

An analysis of lexical errors in the English compositions of Thai learners

SAENGCHAN HEMCHUA – Srinakharinwirot University, Bangkok

NORBERT SCHMITT – The University of Nottingham

ABSTRACT

The importance of vocabulary in second language (L2) writing is widely accepted, but there has been relatively little research into the lexical errors learners produce when writing in their second language. Moreover, the error categorisation frameworks used in some previous studies have addressed only a relatively limited number of lexical error categories. In this paper the authors used a more comprehensive error taxonomy based on James (1998), with some additions from Leech's semantics (1981), to analyse Thai third-year university students' English compositions for lexical errors. The analysis revealed that (a) 'near synonyms' were the most numerous errors, followed by 'preposition partners' and 'suffixes', (b) the students had more difficulty with semantics than the forms of words, and (c) the identified sources of errors were mainly from L2 intrinsic difficulty rather than the first language (L1) transfer. The findings from the Thai students' written lexical errors have implications for L2 vocabulary teaching and learning, which should be of interest to wider English as a Second Language (ESL)/English as a Foreign Language (EFL) contexts.

Research into lexical errors

The language produced by foreign language (FL) learners almost inevitably contains errors of various types. This is the process of learning a language. Empirical evidence suggests that lexical errors are the most frequently occurring category of errors in written English (for example, Grauberg 1971; Meara 1984; Lennon 1991). They significantly affect the quality of academic writing (Astika 1993; Ellis, R 1994; Engber 1995) and native speakers consider them the most irritating (Santos 1988). Moreover, Carter (1998: 185) suggests that 'mistakes in lexical selection may be less generously tolerated outside classrooms than mistakes in syntax'. This is probably because lexical selection consists mainly of content words, which convey the intended message. When inappropriate lexical choices are made (lexical errors), this can lead directly to misunderstanding of the message, or at least to an increase in the burden of interpreting the text. Since lexical errors are potentially so disruptive, they

deserve attention. This paper will first present a taxonomy of lexical errors in written language, and then use this taxonomy to analyse the errors that Thai university students made when writing English compositions. Although the Thai language is quite different from some other languages, we believe the error taxonomy and methodology we present can be applied to the analysis of lexical errors in the academic writing of other L2 learners, and that our findings have implications for vocabulary learning and teaching in a wide range of language contexts.

Despite the frequency and seriousness of lexical errors, they remain under-researched. This may be attributable to the fact that they are complicated; for example, open system (many items and not rule-based [McCarthy 1990]). Among the few studies that have been undertaken in this area, Duskova (1969) found four types of lexical errors when she analysed the writing of 50 Czech postgraduate students: confusion of words with formal similarity, similar meaning, misuse of words caused by one or several equivalents between Czech and English, and distortions among lexical nonce mistakes (for example, throw <through>, solve <dissolve>). These findings are congruent with those of Henning (1973), who found that FL learners stored FL vocabulary in connection with form and meaning. This study showed that beginners tended to make form-based associations, while more advanced learners tended to make meaning-based associations. Laufer (1991) provided further empirical evidence on FL lexical confusion where there was similarity between the lexical form of target words and errors, that is, 'synformic confusions'. These are words with the same root but different suffixes or prefixes, words identical in all phonemes except one vowel, diphthong or consonant, a vowel or a consonant present in one synform but not in the other, and words identical in consonants but different in vowels (see the 'Formal errors' section for further explanation and examples). Laufer (1990) also found that L1 and L2 learners encountered a similar hierarchy for difficulty in lexical forms, with the most difficult or confused synformic distinction being suffix synforms (for example, *considerable/considerate*). Thus we find that appropriate L2 morphological use is not always an automatic process (Jiang 2000), and morphological errors in L2 production seem to be common among learners. For example, L2 learners normally do not know all derivative forms in a word family (Schmitt and Meara 1997; Schmitt and Zimmerman 2002).

There is not always a linear relationship between proficiency and number of errors. Sometimes the greater fluency of advanced learners seems to lead to an increased number of lexical errors in their productive use (Martin 1984). Synonym errors are a case in point, and advanced learners at university level

seem to have considerable problems with synonyms and productive vocabulary choice involving style, syntax, collocation and semantics. Taylor (1986) suggests that synonym or near synonym errors may be the consequence of error-avoidance. Students may adopt a word that is synonymous with the target word because they are more confident about using that word. Therefore, students' choice of near synonyms can also provide evidence of avoidance as a writing strategy (see Sonaiya 1991 for exercises to promote the knowledge of semantically related items). Collocation, the way in which words are used together regularly, requires intuitions about lexical patterning, and these stem from grammar, semantics and register and may even constitute a discrete type of knowledge (Halliday 1966; Sinclair 1991). It probably takes a great deal of exposure to language to acquire this knowledge, and even advanced learners may take extended periods of time to master collocational appropriacy, as many studies have revealed that EFL learners are generally deficient in collocation knowledge (for example, Bahns and Eldaw 1993; Farghal and Obiedat 1995; Granger 1998; Howarth 1998; Schmitt 1998).

Lexical errors may be influenced by the writer's L1 (Swan 1997) or from difficulties stemming from the words themselves (Laufer 1997). Corder (1992) comments that the role of the mother tongue may play a role in the cognitive process of L2 language learning and language use, whereas Schachter (1992: 32) considers the process of transfer as 'evidence of a constraint on the learner's hypothesis learning progress'. However, Scholfield (1987) regards L1 transfer as 'communication strategies' when the learners decide to express their required meaning (for example, false friends, loanwords, and cognates). Regardless of the precise nature of the process, there is some empirical evidence that the native language influences target lexical learning (for example, Zughoul 1991; Ard and Homburg 1992; Broselow 1992; Jarvis and Odlin 2000). In addition, intralingual errors resulting from the intrinsic difficulty of words (the difficulty resulting from the intrinsic properties of target words related to their form, meaning and usage) have been found (for example, James et al 1993; Tinkham 1993; Waring 1997).

In an attempt to gain more insight into the various types of lexical errors and their frequency, this study explored the error types of Thai university students. The following research questions were addressed:

- 1 What lexical errors do the third-year Thai university students make in their English compositions?
- 2 Which of the errors are most frequent?
- 3 How many of the errors are attributable to L1 transfer?

In the process of answering these questions, a lexical error taxonomy was developed, which we feel will be useful for both teachers and researchers.

Method

PARTICIPANTS

The participants were 20 Thai English majors in their third year of study at a university in Bangkok. They had been taught EFL for approximately ten years for three to five hours a week. Thai was the only language they spoke at home, as well as at school, with the exception of the English classes. Overall, the participants were similar in age, ranging from 19 to 20 years old, but factors such as sex and age were not controlled in this study.

The participants had had little previous English-writing experience in their primary and secondary schools, since writing is not emphasised at these levels. They enrolled on a basic writing course in the first year at the university and took a paragraph-writing course in the second year. The basic writing course focused mainly on grammatical structures. For the paragraph-writing course, the objectives were to develop students' writing abilities beyond sentences and also to familiarise them with various styles of writing, such as description, narration and argumentation. Students were systematically guided to improve their writing skills, including rhetorical stance and rhetorical pattern, and the conciseness, precision and logical flow of their writing. They also were encouraged to use complex sentences in their writing and guided to improve in the areas of syntax, word choice, organisation and writing conventions. However, explicit vocabulary teaching was not emphasised in the course. Besides their formal instructional input, participants were sometimes exposed to English language through the Internet, English movies and other media. However, they rarely associated or communicated with English native speakers outside the classroom. Although no standardised test results were available for all of the participants, some students took the TOEFL (Test of English as a Foreign Language) and the scores ranged from 450 to 550. The participants were able to give English oral presentations of between five and seven minutes in class, have brief, general conversations with native speakers, write English formal letters, and compose 300–500 word descriptive and argumentative essays on general topics within two hours (with some errors).

PROCEDURE

To examine the type and extent of lexical errors in students' written English, the participants were asked to write an argumentative composition of about 300–350 words, without consulting their dictionaries, within one-and-a-half

hours on a topic related to the advantages of urban or country living. The mean length of the 20 compositions was 345.30 words (SD = 81.43, min = 218, max = 578). The limited number of participants (20) allowed for an in-depth analysis of each of these essays.

In order to determine what types of lexical errors were made by the participants, their compositions were corrected by two experienced native-English teachers. The teachers were found to agree on the identification of almost all of the errors. Subsequently, the first author (a native speaker of Thai) categorised the errors into the error framework described below, consulting with the second author and an experienced teacher of English for Academic Purposes (both native speakers of English) in cases where the classification was problematic. The lexical errors were classified into 24 sub-categories under two main categories: formal and semantic features. It took some time to identify and classify the errors, and this unfortunately precluded the possibility of interviewing the participants retrospectively regarding the errors and their intended meanings.

Each error was assigned to one category, although in some cases more than one categorisation was possible. In these ambiguous (relatively few) cases, the three judges discussed the best classification, relying on the first author, who has a Thai L1 background, to identify any obvious L1 influence. The following sentence is an example of L1 influence that was missed by one of the English native-speaker raters, but recognised by the L1 rater: *There are many comfortable things to do everything such as machines, buses, trains, airport, etc.* 'Comfortable things' was considered by the L2 rater as a collocation error in 'semantic word selection'. The subject intended to describe life in the city as comfortable because of the facilities. 'Comfortable things' was actually translated from Thai words that are equivalent to 'facilities'. Therefore, it was finally classified as a 'calque' (L1 translation).

Some errors proved rather difficult to classify with confidence, as they could be plausibly placed into more than one category. For example, *You will wake up in the morning because of voice's bird <the sound of bird/bird's song>*. 'Voice's bird' was assigned to 'connotative meaning error' because the student used 'voice' rather than 'noise', which indicates that she wanted to suggest more than just a sound. However, there we could also have categorised this error into 'near synonyms', since 'voice' and 'noise' can be used interchangeably in some contexts.

It was also occasionally difficult to decide whether errors were lexical or syntactic in nature. For example, *What's a <the> matter?* 'A' was considered as a 'collocation error' rather than a syntactic error because, in this case, it is not a matter of a definite versus indefinite article. The point is that 'What's

the matter?' is a fixed phrase. Therefore, it indicates that the student failed to acquire this fixed phrase. This study was confined to lexical errors, therefore the following syntactic errors were disregarded.

- 1 Phrase structure errors (for example, *There is an air pollution*).
- 2 Clause errors (for example, *Furthermore, it's not difficult for getting <to get> to a hospital*).
- 3 Sentence errors (for example, *At first, I didn't think how kind they were <that they were kind>*).
- 4 Intersentence errors or cohesion (for example, *When someone wants one's help, he <they> will help each other*).

In contrast to syntactic errors, which involve more than word structure, morphological errors are confined to word structure and therefore cannot be ignored when lexical errors are discussed. In this study, however, only derivational affix errors (for example, prefixes and suffixes) were included. In the sentence, *Bangkok is pollution <polluted>*, the student showed her knowledge of the word, but she could not use the proper derivative form. On the other hand, inflectional affix errors (for example, plurality, genitive, tense, third-person singular, comparative and superlative) are mainly governed by grammatical processes and were excluded from this analysis.

Lexical error classification

Previous studies on lexical errors have used a variety of error classifications, most with a relatively limited number of categories. For example, Duskova (1969) used only four categories of lexical errors. Similarly, Engber (1995) used a system with nine categories. Given the current understanding of the complexity of lexis, the use of compact classification systems to explain learners' errors can result in unclear boundaries and arbitrary classifications, while the lack of specificity in the definition of error categories may obstruct the analysis from its full potential. Though overlap between categories is an ongoing problem in the field, overall we feel that a more comprehensive categorisation framework can contribute to more precise identification and discussion of error types. Our expanded framework for lexical error classification was mainly drawn from James's (1998) lexical error taxonomy, which was compiled from various sources from previous studies. It was modified to include two types of 'meaning' by Leech (1981). James classifies lexical errors into two main categories: formal and semantic features. This approach is based on the classic word knowledge framework suggested by Richards (1976), that is, seven types of knowledge are necessary to know a

word: (a) morphology including pronunciation and spelling, (b) syntactic behaviour in a phrase and a sentence, (c) functional or structural restrictions or collocations, (d) semantic values, (e) secondary meaning or connotations, (f) word association and (g) frequency of use. For the reader's information, the error categories are described below in some detail.

FORMAL ERRORS

Formal errors are classified into three types: (1) formal misselection, (2) misformations and (3) distortions (James 1998). The sub-types and examples of each type are as follows:

1 Formal misselection

The first four sub-types are based on James (1998), who in turn drew from Laufer's (1991) 'synformic confusions'. They involve similar lexical forms (visual and sound similarity). James refers to them as the mala-propism type. Synforms share some phonemes/graphemes. The four main types of synforms are:

- 1.1 The suffix type. They have the same root but different suffixes (for example, *considerable/considerate*, *competition/competitiveness*).
- 1.2 The prefix type. They have the same root but different prefixes (for example, *reserve/preserve*, *consumption/resumption/assumption*).
- 1.3 The vowel-based type (for example, *seat/set*, *manual/menial*).*
- 1.4 The consonant-based type (for example, *saves/safe*, *three/tree*).

A fifth sub-type was drawn from our Thai data:

- 1.5 False friends caused by divergent polysemy, partial semantic overlap, or loan words that have been taken from English words and which sometimes have meaning overlaps (for example, Thai 'bank' = bank/banknote). Occasionally, the meanings are divergent (for example, Thai 'serious' = stressed).

2 Misformations

These are words that do not exist in the L2. The source of errors is from the learner's mother tongue. They are, therefore, called 'interlingual misformation errors'. James (1998) classifies misformation errors into three types as follows:

- 2.1 Borrowing. L1 words are used in the target language without any change (for example, *I shoot him with gun in kopf* <In German *kopf* = head>.*

- 2.2 Coinage. Inventing a word from L1 (for example, *Smoking can be very nocive to health* <In Portuguese *nocivo* = harmful>).*
- 2.3 Calque. Translation of a word or a phrase from L1 words (for example, *We have to find a car to bring us go to* <bring us to> the hospital).

3 Distortions

These words also do not exist in the L2. However, the errors are the result of misapplication of the target language without L1 interference or misspelling. James (1998) classifies distortions into five types as follows:

- 3.1 Omission (for example, *intresting* <interesting>).
- 3.2 Overinclusion (for example, *dinn~~ing~~ room* <dining room>).*
- 3.3 Misselection (for example, *delitouse* <delicious>).*
- 3.4 Misordering (for example, *littel* <little>).*
- 3.5 Blending (for example, *travell* <travel + travelled>).

Note that L1 inference (interlingual errors) and interference within L2 (intralingual errors) can be at play at the same time in the same error. That is, L1 effects may lead to L2 effects.

SEMANTIC ERRORS

James classifies semantic errors in lexis into two main types. Their sub-types and examples are as follows:

1 Confusion of sense relations

Psycholinguistic evidence suggests that humans store words in terms of sense relations in their mental lexicon, at least to some extent (for example, Deese 1966; Aitchison 1987). Vocabulary meaning normally involves concepts and their relations in lexical fields (for example, a *woman* and a *girl* belong to the lexical field of 'gender'). The following four main types of errors are classified accordingly.

- 1.1 Using a superonym for a hyponym. A more general term is used where a specific one is needed. Therefore the meaning is under-specified (for example, *We have modern equipment* <appliances> in our house).
- 1.2 Using a hyponym for a superonym. An overly specific term is used (for example, *The colonels* <officers> live in the castle).*

- 1.3 Using inappropriate co-hyponyms (for example, *I think the city has good communication <transportation/public transport> such as a lot of buses*).
- 1.4 Using a wrong near synonym (for example, *a regretful <penitent/contrite> criminal or sinner*).

2 Collocation errors

Collocation is a word or phrase that is frequently used together with another word or phrase and sounds natural and correct for native speakers. Inappropriate collocation may not be absolutely wrong, but rather infelicitous. James specifies the following three degrees of the misuse of collocation.

- 2.1 Semantically determined word selection (for example, *The city is grown <developed>*).
- 2.2 Statistically weighted preferences (for example, *An army has suffered big losses <heavy losses is preferred>*).
- 2.3 Arbitrary combinations and irreversible binomials (for example, *hike-hitch <hitch-hike>*).

Apart from the above three types of collocation, James also mentions Leech's (1981) associative meanings. One type of stylistic-meaning infelicity he identifies is 'verbosity' (for example, *I informed my girlfriend of the party through the medium of telephone*).

In addition, L2 learners sometimes do not convey sufficient meaning in their writing. That is, the sentence is too brief and the meaning is unclear. This underspecification can also be due to poor choice of words (for example, *Although cars in the country are lower <Although there are fewer cars in the country/Although car numbers in the country are lower>*). Both of these categories (James's 'verbosity' and the authors' inclusion of 'underspecification') were added to the taxonomy as semantic errors.

In order to form a more comprehensive taxonomy of lexical errors, elements from Leech's (1981) semantics were also included because meaning is central to communication. Therefore, 'meaning' inevitably plays a role in distinguishing semantically odd lexical choices from those that are more appropriate. Drawing on a broad sense of all meaning communicated by language, Leech identifies seven types (conceptual, connotative, social, affective, reflected, collocative and thematic meaning). From the seven types of meaning, only connotative and collocative were incorporated into the present error taxonomy. Connotative meaning (for example, *There are too*

<many> other advantages of living in the city) seems to add something new and also covers conceptual meaning. Collocative meaning had already been counted by James, but not all the aspects were covered. Thus we added 'preposition partners' (verb or noun + particular preposition, preposition + noun) (for example, *some channels in <on> television, surrounded with <by> nature*). Leech's other meaning categories are beyond the scope of this study (for example, thematic meaning relates more closely to syntactic errors; social meaning and affective meaning can be grouped into 'near synonyms'; and reflected meaning is not related to error, as it reflects reader or listener feeling or reaction).

Overall, the present error classification consists of two main categories: formal and semantic errors. A summary can be found in Figure 1.

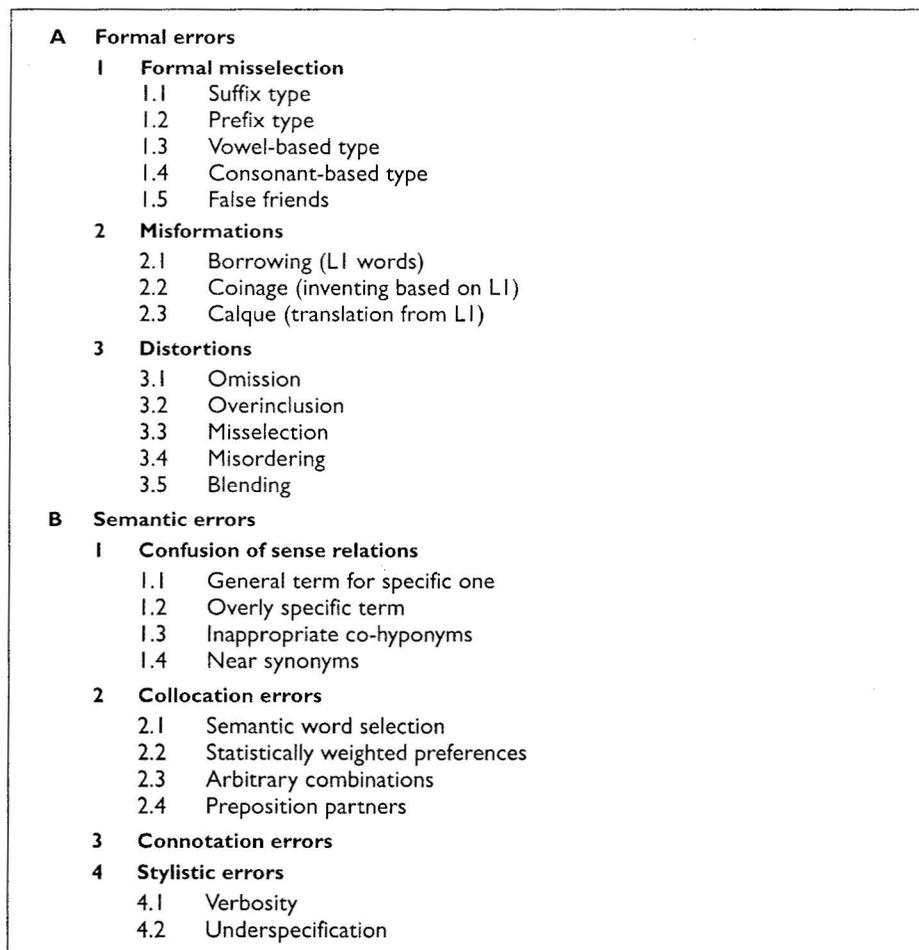


Figure 1: Summary of lexical error taxonomy

Error count

We used the above error framework to gain a quantitative overview of the lexical errors in the participants' writing samples. This entailed counting errors connected with words. However, lexical errors are not always restricted to individual words. For example, inappropriate collocations are seen as lexical errors, but operate at the phrasal level. In making our error count, individual cases of lexical error were counted at the word, phrasal and sentential levels:

- 1 Individual lexical items (for example, *But country people are sincere and share <generous>*).
- 2 Word combinations
 - 2.1 Two lexical items (for example, *They will travel by cart or by legs <on foot/by walking>*).
 - 2.2 Phrases (for example, *He said that my mother and he had mutual consensus <reached a mutual consensus>*).
 - 2.3 A whole sentence (for example, *Every time that I hitch-hiked, it will be a car of country people <I was picked up by a car driven by country people>*).
- 3 Multiple errors in a phrase were counted separately (for example, *In my own idea <In my opinion>*).
This was counted as two errors, 'opinion' for 'idea' was one error, and 'in my opinion' for 'in my own opinion' was another error.
- 4 Identical errors made by the same student were counted as one error. This is because the study is an attempt to explore which errors were common across the students, and the repeated occurrences of a particular error from the same student would exaggerate the error frequency.
- 5 To qualify as identical errors, both the erroneous form and the likely target form had to be identical (for example, *In the city, we can find important or famous places, famous and good schools, department stores, theme parks and the electric instruments <appliances> | They find the electric instruments <appliances> to help them*). The phrase 'the electric instruments' of these two sentences were identical errors and were counted as one error since they were made by the same student.

It is important to stress that there is some overlap between error categories and that precise categorisation was not always possible. Likewise, attributing the source of errors is imprecise and sometimes it is impossible to be certain

of the reason(s) for a particular error. Nevertheless, it is often possible to establish the source, particularly if it involves L1 transfer. We examined the lexical errors for any sign of L1 influence, to compare the relative influence of L1 transfer versus intrinsic difficulties with the words themselves.

Results and discussion

LEXICAL ERRORS IN THIRD-YEAR THAI UNIVERSITY STUDENTS' COMPOSITIONS

As shown in Table 1, the 20 compositions on the same topic by different individuals yielded 261 lexical errors. That is, on average, each paper contained 13.05 errors, which equates to one error per 26.46 running words. Some error types were common (for example, near synonyms, preposition partners and incorrect suffixation), others were relatively infrequent (for example, prefix type, false friends, blending and inappropriate co-hyponyms) and there were no occurrences of others at all (the categories of vowel-based formal errors, L1 borrowing, coinage and using overly specific terms contained no errors). This indicates errors are not evenly distributed across the error-type spectrum; rather, certain error types appear to be particularly problematic.

Formal errors

As far as formal errors are concerned, the formal misselection of words was the most problematic error category in the data (15.33% of all errors), followed closely by intralingual 'distortions' (14.56%). L1-influenced errors ('misformations') were less of a problem (6.90%). This finding suggests that the similarity of form and parts of speech (for example, verb, noun, adjective and so on) remained a serious problem in the Thai students' writing. The possible reason for the formal misselection is lack of knowledge of words in a word family (for example, Schmitt and Meara 1997; Schmitt and Zimmerman 2002). When the meaning and the spelling of an intended word were acquired, knowledge of the other members of a word family or grammatical patterns may have been incompletely acquired, which might result in a wrong derivative or grammatical pattern being used in a particular context (for example, *the people who live in the country are honesty <honest>*).

Although 'distortions' or misspellings were the second most frequent formal errors, they occurred only slightly less frequently (as 14.56% of the total errors) than 'misselections'. Thus misspellings were a problem for the Thai students, and it is therefore important to consider why the participants misspelled even short or monosyllabic words, for example, 'frist' <first> and 'shools' <schools>. Raimes (1985: 247) suggests that such misspellings occur

Table 1: Frequency of lexical errors

Error types	No of errors (Total=261)	%	No of papers containing the error (N=20)	%
A Formal errors				
1 Formal misselection				
1.1 Suffix type	24	9.20	12	60
1.2 Prefix type	1	0.38	1	5
1.3 Vowel-based type	0	0.00	0	0
1.4 Consonant-based type	12	4.60	9	45
1.5 False friends	3	1.15	2	10
2 Misformations (interlingual)				
2.1 Borrowing (L1 words)	0	0.00	0	0
2.2 Coinage (inventing)	0	0.00	0	0
2.3 Calque (translation)	18	6.90	12	60
3 Distortions (intra lingual)				
3.1 Omission	14	5.36	8	40
3.2 Overinclusion	6	2.30	6	30
3.3 Misselection	10	3.83	7	35
3.4 Misordering	5	1.92	4	20
3.5 Blending	3	1.15	3	15
B Semantic errors				
1 Confusion of sense relations				
1.1 General term for specific one	11	4.21	7	35
1.2 Overly specific term	0	0.00	0	0
1.3 Inappropriate co-hyponyms	3	1.15	2	10
1.4 Near synonyms	51	19.54	16	80
2 Collocation errors				
2.1 Semantic word selection	10	3.83	6	30
2.2 Statistically weighted preferences	9	3.45	5	25
2.3 Arbitrary combinations	16	6.13	10	50
2.4 Preposition partners	33	12.64	15	75
3 Connotative meaning	11	4.21	7	35
4 Stylistic errors				
1.1 Verbosity	17	6.51	10	50
1.2 Underspecification	4	1.53	4	20

because unskilled L2 writers 'concentrate on the challenge of finding the right words and sentences to express their meaning instead of editing'. Although 'distortions' may not significantly confuse readers in relation to communicative meaning, they do create a bad impression.

Semantic errors

Of the four main semantic types of error, 'collocation errors' were the most frequent (26.05% of the total), followed by 'confusion of sense relations' (24.9%), 'stylistic errors' (8.04%) and 'connotative meaning' (4.21%). Together, errors in the categories 'collocation errors' and 'sense relations' accounted for about half of all errors in the study. Overall, semantic errors were roughly twice as frequent as formal errors (63.22% versus 36.78%), as illustrated in Table 2. Likewise, the importance of semantic errors is illustrated by the fact that the single most frequent type of error, 'near synonyms', occurred in the 'confusion of sense relations' category.

Table 2: Summary of frequency in formal and semantic errors

Formal errors		Semantic errors	
No of errors	%	No of errors	%
96	36.78	165	63.22

These results suggest that formal errors were less problematic for the students than semantic errors. However, they made up a considerable percentage of the total errors, and this implies that the students would benefit from developing their morphological and formal knowledge, as Schmitt and Zimmerman (2002) suggest. Semantic knowledge might be more difficult to acquire, since various word knowledge facets (for example, sense relation, collocation, connotation and register) are required. It can be argued that these aspects are among the more 'advanced' elements of word knowledge, and that learners are not likely to master them until later in the acquisition process (Schmitt 2000: 117–18). If so, these results might indicate that most of the Thai learners in this study had largely mastered the formal aspects of the words they know productively (except perhaps the suffix type of synformy), but still had not mastered all of the semantically-related aspects. In other words, their mastery over semantic types of word knowledge was insufficiently developed.

ERROR FREQUENCY

Table 3 ranks the 24 error types in terms of frequency. As can be seen from this table, 'near synonyms' was the most frequent type of error (19.54% of

the total errors) and it also had the most papers containing this error type (80% of the total). The second and the third most frequent errors were 'preposition partners' and 'suffix type', which accounted for 12.64% and 9.2% of the total respectively. If we add the errors made in these three types together, they account for over 40% of the total errors.

Table 3: Rank-order frequency of lexical errors

Error types	No of errors (Total=261)	%	No of papers containing the error (N=20)	%
1 Near synonyms	51	19.54	16	80
2 Preposition partners	33	12.64	15	75
3 Suffix type	24	9.20	12	60
4 Calque (L1 translation)	18	6.90	12	60
5 Verbosity	17	6.51	10	50
6 Arbitrary combinations	16	6.13	10	50
7 Omission	14	5.36	8	40
8 Consonant-based type	12	4.60	9	45
9 General term for specific one	11	4.21	7	35
10 Connotative meaning	11	4.21	7	35
11 Misselection	10	3.83	7	35
12 Semantic word selection	10	3.83	6	30
13 Statistically weighted preferences	9	3.45	5	25
14 Overinclusion	6	2.30	6	30
15 Misordering	5	1.92	4	20
16 Underspecification	4	1.53	4	20
17 Blending	3	1.15	3	15
18 False friends	3	1.15	2	10
19 Inappropriate co-hyponyms	3	1.15	2	10
20 Prefix type	1	0.38	1	5
21 Vowel-based type	0	0.00	0	0
22 Borrowing (L1 words)	0	0.00	0	0
23 Coinage (inventing)	0	0.00	0	0
24 Overly specific term	0	0.00	0	0

Near synonyms

The inappropriate use of near synonyms can be grouped into three cases as follows:

- 1 The use of informal words instead of formal ones. For example: *We can communicate with people and get <gain/acquire> knowledge from other*

countries by using computers. 'To get knowledge' is not entirely incorrect, but the use of 'get' seems more appropriate in informal writing. The student appears to be unaware of this register choice when engaging in formal written language.

- 2 The meaning of the synonym used and the appropriate synonym were not exactly identical. Thus the intended meaning was not expressed by the synonym used. For example: *You will get up <wake up> in the morning because of the sound of birds.* The intended meaning of the underlined words in the context was 'to become awake after sleeping' not 'to leave the bed'. Therefore, 'wake up' was required.
- 3 Two words were close in meaning but were different in usage. For example: *Because in the city has <there are> many hospitals.* To refer to something for the first time and to refer to a quantity as 'there are' is common usage. The student, in fact, did not mean 'to own or possess something'. Thus 'has' was used erroneously in this context.

Most 'near synonym' errors appear to result from the intrinsic difficulty of the words themselves (38 out of 51 errors). Often this had to do with the semantic features of words (29 errors): *For me, I can live <stay> in the country only for relaxing because this life style is very boring.* In this case, the student could not distinguish the difference between 'stay' (be in the country only for a short time as visitor) and 'live' (have her home there). Other errors had to do with the specific usages of words (seven errors), and the overuse of shorter, high-frequency words (two errors).

L1 transfer was identified as the cause of 13 out of 51 near synonym errors, mainly due to 'divergent polysemy' (where a single Thai word corresponds to two or more words in English). In writing, this resulted in an error when the students applied their knowledge of Thai and then translated it into English synonymous but the use and meaning were not the same: *We will have good chances to get a good work <job>.*

Our analysis suggests that near synonym errors have variable causes (including register restriction: formal and informal situations, different underlying meanings and overlapping meanings), but that L1 transfer had a relatively minor effect. This finding is consistent with Zughoul (1991), who found that assumed synonymy was the most frequent error type in Arabic-speaking learners' English compositions.

Preposition partners

As noted above, the second most frequent error was 'preposition partners', which accounted for 12.64% of the total errors. In English only particular

prepositions can be used with verbs, nouns and adjectives. Changing the preposition changes the meaning and therefore results in an error. These errors referred to the omission of a preposition where it was required, the addition of a preposition where it was not required and an incorrect substitution for a preposition. For example:

Omission: *Also, you have many things to do and to think <think about>.*

Addition: *I don't have to face up to <face> the traffic congestion.*

Substitution: *There are no traffic jams that result of <from> having many cars.*

Among these three incorrect uses of prepositions, substitution (18 errors) was the most frequent, while omission (nine errors) and addition (six errors) occurred less frequently. This indicated that the students were conscious of using prepositions, but that they had not mastered which correct prepositions should accompany particular verbs, nouns or adjectives. Additionally, the Thai students were sometimes not aware that a particular preposition was required and this resulted in errors of omission. Note that prepositions seem to be difficult to acquire even among native speakers. Young native speakers take years to master the similar, but not identical, meanings of words like 'near' and 'next to' (Durkin et al 1985).

Suffix type

Suffix errors consist of the formal misselection of words consisting of syn-forms: (a) the same word class with similar form (for example, *There is a lot of competition <competitiveness>*) and (b) the wrong use of a particular word class (for example, *It is said that today our world is globalisation <globalised>*). It was found that 19 out of 24 suffix errors were due to the misselection of a particular word class, while only five errors arose from the confusion of similar forms. This suggests that the students had more problems in using words with the right word classes or derivative forms (for example, noun, verb, adjective and adverb) than the similarity of forms within the same class. This supports Jiang's (2000) view of L2 lexical development where morphological knowledge is not composed in the lexical entry until learners have reached stage three (when the semantic, syntactic, morphological and formal specifications of a L2 word are integrated and formed in the lexical entry). In addition, Laufer (1991) found that the initial part of words was stored correctly in FL learners, but the final portion of words frequently confused learners.

One source of suffix errors is potentially the wide use of bilingual

(English–Thai and Thai–English) dictionaries, particularly pocket dictionaries. For example, ‘compete’, ‘competition’ and ‘competitive’ are translated as a single word in Thai. Though the particular word class is given, there is no explanation or examples of how to use these word classes in context. The students, therefore, tend to memorise a core word or a frequently used one and apply it in every context, even where it is not the one required. Potentially, the wrong use of word class is grammatically driven. However, nouns, verbs, adjectives and adverb are also lexical words, which deal with word structure and the constraint of word structure resulted in suffix errors.

L1 TRANSFER IN LEXICAL ERRORS

It must be stressed that identifying the underlying cause of errors is inexact and problematic. Nevertheless, in many cases a likely cause can be identified. With this in mind, we calculated that 23.75% of the errors could be reasonably attributed to L1 influence. This should be seen as a conservative estimate, as some L1 influence undoubtedly remained hidden, and so was not identified in our count. Nevertheless, our estimate strongly suggests that the Thai language was not the major factor in the lexical errors that the Thai learners made. Rather it was largely the intrinsic properties of the English words (abstractness, specificity and register restriction, multiplicity of meaning, similarity of lexical forms, sound-script incongruence, inflectional and derivational complexity, long words, and low-frequency words occurring infrequently in textbooks – see also Ellis, N C 1997) that prevented the learners from deploying the words appropriately and resulted in the lexical errors. This lack of adequate vocabulary knowledge can become a major hurdle to acquiring and using language (Nation and Meara 2002). This result mirrors many findings from L2 grammar studies, where L1 influence was not found to be the major source of error (see Ellis, R 1994 for one overview).

Although L1 transfer was not the major source of error, it was still the cause of a considerable number of errors. Errors of L1 transfer mainly involve phrases or clauses (mostly not with a single word). Accordingly, the sentences tended to be longer and redundant. Also, the Thai language structure (topic-prominent language) was frequently adopted when writing in English (for example, *For myself, I think there are many interesting things*).

Implications

Although this study focuses on Thai learners, the lexical errors discussed are likely to be problematic for a wide range of L2 learners. Our findings suggest the following implications for L2 vocabulary teaching and learning in general.

The students had considerable difficulty in semantics, particularly near synonyms and preposition partners. This may not only apply to Thai students, but also to other L2 students, as the results are consistent with previous findings (for example, Lennon 1991; Zughoul 1991). Martin (1984) suggests that to avoid misunderstanding between new words and already learned synonyms, teachers should (a) identify stylistic level (formal/informal, colloquial/slang and so on), (b) provide examples in grammatical frames, (c) give their most common collocates and (d) supply additional information of semantic distinctions (hyponym/superordinate, animate/inanimate and so on). Sonaiya (1991) points out that learners normally know the likely target lexical items, even when they apply the erroneous ones. This is due to over-generalisation and the failure of distinctive awareness. She therefore suggests exercises to promote the knowledge of related lexical item differences by teaching them in context with meaning structure instead of as individual, isolated items. However, introducing words in the same semantic field may also cause some degree of confusion (cross-association) if done so concurrently (Higa 1963; Tinkham 1993; Waring 1997).

Apart from near synonyms and preposition partners, the students had problems in suffix use. As mentioned in Jiang (2000), morphological specifications are not automatic. Classroom teaching should be designed to raise students' awareness of this issue.

Although L1 transfer may not be the main source of errors, it may be beneficial to point out some differences between L1 and L2. These include divergent polysemy (a single word in L1 corresponds to two or more words in L2), lack of equivalence in formality and style (for example, respectful and intimate styles), cultural differences (for example, abstract concepts), phonological difficulty (resulting from a learner's L1 system, which might cause confusion in L2 words) and false friends (using L1 false friends in a L2 context).

Vocabulary learning strategies can be effective and should be introduced into the classroom so that the students can find their own ways to cope with problems. For university students from intermediate level onwards, students can be trained to use dictionaries effectively. For example, using monolingual dictionaries of synonyms (Rutledge 1994) to distinguish near synonyms is one way; looking up parts of speech is another. The use of a word-part strategy (Nation 2001) or breaking words into parts (affixes and stems) will raise students' awareness and consciousness about appropriate parts of speech, particularly derivational suffixes in productive use (Schmitt 1994). Using corpora, which are now accessible online or for sale (for example, the British National Corpus, the COBUILD Corpus), is a powerful way to gain

awareness of collocations, particularly preposition partners (see Thurstun and Candlin 1998 for academic writing and concordancing, and Wichmann et al 1997 for more implications of corpora in language teaching and learning). It might even be useful for students to memorise unanalysed chunks (for example, the chunk 'subject to'), either writing each chunk on a small card or repeating it aloud, in order to become more accurate in productive use (Nation 2001).

Although the percentage of lexical errors made by learners in each category may vary according to L1, degree of cognateness and proficiency, we believe that the error results of the Thai students reported here may be indicative of the lexical problems of other L1 learners. The results of this study can therefore usefully inform ESL/EFL teachers working with learners from many different backgrounds and can help in highlighting the range of lexical errors that need attention, particularly in the light of our finding that the majority of the lexical errors were not L1-specific. We hope that the error categorisation system can help teachers and researchers make more sense of the vast range of lexical errors, and would like to believe that the categorisation system could also be used with students. As it stands, the taxonomy is probably too complex for students to employ, but, with experience, teachers could discover the error types that their particular students are prone to, and present a shorter taxonomy to their classes. Such a student-specific taxonomy could prove to be a useful tool for awareness-raising of, and instruction in, lexical errors.

NOTE

- * These examples are taken from James (1998). All other examples in the article are taken from the Thai data in the study.

ACKNOWLEDGMENT

Our thanks go to the anonymous reviewers for their detailed and insightful comments on an earlier draft of this paper. We would also like to thank Sue Roberts for her input into the error classification.

REFERENCES

- Aitchison, J. (1987). *Word in the mind*. Oxford: Blackwell.
- Ard, J., & Homburg, T. (1992). Verification of language transfer. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language teaching* (pp. 47-70). Amsterdam: John Benjamins.
- Astika, G. G. (1993). Analytical assessment of foreign students' writing. *RELC Journal*, 24, 61-72.

- Bahns, J., & Eldaw, M. (1993). Should we teach EFL collocation? *System*, 21, 101–114.
- Broselow, E. (1992). Transfer and universals in second language epenthesis. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language teaching* (pp. 71–86). Amsterdam: John Benjamins.
- Carter, R. (1998). *Vocabulary: Applied linguistics perspectives* (2nd ed.). London: Routledge.
- Corder, S. P. (1992). A role for the mother tongue. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language teaching* (pp. 18–31). Amsterdam: John Benjamins.
- Deese, J. (1966). *The structure of associations in language and thought*. Baltimore: John Hopkins.
- Durkin, K., Crowther, R. C., Shire, B., Riem, R., & Nash, P. R. G. (1985). Polysemy in mathematical and musical education. *Applied Linguistics*, 6, 147–161.
- Duskova, L. (1969). On sources of errors in foreign language learning. *International Review of Applied Linguistics*, 7, 11–36.
- Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–139). Cambridge: Cambridge University Press.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139–155.
- Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics*, 33, 315–331.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology* (pp. 145–159). Oxford: Clarendon Press.
- Grauberg, W. (1971). An error analysis in German of first year university students. In G. Perren & J. Trim (Eds.), *Applications of linguistics*. Cambridge: Cambridge University Press.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robin (Eds.), *In memory of J. R. Firth*. London: Longman.
- Henning, G. H. (1973). Remembering foreign language vocabulary: Acoustic and semantic parameters. *Language Learning*, 23, 185–197.
- Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 2, 170–175.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology* (pp. 161–186). Oxford: Clarendon Press.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. New York: Longman.

- James, C., Scholfield, P., Garrett, P., & Griffiths, Y. (1993). Welsh bilinguals' English spelling: An error analysis. *Journal of Multilingual and Multicultural Development*, 14, 287-306.
- Jarvis, S., & Odlin, T. (2000). Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition*, 22, 535-556.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21, 47-77.
- Laufer, B. (1990). Sequence and order in the development of L2 lexis: Some evidence from lexical confusions. *Applied Linguistics*, 11, 281-296.
- Laufer, B. (1991). Some properties of the foreign language learner's lexicon as evidenced by lexical confusions. *International Review of Applied Linguistics*, 29, 317-330.
- Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.
- Leech, G. N. (1981). *Semantics* (2nd ed.). Middlesex: Penguin Books.
- Lennon, P. (1991). Error and the very advanced learner. *International Review of Applied Linguistics*, 29, 31-44.
- Martin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal*, 68, 130-137.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Cramer & A. Howatt (Eds.), *Interlanguage*. Edinburgh: Edinburgh University Press.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P., & Meara, P. (2002). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 35-54). London: Arnold.
- Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19, 229-258.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77-89.
- Rutledge, H. (1994). Using a dictionary of synonyms. In P. Nation (Ed.), *New ways in teaching vocabulary* (pp. 195-196). Virginia: TESOL.
- Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaker students. *TESOL Quarterly*, 22, 69-90.
- Schachter, J. (1992). A new account of language transfer. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language teaching* (pp. 32-46). Amsterdam: John Benjamins.
- Schmitt, N. (1994). Word family practice. In P. Nation (Ed.), *New ways in teaching vocabulary* (p. 182). Virginia: TESOL.

- Schmitt, N. (1998). Measuring collocational knowledge: Key issues and an experimental assessment procedure. *International Review of Applied Linguistics*, 119–20, 27–47.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19, 17–36.
- Schmitt, N., & Zimmerman, C. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36, 145–171.
- Scholfield, P. (1987). Vocabulary problems in communication: What determines the learner's choice of strategies? *Banger Teaching Resource Materials in Linguistics*, 1, 56–75.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sonaiya, R. (1991). Vocabulary acquisition as a process of continuous lexical discrimination. *International Review of Applied Linguistics*, 14, 269–284.
- Swan, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 156–180). Cambridge: Cambridge University Press.
- Taylor, G. (1986). Errors and explanations. *Applied Linguistics*, 7, 144–166.
- Thurstun, J., & Candlin, C. N. (1998). *Exploring academic English*. Sydney: NCELTR.
- Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, 21, 371–380.
- Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, 25, 261–274.
- Wichmann, A., Fligelstone, S., McEnery, T., & Knowles, G. (1997). *Teaching and language corpora*. London: Longman.
- Zughoul, M. R. (1991). Lexical choice: Towards writing problematic word lists. *International Review of Applied Linguistics*, 29, 45–60.