table below shows both studies with the t-values indicating how each group improved or regressed compared to its own pretest score. The criteria are indicated in the table. ("Sentence boundaries" appears to refer to what are often called run-on sentences.)

Table 2: Comparison of Dicker and Sheppard (1985) with Sheppard (1992)

Sheppard's 2 studies on Error Correction	1985 Study – 36 ESL students over 5 weeks involving 4-6 drafts of work		1992 Study – 26 ESL students over 10 weeks involving 14 drafts of work for each group.	
	Experimental Group (Garrison Sequence including error correction)	Control Group (Traditional Correction)	Group A (Coded error correction +conference with rewrite.)	Group B (Requests for Clarification + conference with rewrite.)
% gain in correct verb forms: before/after	-1.01	-.09	3.77*	4.78**
% gain in correct sentence boundary markers: before/after	1.83*	1.96*	0.54	3.38*
% gain in ratio of subordinations to total number of sentences: before/after	-4.10**	-1.06	-2.34*	-0.54

[*p <.05; **p <.001 -the numbers represent t-values. Negative values indicate a regression.]

What strikes one about these results is their essential congruity. This is not surprising if we view them as four groups that have all received some form of error correction. The difference is that the length of the 1985 study was only half that of the 1992. The 1985 study shows no improvement in either group with respect to accuracy in verb forms after 5 weeks of treatment but after 10 weeks the subjects of the 1992 study show significant improvement. This accords with Ferris (2006) who also found a main effect of error correction was in the use of verbs. As far as run-on sentences are concerned, there is a general improvement – mostly significant when correction is given. The lack of significant improvement for Group A in the 1992 study may be because of a ceiling effect. Run-on sentences account for a small percentage of errors – only 3% in Ferris’s (2002) study, and they tend to cluster in just a few students.

The lack of a significant observable improvement for group A may result from the small sample size (13 students in each group), lack of random assignment, low frequency of error and the fact that less scope for improvement was available for Group A.³ On complexity, measured by ratio of subordinations to the number of sentences, there is an initial decline evident in the 5-week study but with longer treatment that decline appears to diminish, which argues against Truscott’s contention that error correction will result in loss of complexity. If that were the case, we would expect to see a more significant loss with longer error treatment instead of less.

Ashwell’s study was much more carefully planned and fully reported than Sheppard’s (1992) study and his content comments are, on the whole, responses to larger issues of organization and meaning. However there is still a problem with differentiating some content comments from his form comments. His annotation to a student that “You don’t say very much about how you think your father compares with a ‘typical Japanese father’” is, I would agree, an unambiguous comment on content. But when a student writes: “Japanese lady are very hard” and Ashwell underlines ‘hard’ and asks: “What do you mean?” he indicates a problem of lexical choice.⁴ This, I argue, is included in the term error correction, though at the word level, rather than content comment at the level of discourse. If a student selects a word whose denotation or connotation is inappropriate in the context this clearly constitutes a lexical error. Ashwell states that his content feedback was “aimed

principally at multiple-sentence level issues” but admits that there was “some cross-over with form feedback at the clause or sentence level” (Sheppard, 1992, p. 234). Having considered the issue he concludes that “generally, the distinction” (between the two kinds of feedback) “was maintained throughout” (Ibid.). This is not unreasonable. However, if the whole *raison d’être* of a study is to compare two different types of response to student writing it would seem to be essential to ensure a clear boundary be established between the two even if such separation were not possible in a normal teaching situation.

While there is a problem with the controls in the Sheppard and Ashwell studies they do nevertheless illuminate the process of correction. Taken together they suggest that error correction will have a dramatic effect on redrafts but that effect does not as easily translate into long term improvement in the student’s ability to write accurately. The effect will be more attenuated as we see in Sheppard (1992). Ashwell’s study also shows that content comment has little impact on complexity of content. Nonetheless, it may have an important role in showing students that the instructor is also paying attention to what they mean as well as to how they express it.

Other pre-2001 studies

The other pre-2001 studies, like Sheppard and Ashwell, generally have two groups of participants, one of which gets error correction and one of which receives some other kind of response from the instructor. Often the language used, and this is the case with Sheppard too, appears to suggest error correction is a response to a less important aspect of the text. Commonly, the writer will refer to correction as being focused on “surface error”. Generally, what would be “deep error” is left unspecified. It is probable that the distinction coincides with that made by Zamel (1985) when she argued that L2 instructors were too concerned with “surface-level features” and tended to ignore “larger issues of rhetoric and content” (p. 81-2). She had a point and it is noticeable that writers favoring error correction tend now to ensure that students also receive content comment (Chandler 2003; Ferris 2006).

The amount that students wrote in these studies is not large and what is normally the finding is that the error correction group has performed more poorly than the comparison group with respect to an aspect such as fluency or ideational content but

has failed to do any better on accuracy. The implication is that error correction offers no advantage but the alternative treatment does. Thus Semke (1984) found that students who were rewarded with higher marks for longer texts were able to write significantly more than the error correction group but with no loss of accuracy. Kepner (1991) found students who received comment on content had significantly more high level propositions in their work than a correction group but with similar accuracy. Sheppard (1992) was interpreted similarly. Frantzen (1995) had two groups, one of which had more grammar input, but got mixed results that were inconclusive overall. Finally, in a study by Polio, Fleck and Leder (1998) the comparison group received no feedback of any kind yet improved in accuracy as much as the correction group who also received extra grammar lessons.

If error correction works, why do these studies not show it? It is unlikely that comprehensive error correction will have a discernable effect if students do not write a large amount. A paper by Truscott and Hsu provides a case which may illustrate this point. Their 2008 paper appended three samples of one student's work. These included the first draft of a narrative that was corrected by the author and returned to the student for redrafting and the first draft of a subsequent essay. Truscott's point was that though the student had received error correction on the first paper he still made the same number of mistakes on the subsequent one. But the three samples can make a different point, namely that students need to write a considerable amount for comprehensive error correction to have an effect.

Table 3: Truscott and Hsu (2008) Comparison of 2 First Drafts Truscott

<u>Errors: Narrative 1</u> Total errors: 13 (152 words)	<u>Errors: Narrative 2</u> Total errors: 13 (143 words)
<u>Omissions:</u> 'a' x 2, 'about'	<u>Omissions:</u> 'for' x 2, 'to go', 'in', 'of', 'looking'
<u>Wrong word:</u> 'for' → 'about', 'hoped' → 'wished'	<u>Wrong word:</u> 'a' → 'some', 'found' → 'arrested', 'caught' → 'put'
<u>Wrong form:</u> 'a' → 'an', 'significance' → 'significant', 'chose' → 'choose', 'jointed' → 'joined'	<u>Wrong form:</u> 'rung' → 'rang' 'stealing' → 'to steal'
<u>Tense:</u> 'is' → 'was' x 3.	<u>Tense:</u> 'is' → 'were'
<u>Plurals:</u> 'gift' → 'gifts'	<u>Spelling:</u> 'favorit'

Table 3 above shows the errors in the first draft of essay 1 and of essay 2. Perusal of the two columns shows the student made quite different errors in essay 2. There is one partial overlap where the present tense of the verb “to be” was used where past tense was required.⁵ It was partial because there was a possible further complicating error in the misuse of a determiner (“pair of”) in the second essay. But apart from that one case there is no overlap in errors between the two essays. A new topic generated new vocabulary and new errors.

In 450 words (2 drafts of the first essay and one of the second) even if the student did learn from the first essay’s corrected mistakes, he had insufficient opportunity to demonstrate benefit from it in the second essay. And certainly we would not expect a student to remember every correction anyway. Where the error correction is not focused but comprehensive, it is likely that students will have to receive a considerable amount of correction on a large volume of writing before the effects of comprehensive error correction can be discerned. It also might be expected that students need to encounter correction of the same error more than once before the difficulty is overcome. Most of the pre-2001 studies in question have students write only one to two thousand words which is not likely to be sufficient to have much effect on accuracy. Semke’s correction group, for example, appears to have written only a total of 300 words in the first draft and 300 in the second. Studies which do show an effect, Chandler (2003) for example, have students write much more.

Other than Sheppard (1992), in which students did produce a considerable amount of written work, the only pre-2001 study where students wrote an adequate quantity was Polio, Fleck and Leder (1998) which found no value in feedback on L2 student writing. Thus, it is a study requiring attention. Sixty-five mostly Asian ESL students in Michigan University were randomly distributed into experimental and control groups. Having all failed to satisfy the university’s minimum English language requirement on the placement exam and failing to achieve the minimum Test of English as a Foreign Language (TOEFL) scores, they were required to undertake a remedial English course while continuing their university studies. The control group wrote four journal entries a week for seven weeks and received no feedback at all. The experimental group wrote two journal entries per week plus one redraft after receiving error correction on both entries. The experimental group also had a weekly

lesson in grammar and editing. On the post test both groups had significantly improved in accuracy measured as error-free T-units per total T-units. But neither group outperformed the other. As the control received no feedback of any kind it has to be assumed their improvement resulted from the sheer volume of writing practice itself with the probable confound of exposure to English in their classes and in their living environment. The experimental group did no better than the control, so in this case, it appears no measurable benefit was derived from correction or explicit grammar instruction vis-a-vis the accuracy of students' written assignments.

The results reported by Polio, Fleck and Leder (1998) are similar to those of a 1971 study done by Mason involving 25 students at the University of Hawaii. Mason also compared two groups of 24 foreign freshmen who had not met the university's English requirements and would normally have been required to take remedial English courses. In this case the requirement was waived for the experimental group and this permitted them to take additional courses in their major area of study. The control, however, were required to take a traditional sequence of EFL courses meant to assist them to cope with studying in an English environment. The two groups had achieved equivalent results in the English section of the university entrance test but further testing, done before treatment began, found the experimental group was significantly better in listening and writing. To eliminate these differences, the researcher drew matched pairs to form two smaller groups of six subjects each, a refinement that would enable him to better measure the effects of the treatment. After one semester, post-tests showed no significant differences between the two matched groups leading Mason to conclude that the experimental group had not suffered from their exemption from remedial English courses. Moreover, when the post-test results of the original larger groups were compared, the initially significant differences in writing and listening abilities were no longer evident.

Mason does not provide any figures on how either group improved compared to its pre-test scores but it is clear from his discussion and his statement that the 'mean gains' of the groups were compared, that both groups did improve (p. 21). Probably the large positive effect of initial immersion of these students in an English language environment largely obscured the impact that EFL courses may have had on the control and this would account for fact that the matched groups were equal on pre and

post tests. This may also have been the case with the Polio, Fleck and Leder (1998) study where the experimental group received grammar lessons and correction but did no better than the control which did not. It is not surprising that immersion in an L2 university environment, where the student cannot avoid receiving vast amounts of high level L2 input, has a very powerful initial effect in boosting L2 skills. Yet the Critical Period Hypothesis clearly implies that this effect will ultimately be insufficient to lift these mature students to native-like competence without the adoption of analytical and explicit techniques of language learning. This is supported by Scarcella's observations on the failure of American ESL students to achieve satisfactory competence despite receiving vast amounts of PLD (1996).

Post 2001 Comparison Studies

Studies in error correction by researchers whose alternative to the null hypothesis predicts correction to be beneficial have had problems in the provision of genuine control groups. Three post-2001 studies that have been able to do this provide positive evidence for error correction. There appear to have been no studies in this period that find against error correction as Truscott has noted (2007, p. 255). The first is a study in comprehensive error correction and the others are examinations of focused error correction.

Chandler (2003) had 32 high intermediate ESL conservatory students divided into two groups. Both groups received indirect correction. All errors in their first drafts were underlined and the papers returned to the students. The experimental group then corrected their errors and returned the papers to the teacher who checked their corrections, amending them if necessary. The control was advised not to attend to their errors until the experiment was finished. Later, after the experiment was complete, both groups did final drafts. The crucial difference in treatment was that the control did not correct their errors while the experimental group did. Both groups wrote the same amount over the duration of the experiment. The treatment involved five autobiographical essays, totaling an estimated 6,000 words for each group. The error rate for the experimental group dropped from 7.8/ 100 words to 5.1, significantly better than the control which did not improve, and with an effect size of 1.08 which was very large as Truscott conceded (2007, p. 269). Chandler repeated

the treatment the following year and again found a comparable reduction in error with her new students.

Truscott expressed skepticism and reinterpreted her data (2004). In his rebuttal he posits two effects, both harmful, to explain Chandler's results. First he argues that the control, who wrote a lot but declined in accuracy, must have been "harmed by correction" (p. 339). Otherwise, he argues, the work they did would have improved them. Here he ignores the fact that the decline was not significant and assumes that students must unfailingly improve in accuracy if they practice writing. The nature of the harm he suggests was done to the control group is vaguely sketched as "additional uncontrolled factors" (Ibid.). We can only guess as to what this might mean. Logically this unspecified harm must have affected both groups as both received correction. So having accounted for the lack of improvement in the control he must now find another factor to explain why, despite the initial harm done by correction to both groups, the experimental group nonetheless showed clear improvement. He suggests harm was also done in "the form of avoidance" (p. 338). Here he relies on a suggestion made in Sheppard (1992) that error correction provokes avoidance, i.e. in subsequent writing tasks, students somehow try to avoid using structures that were marked for correction in previous writing. Here, he ignores Robb, Ross and Shortreed (1986) whose error correction groups all showed rising measures of complexity, and explains away Chandler's second 2003 study where the holistic measure indicated no loss of complexity despite error correction. So Truscott speculates that in order to avoid the work of correcting their mistakes, Chandler's experimental group learned which structures they were likely to err in using to the extent that they avoided using them and so began to "simplify their writing" (p. 339). This is obviously implausible as it attributes a degree of analytical, recognition and recall skill which, if possessed by student ESL writers would be of great benefit in improving their writing ability, not diminishing it. Further, this harmful avoidance behavior was so effective that it not only counteracted the initial negative effect of correction on the experimental group, but created the "appearance" of improvement with respect to their own writing ($p = .001$) and with respect to the control ($p = .005$). Thus he asks his reader to ignore Occam's razor and entertain the paradox that the group that demonstrated measurable and significant improvement were in fact, doubly harmed. By contrast,

Chandler’s interpretation of her own results is neither fanciful nor strained. Two comparable groups received similar treatment. There is one difference – one group did not process its errors while the other did, and this latter group improved. She is entitled to argue that the group which significantly improved did so because it processed its corrections.

More recently, researchers have moved away from broad correction of every error (what I have termed “comprehensive” error correction) and turned their attention to focused correction of error, in particular mistakes in the use of definite and indefinite articles. One study by Younghee Sheen (2007) involved ESL students in America who were placed in three groups, two of which received treatment while the third, which constituted the control, participated only in the pre, post and delayed testing. The treatment consisted of students reading and then writing from memory short fables such as “The Fox and the Crow” and then receiving corrective feedback on their texts before reading and writing a new fable. This was done three times with one experimental group receiving direct feedback while the other received metalinguistic comment as well as the feedback. The language feature targeted was the use of indefinite and definite articles to denote first and anaphoric reference in the narratives e.g. “I saw a movie last night. The movie made me sad.”

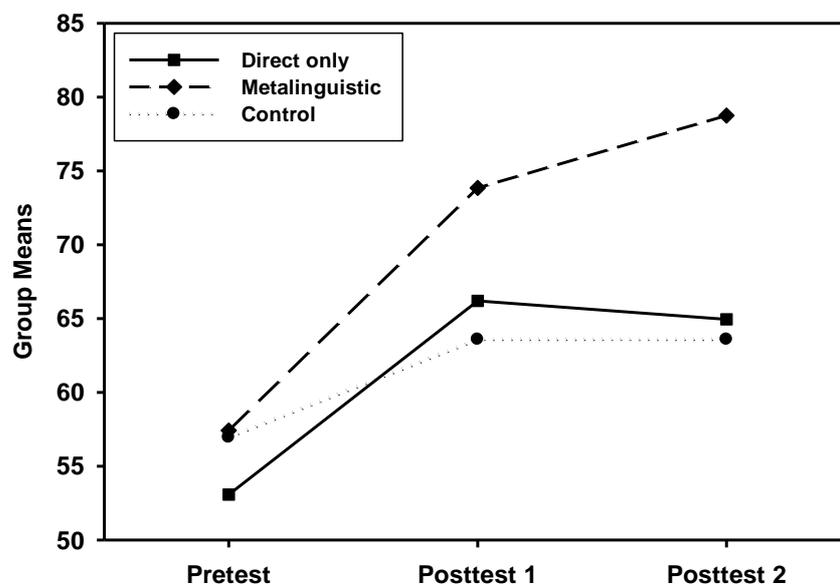


Figure 1: Group Means on Writing Test -Younghee Sheen, 2007

The pre, post and delayed tests included a speeded dictation test, a writing test and an error correction test. A further test measured the ability to analyze language. Results on the post tests favored the experimental groups over the control and the results from the delayed post tests indicated that significant longer term benefits were retained by the experimental group which received metalinguistic feedback but not by the experimental group which received only direct feedback. Here there seems to be clear evidence that error correction that is focused on specific grammatical features can benefit experimental groups. Moreover, the superior performance of the group which received metalinguistic feedback on the delayed post-test which was given 3 to 4 weeks after the post-test is persuasive evidence that explicit knowledge can provide longer term benefits. There was also found to be a fairly high correlation between the score on the test measuring language analysis ability and improvement in performance, a finding which echoes DeKeyser's contention that success in post-critical period L2 learning correlates with language aptitude. (The students were aged 20 or over.)

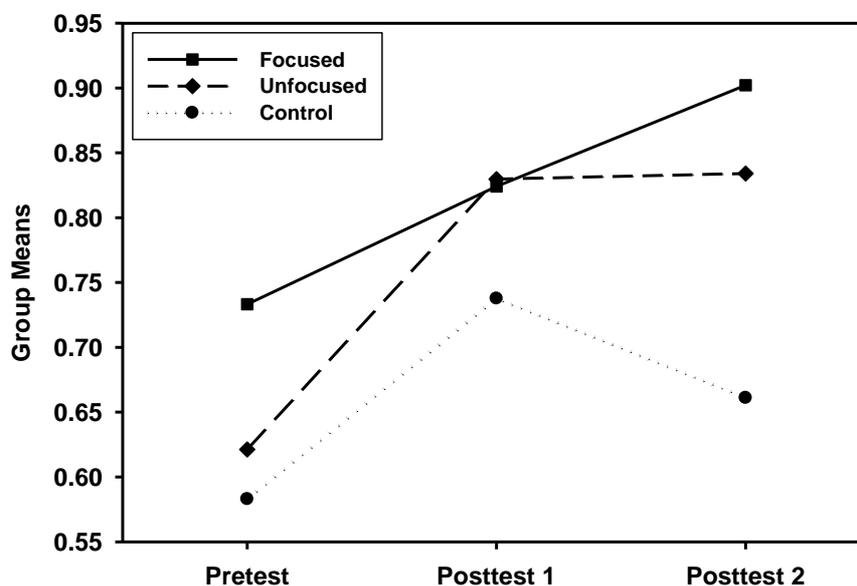


Figure 2: Means for narrative writing Scores– three groups – Ellis et al., 2008

This experiment was followed up with a study by Ellis et al. (2008) in Japan on the same feature of language – indefinite/definite article use. Again a control was contrasted with two experimental groups one of which received focused treatment

(only article errors were corrected) and one of which received unfocused treatment (other errors were also corrected). The experimental groups were significantly better with respect to the target variable than the control on the delayed post-test but the group that received focused correction did not outperform the group receiving unfocused correction. However, the trend in the experimental groups (note the rising trend line for the focused group between the 2 posttests) suggests that focused feedback ‘may be more effective in the long run’ (p. 367).

The importance of these results is that they provide evidence that focused correction can result in clear benefits when measured against a control thus rebutting Truscott’s longstanding criticism. However, all the papers of this sort have focused on just one limited aspect of English grammar, namely, one aspect of article use. It has yet to be demonstrated that the technique can be used with other language problems. If its scope can be widened, focused correction may prove to be a very effective way to increase the range and precision of error correction.

In summary, comprehensive correction of students’ texts is supported when students write large amounts. However, more focused correction or focused correction supported by metalinguistic comment may allow students to acquire some grammatical features of a second language in a more limited period of time and without requiring them to write large amounts of text.

Pedagogical Implications

What does the research evidence imply for the instructor in second language writing presuming he or she proceeds on the assumption that error correction is conducive to learning?

Which students should receive correction?

Van Patten argued 20 years ago that the evidence was ‘against any beneficial effect of a focus on form in the early stages of learning’ (1988, p. 255). This seems sensible when we examine texts like that below produced by one of Hendrickson’s students in 1979. We can see that comprehensive correction of such a piece of writing might produce an excessive amount of red ink. Not surprisingly most of the

research into corrective feedback has been done with intermediate and advanced students.

In the sommer Mr Smith and her wife leave the camping. They fishing in the lake. She lost her watch in the lake. She is sick. Mr Smith fishing one big fish. They walking at home. She craining. He said: 'No problem, I buy new watch.

Figure 3: Sample of one of Hendrickson's student's texts (1979)

However, the recent studies of Sheen (2007) and Ellis et al. (2008) using focused correction targeting article use may offer a means to help weaker students if it can be extended to other functional aspects of language such as understanding and acquiring the tense system on verbs. For example, students could be given a sample narrative with the title: "How I met my best friend and how we keep in contact now." It could be constructed to contain examples of verbs in the simple past and present tense. It would be read as a model by the learners who could then be asked to write their own stories on how they met their best friends. The teacher would correct the errors focusing only on verbs and tense. The students would then write their corrections above the errors. If a more competent student had no problem with verbs another aspect of the writing might be highlighted or the student could be given comprehensive correction. Two examples are given in Appendix 1, one where focused correction with metalinguistic comment was given to a weaker student and the other where comprehensive correction was considered appropriate for a student with greater proficiency. Thus feedback can be differentiated for students as Han recommends (2008).

How should instructors indicate students' errors?

Whether teachers should give direct correction (writing the correct form on the student's work), coded correction or indirect correction (highlighting the error by circling or underlining) has long been a topic of discussion. Lalande (1982) undertook a study with American students learning German to discover whether coded correction was superior to direct correction. He thought making students work

harder to decode his correction might be more beneficial than his writing the correction directly on their work. He found no advantage for coded correction except perhaps with respect to spelling errors. However, this is another study where students did not write much (perhaps 750 words on the first draft and 750 on the second) so it is not surprising its results were inconclusive. Chandler (2003), in her second study, compared four different kinds of feedback by rotating her students through all four types and tentatively concluded that direct correction and underlining appeared to be more effective than coded or marginal description and may be preferable to students as well.

Robb, Ross and Shortreed in their very careful study compared indirect, direct, coded and marginal correction (a number in the margin indicating the number of mistakes in a line) of Japanese EFL students' writing. At the end of a year of this kind of treatment they concluded that, "improvement was independent of type of feedback" (Robb, Ross and Shortreed, 1986, p. 91). One can see a slow but steady improvement in the students' accuracy with all four types of correction. (A decline in the scores of these students on the fourth test was attributed to a two-week holiday interruption after the third test.) Statistical analysis showed no significant differences in the effects of the four kinds of correction over the period of the study.

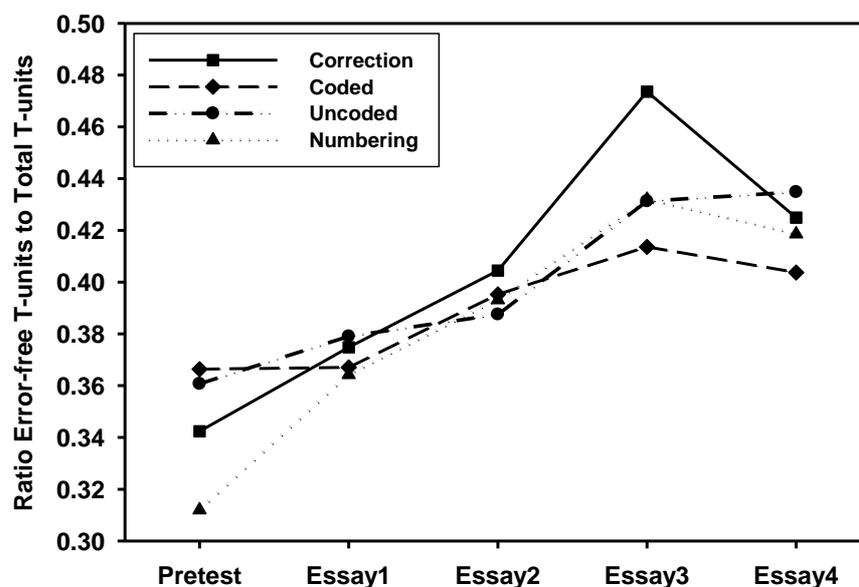


Figure 4: Improvement in Accuracy in Robb, Ross and Shortreed, 1986.

On balance, there is no conclusive evidence that any one form of correction is superior to any other for comprehensive correction. Though different reasons are given by researchers for preferring one or the other, the results do not clearly favor one over the others. However, from the perspective of the practitioner, the amount of time taken up by correction may be a deciding factor in the decision to use one or another method. Chandler found that underlining required .8 minutes per 100 words while other methods took .9 or a full minute per 100 words (2003, p. 289).

If students merely have errors highlighted or underlined, instructors may worry that students will not understand how to correct them. This is a potential problem. When one examines the three drafts of a student's essay in Ashwell (2000) it can be seen that in some cases the student fails to correct the error by the third draft and sometimes solves the problem by paraphrasing or deleting problematic passages. However, Ferris's 2006 study found the success rate for error correction on the second draft was 75% for highlighting (which is equivalent to underlining) which is only 13% less than for direct correction.

With focused correction the case may be different. Students' texts are often short in this kind of treatment and one experiment (Sheen, 2007) has found a clear advantage when a metalinguistic explanation accompanies correction. Generally, researchers write the correct form above the student's error but teachers could choose just to highlight errors, perhaps with a statement of the rule, and ask the students to write the correct form on their texts.

Rate of return for words written- Cost benefit considerations

The gains from error correction are difficult to measure. Figure 4 above shows slow and steady improvement in accuracy in the Japanese students who received error correction (Robb, Ross & Shortreed, 1986)⁶. But more than error correction was involved in the treatment. The roughly 10% improvement in the ratio of T-units to total T-units is comparable to the results obtained by Polio, Fleck and Leder (1998) where the students were in an ESL learning situation. In that study the correction group and the control group that merely practiced writing with no feedback at all, both made gains of about 10% on the same measure. Thus, with 35 hours of intensive EFL tuition over a year the Japanese students showed a similar improvement to that

made by ESL students fully immersed in an American university over seven weeks. It seems a not unreasonable result for a very form-focused EFL class. What is clear is that the students wrote a lot – a weekly essay that was corrected and then redrafted and five occasional essays that were used to test their progress. Assuming the weekly essay was about 200 words they would have written about 4,000 words on first drafts and another 4,000 on redrafts. This was on top of focused grammar work and sentence combining exercises.

In the first part of Chandler's (2003) study she had students write 5 autobiographical essays totaling 25 typed pages. This means that students wrote a total of 6,250 words on first drafts. The result was a 20% reduction in the error rate. In the second part of the paper Chandler notes that the following year's class was asked to increase the amount they wrote by 60% but there was no further reduction in the error rate which suggests that over a certain amount there is a diminishing rate of return for increased amounts written.

Ferris (2006) had students write 3,000 words on first drafts and another 6,000 on second and third drafts after receiving content comment and error correction. Results showed a statistically significant reduction in total errors at the .01 level and reduction of errors in the use of verbs at the .001 level. In Ferris's study, improvements in lexical choice approached significance ($p = .07$) but the improvement in errors on nouns, articles and sentence level problems was clearly not significant.

Thus the evidence suggests that comprehensive correction of whole texts can achieve significant results provided the students write in the region of 5,000 words a semester on first drafts but also that certain types of error are resistant to eradication. Improvements accompany large amounts of writing and consequently large amounts of correction. Is this modest benefit worth the effort? Could it not be attained by gradual improvement from more exposure to English through reading? The testimony of Scarcella (1996) suggests not. The thrust of her paper is that college level students in California who received massive input in L2 English and engaged in many communicative activities were seriously disadvantaged because their English was not near the standard required in a post-secondary educational context. She believed correction was one means for students to move closer to target forms required for

tertiary education. Many of the students in an EFL situation also expect to use English for higher education where they have to be able to read and produce accurate and complex texts. So, for students who hope to use English in tertiary education and professional activities, the price of the extra effort is probably worth paying.

Combining content comment with correction

There is some evidence as to how content comment and form correction should be incorporated. Zamel (1985) analyzed teachers' responses to ESL students' texts and found them "confusing, arbitrary, and inaccessible" (p. 79). She further complained that form correction and content comment were often given in a contradictory way. One way instructors have reacted to her critique has been by reserving a first draft of student writing for content feedback and then giving error correction on a second draft, a practice Ferris followed in the study reported in her 2006 paper. Fathman and Whalley (1990) and Ashwell (2000) investigated whether this was necessary by looking at how re-drafts fared depending on whether students had received form or content feedback first or a mix of the two. They measured accuracy improvement by the decline in the number of errors from first draft to second and content quality by using a holistic scoring guide. Fathman and Whalley's conclusion was that "Grammar and content feedback can be provided separately or at the same time without overburdening the student" (p. 187). Ashwell (2000) concurred. So, the evidence seems to indicate that teachers may well undertake both error correction and content feedback on one draft provided the latter does not become perfunctory. In a recent lecture by Zhao Hong Han (2008), following Van Patten (1988), it was argued that if teachers only respond to errors, students will think that instructors are not interested in what they have to say, only in how they say it. The ultimate result may be silence. Therefore, content comment may be important even if its effects are hard to measure.

Must students write multiple drafts for error correction to be effective?

With focused correction, redrafting seems not to be required. Researchers have merely asked students to examine their corrected texts (Ellis et al. 2008; Sheen 2007) and still achieved significant improvements on subsequent writing tasks. With comprehensive correction, a commonly held belief among instructors is that students

need to redraft to profit from the feedback that has been provided. In her first study Chandler (2003) provided compelling evidence that students need to process their errors for correction but not that they needed to redraft. It appears that the experimental group in that study had to write their corrections above those words and phrases that the instructor had underlined to indicate a problem and she later amended any corrections that were unsatisfactory. The control did not correct their errors and their accuracy failed to improve. The accuracy of the experimental group improved significantly on new tasks. Other researchers required full redrafting (Ferris, 2006; Robb et al., 1986; Sheppard 1992) but did not collect evidence showing redrafting was superior to merely having students correct their papers. We can conclude that while instructors may have good reasons to require redrafting of complete texts it is possible to improve L2 students' accuracy merely by having them amend their original drafts.

Conclusion

For a considerable period of time the value of error correction of L2 students' texts has been highly contested and its practice often neglected. As Scarcella (1996) recounts, the writings of Stephen Krashen were a contributing factor to this for he argued the practice was irrelevant to language acquisition. Professor Truscott went further insisting it was harmful and for a time, he largely dominated the debate. Gradually, however, testimony from teachers (Chandler, 2003, and Ferris, 2006) concerned more with pedagogy and evidence from researchers in the field (Sheen, 2007, Ellis et al., 2008, and others) have come to demonstrate that error correction is effective and its outcomes measurable. These efforts need to be consolidated both by replication and by widening the scope of focus. Chandler (2003) has made a persuasive case for comprehensive correction and bolstered her conclusions by repeating her treatment with a new intake of students. Sheen (2007), Ellis et al. (2008) and others have reported results which support focused correction albeit on only one aspect of the English article system. Replication of Chandler's results in a new context and expansion of Sheen (2007) and Ellis et al.'s (2008) work in other areas of grammar and syntax would make the case for error correction more compelling and could show instructors how correction might be better implemented.

End notes

1. Some errors on the first draft are not highlighted. Time for teacher correction was limited to 12 minutes so not all errors were found in one marking.
2. DeKeyser's support for error correction is implied in his 1993 paper and made explicit in his introduction to *Practice in a Second Language* (2007) where he states the literature shows feedback (including explicit error correction) has a "substantial positive effect" (p. 9). He notes that what remains at issue is exactly "when and how to give feedback" (p. 9).
3. It is true that the posttest mean for group A was measured against B's, and a t-test showed B was significantly better on the last essay on sentence boundaries. But the arguments given here, especially the lack of random assignment still hold. Further, a key issue was whether A and B were equivalent on "sentence boundaries" on the pretest and B was significantly better on the posttest or whether B was significantly better on a posttest weighted against the score on the same variable in the pretest. A simple t-test showing difference of means only on the posttest is not enough to show this.
4. The corresponding idea in Japanese is probably rendered by "tsuyoi" which could be translated as "patient" "strong" or "hard". The Japanese student mistranslated the word "tsuyoi" which sometimes may be equivalent to "hard" in English but here should be rendered idiomatically as either "patient" or "strong".
5. Bruton (2009) analyzed these same essays and considered the reduction in errors of tense was evidence of improvement on use of past tense. However, the particular tense error was in using present tense for something which was still probably true at time of writing but which should have been rendered in past tense in English: "In the bookstore they thought the storekeeper is cute...". This error did recur in the second essay on a similar occasion. ("...John picked up a pants which is his favorit style".) So there was no real improvement.
6. The Robb, Ross and Shortreed (1986) study, according to Truscott, showed the "futility" of error correction (Truscott, 1996). It is a surprising claim because all four treatment groups improved, though we do not know to what extent correction alone was responsible for the improvement. He seeks to justify his claim with two arguments. The first states that the fact that the four methods of correction had similar

effects despite the differences between them suggests that correction had no real effect at all. This is not a compelling argument for it may be that the different methods are all equally effective. What is lost in salience in more indirect correction may be made up for by having the students work harder to process their errors. Of course it is implausible to argue they must be exactly equal and examination of Figure 4 above does show a reasonable amount of variation in the effect of the four methods. His second and stronger argument is that because the fourth method of correction (numbers in the margin indicating the number of errors in a line) was so obscure, we can assume it provided negligible help to students in finding their errors. If so, this would allow the fourth group to be regarded as a control. Seeing they fared no worse than the other three groups they demonstrate that error correction is futile. His key assumption is that the fourth group would not be able to find their errors. However this assumption is not supported by the researchers. Tom Robb affirms that for ethical reasons they could not have a control receiving no feedback and he noted that the students in that group “were able to find many of their errors” (Personal communication, 7th June, 2009). Steven Ross supported Robb’s account and stated that the fourth group was “certainly not a ‘do nothing’ control” (Personal communication, 7th June, 2009). If the authors are right, Truscott’s argument is not.

References

- Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*(3), 227–257.
- A.U.A. (n.d.). AUA Thai Language Program, Retrieved June 20, 2010 from <http://www.auathai.com/>
- Bruton, A. (2009). Designing research into the effects of grammar correction in L2 writing: Not so straightforward. *Journal of Second Language Writing, 18*, 136-140.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267–296.

- Cohen, D. & Robbins, M. (1976). Toward assessing interlanguage performance: The relationship between selected errors, learners' characteristics, and learners' explanations. *Language Learning*, 26(1), 45-63.
- DeKeyser, R.M. (1993). The effect of error correction on L2 Grammar Knowledge and Oral Proficiency. *Modern Language Journal*, 7 (4), 502-503.
- DeKeyser, R.M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-553.
- DeKeyser, R.M. (2003). Implicit and explicit learning. In Doughty, C.J. & Long, M.H. (Eds.), *The handbook of second language acquisition*. Malden: Blackwell Publishing.
- DeKeyser, R.M. (2007). Introduction: situating the concept of practice. In DeKeyser, R. M. (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press: Cambridge.
- Dicker, S. & Sheppard, K. (1985). The effect of multiple drafts on structural accuracy in writing. *TESOL Quarterly*, 19(1), 168-170.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36, 353-371.
- Fathman, A, & Whalley, E. (1990). Teacher response to student writing: Focus on Form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge: Cambridge University Press.
- Ferris, D. R. (2002). *Treatment of error in second language student writing*. Ann Arbor, MI: University of Michigan Press.
- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues*. Cambridge: Cambridge University Press.
- Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal*, 79, 329-344.
- Han, Z. H. (2008). Error correction in Second Language Teaching. Towards a differential approach. A Queensborough Community College Seminar. Retrieved March 30, 2010

from,

<http://www.qcc.cuny.edu/Tigermedia/detailView.aspx?mediaID=853&CategoryID=9>

- Hendrickson, J. M. (1979). *Error analysis and error correction in language teaching*. Singapore: SEAMEO Regional Language Centre.
- Ho, C. M. L., (2008). Toward exploring errors in grammar: A systematic approach for language teachers. *TESL Canada Journal*, 252(2), 85-102.
- Hudson, R. (2001). Grammar teaching and writing skills: the research evidence. *Syntax in the Schools*, 17, 1-6.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal*, 75, 305–313.
- Krashen. S. (1981). *Second language acquisition and second language learning*. New York: Pergamon Press.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*, Retrieved June 20, 2010 from: http://www.sdkrashen.com/Principles_and_Practice/index.html
- Krashen, S. (1998). Comprehensible output. *System*, 26, 175-182.
- Krashen, S. (2004, November). *Applying the Comprehension Hypothesis: Some Suggestions*. Paper presented at 13th International Symposium and Book Fair on Language Teaching (English Teachers Association of the Republic of China), Taipei, Taiwan.
- Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal*, 66, 140–149.
- Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal*, 22(2), 1-16.
- Mason, C. (1971). The relevance of intensive training in English as a foreign language for university students. *Language Learning*, 21(2), 197-204.

- Polio, C., Fleck, N., & Leder, N. (1998). "If only I had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7 (1), 43–68.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83–93.
- Scarcella, R. (1996). Secondary education in California and second language research: Instructing ESL students in the 1990s. *The CATESOL Journal*, 9(1), 129-151.
- Semke, H. (1984). Effects of the red pen. *Foreign Language Annals*, 17, 195–202.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly* 41(2), 255-283.
- Sheppard, K. (1992). Two feedback types: do they make a difference? *RELC Journal*, 23, 103–110.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Gass, S. M. & Madden, C. G. (Eds.), *Input in second language acquisition*, Boston: Heinle & Heinle.
- Swain, M. (2005). The output hypothesis: Theory and research. In Eli Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371-391.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369.
- Truscott, J. (1998). Noticing in second language acquisition: a critical review. *Second Language Research*, 14(2), 103-135.
- Truscott, J. (1999a). The case for "the case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122.
- Truscott, J. (1999b). What's wrong with oral grammar correction. *Canadian Modern Language Review*, 55(4), 437-456.

- Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics*, 27(2), 93-108.
- Truscott, J. (2004). Evidence and conjecture on the effects of correction: a response to Chandler. *Journal of Second Language Writing*, 13, 337–343.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16, 255–272.
- Truscott, J. (2009). Dialogue: Arguments and appearances: A response to Chandler. *Journal of Second Language Writing*, 18, 59-60.
- Truscott, J. & Yi-Ping Hsu, A. (2008). Error correction, revision and learning. *Journal of Second Language Writing*, 17, 292–305.
- VanPatten, B. (1988). How juries get hung: Problems with the evidence for a focus on form in teaching. *Language Learning*, 38, 243-260.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19(1), 79–97.

Appendix 1

Examples of focused correction and comprehensive correction for Thai freshmen in Chulalongkorn University, Bangkok (Extracts)

1. Correction focused on tense with meta-linguistic comment. The student writes his or his corrections above the mistake. Alternatively a teacher could write in the correct forms and ask students simply to examine them.

You are writing about your memories. So it is better to use past tense here even if some of these things are still true now.

I studied in Phuket witayalai school. ~~This's~~ the best school in my city and in my opinion. In my class, We had 40 peoples. Everyone knows each other very well because this school teach the student to love school and their friends. I had 11 best friends...

We lived together in the class. When we studied, we sat together. .. The common activity that we ~~did~~ every finish class is play football. It was very fun and very tried... Some weekends we will travel around the city like cinema, beach. The sea was the most place that we went to... we usually ate seafood. It's very delicious...

Now, everybody in my group study in various universities. It's hard to meet but we call to each other last month. 3 my best friends call to me. we went to cinema and then went to have dinner.

2. For a more skilled student in the same class comprehensive indirect correction (here underlining) and content comment to be followed by redrafting seems more appropriate.

"Gug, gug", the noise of a pen falling onto the ground startled me ... again. And I always picked it up and gave it back to its owner. Yes, the pen was not mine. It belonged to a girl who sat in front of me. At that time, I never imagined this girl would become a friend who I could not forget... But thanks to this clumsiness of her, it lead us toward the wonderful friendship....

The falling pencil is an interesting way to introduce your friend. It catches the reader's attention and tells us a little about her character.