عَرِينِي
Analysis of Lexical Errors in Saudi College Students’ Compositions

Nadia A. Shalaby, Noorchaya Yahya and Mohamed El-Komi*

Abstract
Research on lexical errors made by second/foreign language student-writers is scarce in comparison to research in other problematic areas in writing, such as grammar. More studies are needed to expand/modify the relatively limited number of lexical error categories, and to explore the sources of such errors. This study aims at presenting a comprehensive taxonomy capable of accounting for Saudi EFL students’ lexical errors. It examines the types of lexical errors produced by female Saudi students studying English as part of the requirements of the preparatory year at Taibah University in Al-Madinah Al-Munawwarah, Saudi Arabia. The study addresses the following research questions: a) what types of lexical errors are common in the writings of female EFL students studying in a preparatory year program at a Saudi university? b) Which of these lexical

* Dr. Nadia Abdul Galil Shalaby is Lecturer of Linguistics in the English Department, Ain Shams University. She holds an MA from the American University in Cairo, and a Ph.D. in linguistics from the University of South Carolina. Her areas of interest are sociolinguistics, discourse analysis and language learning. She has published research on textual strategies in Arab American literature, Critical Discourse Analysis, and peer revision in composition classes. nashalaby@yahoo.com

Dr. Noorchaya Yahya is an Associate Professor, fulfilling her sabbatical leave from Florida Atlantic University, Florida, USA, by teaching EFL methods courses and coordinating the English Language Center at Taibah University, Madinah, Saudi Arabia. Co-author of two TESOL books and several journal articles. Noorchaya’s research interests include TESOL education of US pre-service and in-service teachers, second language writing and second language learning strategies. noorchayayahya@yahoo.com

Dr. Mohamed Fouad El-Komi is Assistant Professor, Department of Languages and Translation, College of Arts, Taibah University, Madinah, Saudi Arabia. He is the co-founder and supervisor of the Writing and Recourse Center, English Language Center, Taibah University. He completed a Master's in Linguistics and a Doctorate in curricula and instruction, both at the University of South Carolina, Columbia, South Carolina. mkomii20@yahoo.com
errors are most prevalent? c) Which of these errors are due to the influence of the students’ first language? The results of the analysis of 96 writing samples used for this study yielded a total of 718 lexical errors, with an average number of 7.48 errors in each essay. The wrong choice of a suffix was the highest category of errors, occurring 128 times (17.83%) followed by direct translation from L1, occurring 113 times (15.74%). In general, formal lexical errors were fewer than semantic lexical errors; 39.55% compared with 60.45%. The researchers also discuss the pedagogical implications for the teaching of vocabulary for second/foreign language learners.

1. Introduction

Just how important is the correct choice of lexical items? Perhaps the following signs from around the world can provide the answer.

- Outside a Hong Kong tailor shop: “Ladies may have a fit upstairs”.
- In a Rhodes tailor shop: “Order your summer suit. Because is big rush we will execute customers in strict rotation.”
- In an East African paper: “A new swimming pool is rapidly taking shape since the contractors have thrown in the bulk of their workers”.
- In a Vienna hotel: “In case of fire do your utmost to alarm the hotel porter”.

(The Network: Spring 1995)

Words constitute the building blocks of language. When appropriately chosen, they help language structures hold together and help language learners convey their intended meaning. If, on the other hand, they are improperly and inaccurately selected, they lead to the confusion and segmentation of meaning. These lexical errors, which can in many cases hinder communication, are less tolerated by readers as compared to syntactic errors (Carter, 1998). Also Vahallien and Schoonen (1989) attest to the importance of lexical knowledge in learning and communication in a foreign language. They even state that lexical knowledge is one of the most important factors in academic success. Many L2 teachers find that their students’ wrong choice of words can be quite amusing. We agree. However, what is problematic, according to Zughoul (1991), is that while the wrong lexical choice often leads to the production of funny utterances, they are not easily comprehensible. ESL/EFL teachers as well as researchers are confronted with the daunting task of dealing with their students’ wrong lexical choices, and often they do not know where to begin. Furthermore, lexical errors are quite
widespread. Dealing with such a problematic area first requires that teachers become aware of the source and nature of these errors as this will help them understand the cognitive processes leading to these errors. Such awareness would in turn help teachers in addressing these problems in class. An understanding of the nature of lexical errors calls for the employment of a well founded and comprehensive taxonomy which acknowledges the highly complex nature of the process of ESL/EFL students’ selection of words when writing. This constituted the motivating factor for the researchers in this study to reconstruct the taxonomy proposed by Hemchua and Schmitt (2006). The proposed taxonomy would not only enable ESL/EFL researchers to examine word choice errors but also to classify such errors, investigate their source and eventually suggest teaching implications which could be helpful and effective for ESL/EFL students and teachers.

Research into EFL/ESL lexical errors, with a few exceptions, remains scanty, in comparison with research in other problematic areas of writing, and usually focuses on a limited number of error categories. Many researchers point to this fact as for example Zughoul (1991), Sheshsha (1993), Llach (2005), Hemchua and Schmitt (2006) and Tang, Q. (2006), among others. Moreover, most research on lexical errors neither acknowledges the complexity of these errors nor does it take into account the inherent overlap among them. In comparison research in syntactic errors, for example, deals with a much more clearly defined area, with specific categories that are easier to identify and classify than categories of lexical errors. Due to the significant gap in research in lexical errors, the researchers decided to conduct this study.

2. The Objective of the Current Study

The objective of the current study is to examine the lexical errors Saudi EFL female college students make while writing, adapting Hemchua and Schmitt’s (2006) somewhat comprehensive taxonomy and employing it to classify and quantify the number of lexical errors in the students’ writings, and analyze the sources of these errors. The study, more specifically, attempts to answer the following research questions:

a) what types of lexical errors are common in the writings of female EFL students enrolled in the preparatory year program at Taibah University?

b) Which of these lexical errors are most prevalent?

c) Which of these errors are due to the influence of the students’ first language?
3. Research on Lexical Errors

Most of the studies reported in this section, with a few exceptions, attempted to account for their subjects’ lexical errors using a limited number of error categories. They, therefore, miss on a large number of errors that do not fit their proposed set of lexical error categories, and are in this respect deficient.

Duscova (1969), for instance, examined the writings of fifty Czech postgraduate students using only four categories of lexical errors: confusion of words with formal similarity, similar meaning, misuse of words and distortions among lexical nonce mistakes (throw: through). The categories used by Duscova are obviously too broad. For example, the category “misuse of words” can in fact incorporate many subcategories.

Also using a very limited taxonomy, Akande, Adedeji, and Okanlawon, (2006) stated that their 225 students’ lexical errors fit under only three categories: overgeneralization of rules, wrong analogy and wrong spelling. In their study, they investigated the lexical errors of 225 final year technical college students. A 20-item fill-in multiple choice test and an essay writing exercise were administered to the students. The researchers found that technical college students did not have high competence in the use of words related to their different areas of specialization as they normally made mistakes resulting from overgeneralization of rules, wrong analogy and wrong spelling. They suggested that English teachers in Nigerian technical colleges should tailor their teaching of English to reflect the lexical needs of their students. By failing to provide a comprehensive, detailed taxonomy, however, the study collapses the many different types of errors under a few headings and therefore overlooks many specific types of sources of lexical errors.

Research on the lexical errors in the writings of Arab students has given mixed results, with some researchers finding developmental errors to be more pervasive than L1 transfer errors and vice versa. With few exceptions, researchers have mainly used a limited taxonomy, or focused on a particular aspect of lexical errors. Analyzing the writings of Arab students and working with quite a comprehensive lexical error taxonomy, Zughoul (1991) argues that a detailed typology of EFL college student writers can be helpful in giving “a better understanding of interlanguage and insights into the strategies employed by language learners for lexical choice” (p.46). Zughoul presents a typology of 13 lexical errors. According to his findings, assumed synonymity is the most common type of word choice errors made by his subjects (23.5% of the total number of errors). He states that “first
language interference is a major variable in lexical choice as it takes forms of assumed synonymity, derivativeness, literal translation, and idiomaticity” (p.56).

In another study investigating the lexical errors made by 48 Saudi university student-writers majoring in English, Sheshsha (1993) presented an error classification of five categories, confusion of words with formal similarities, confusion of words with similar meaning, inappropriate collocation, literal translation, and divergence. Sheshsha then collapsed the five error types into two major types, intra-lingual and inter-lingual. The results of the study show that inappropriate collocation errors are the most frequent (38.71%) while literal translation errors are the second most frequent (23.65%). On the other hand, errors resulting from the confusion of words with similar meaning are the least frequent (11.29%). Finally, Sheshsha concludes that intra-lingual errors, which have their source in the target language, are more frequent than inter-lingual errors.

Along the same lines, in her research, reported in an unpublished MA thesis, Al-Jabri (1998) analyzes the lexical errors in the written English of a 110 Saudi, female, college freshman students. Like Sheshsha (1993), she classifies her errors into two main types, intra-lingual and inter-lingual, with the former including 8 categories and a ninth for uninterpretable errors, and the latter including the same two categories as Sheshsha: literal translation, and divergence. She agrees with Sheshsha that intra-lingual errors are more frequent than the inter-lingual errors. However, she waters-down her conclusion by adding that “most categories of lexical errors are caused by interference between lexical items in the same language or between the two languages” (Al-Jabri, 1998, p.iii). Al-Jabri also acknowledges that the categories of lexical errors are neither exclusive nor discrete.

Also working with Arab students, Mahmoud (2005) presented a systematic analysis of Arab EFL learners’ lexical errors in general and of collocation errors in particular. In his study, he presents empirical data verifying the informal observations and theoretical assertions that EFL learners produce ‘unnatural’ word combinations. A total of 420 collocations were found in 42 essays written by Arabic-speaking university students majoring in English. About two thirds of these collocations (64%) were incorrect and 80% of these were lexical collocations as opposed to grammatical ones. He found that 61% of the incorrect combinations were due to negative transfer from Arabic. Mahmoud, however, focused his work on only one type of lexical errors, collocations.
Working with a much broader base of categories, Hemchua and Schmitt (2006) propose a lexical error taxonomy (adapted from James, 1998) which includes two major categories, formal and semantic. Under each category there are classifications and sub-classifications. All in all, their taxonomy accounts for 24 different types of lexical errors. Hemchua and Schmitt acknowledge that there is a need for a comprehensive taxonomy which incorporates a wide range of lexical errors categories. They used their taxonomy to analyze the lexical errors made by Thai L2 third-year university students in their English compositions. They found that first, the most common types of errors were near synonyms, second, that students made more errors related to semantics, rather than formal errors, and third, that the errors were mainly due to the difficulty of L2, rather than L1 transfer. They believe that “the error taxonomy and methodology [they] present can be applied to the analysis of lexical errors in the academic writing of other L2 learners, and that [their] findings have implications for vocabulary learning and teaching in a wide range of language contexts” (p.4). They also warn that “the use of compact classification systems to explain learners’ errors can result in unclear boundaries and arbitrary classifications” (p.8).

Llach (2007) looks at lexical errors from a different perspective than that of the studies cited above and examines the relationship between lexical errors and the proficiency level of the learners. She analyzes the compositions of 79 4th grade Spanish students, aged between 9-11 years, enrolled in a Catholic state supported private school in La Rioja, Spain. Llach identified the lexical errors and quantified them, and then correlated the frequency of errors with the proficiency level of the students. The students’ proficiency level in English was determined by both a cloze test and a reading comprehension test. As would be expected, she found a strong positive correlation between the number of lexical errors in students’ compositions and their proficiency level in general. In her study, Llach does not classify the lexical errors into types, but simply makes a count of all lexical errors in general that she finds in the students’ papers. Her study, however, provides empirical evidence to the general importance of the accurate choice of lexical items and how it reflects on the general proficiency level of a language learner.

Among the studies reported above, only Hemchua and Schmitt (2006) present a taxonomy which incorporates a wide range of lexical error categories. Their proposed taxonomy can definitely account for lexical errors more specifically than the typologies presented in other research
studies. Zughoul (1991) and Al Jabri (1998), while employing half the number of categories that Hemchua and Schmitt used, also attempted to use a broader taxonomy than those who only employed a limited number of lexical error categories. Even though the studies presented above are only examples of the available research on lexical errors and by no means constitute an exhaustive presentation, it is safe to claim that they are a representative sample. Most importantly, it is evident that the studies employing a comprehensive list of lexical error categories are scarce.

4. Design of study

4.1 Participants

The participants in the study are female students in a prep-year program at Taibah University in Al Madinah Al Munawarah, Saudi Arabia. The students’ native language is Arabic. Before the prep-year, students had been taught EFL for a total of six years in intermediary and secondary schools. Since writing skills were not heavily emphasized in the school curriculum, these students can be classified as novice writers of English. The English language proficiency levels of the participants ranged from low beginners to low intermediate.

In the university prep-year, students receive 20 hours of English instruction per week in the four language skills: listening, speaking, reading and writing. The main objective of the writing course was to develop the students’ writing abilities in various rhetorical modes of expository writing such as description, narration, and comparison/contrast. The students received 7 weeks of instruction before taking a mid-term exam and an additional 7 weeks of English instruction before the final exam. In both the mid-term and the final, the students were given 90 minutes to complete the writing exam. For the mid-term exam, students were asked to write a one paragraph composition from a selection of topics. Topics included advantages and disadvantages of living in a big city and a small city, reasons for their choice of an ideal job, and a description of the life of ProphetMohammad (peace and blessings be upon him). For the final exam, students had to write a 3-paragraph essay on topics such as the influence of advertisements, Saudi culture: past and present, and the influence of technology on our life.

4.2 Procedures

The data for this study consists of 96 essays randomly selected from a pool of 563 writing exam papers. The taxonomy employed resulted from the examination of various taxonomies already compiled by other
researchers such as Zughoul (1991), Llach (2005), Hemchua and Schmitt (2006) which is in turn an adaptation of James (1998). After a close examination of various taxonomies, Hemchua and Schmitt’s taxonomy was adopted because of its comprehensive nature to constitute the basis of the present study. However, the writing of the taxonomy developed into a process in which the researchers modified and refined the taxonomy used by Hemchua and Schmitt several times while analyzing the students’ sample papers. The 96 writing samples were divided among the researchers, with a lexical error taxonomy sheet, such as that shown in table 2 below, attached to each one. The researchers read the writing samples, identified the lexical errors, highlighted them, and classified them, making a frequency count for each writing sample using the taxonomy sheet. Prior to individual analysis of papers, several group sessions were held in which sample papers were analyzed and discussed by all the three researchers to ensure that a uniform analysis was achieved. Two of the researchers are native speakers of Arabic, while the third has a rudimentary knowledge of the language. All three researchers are near native speakers of English and are experienced teachers of English as a second/foreign language.

Accordingly, the lexical errors counted for this study included content word, nouns, verbs, phrasal verbs, adjectives and adverbs. These lexical errors were placed into two main categories: formal and semantic errors; each including a number of subcategories, which are discussed in the following section.

4.3. Categorization of Lexical Errors

The following section is a discussion of how Hemchua and Schmitt’s (2006) taxonomy was modified in constructing the taxonomy used in the present study. The whole taxonomy is divided into two main types of
errors, formal and semantic. The formal errors comprise 3 major categories, further divided into 8 subcategories, while the semantic errors are divided into four major categories, subdivided into 13 subcategories. The taxonomy thus included a total of 21 subcategories of lexical errors.

4.3.1. Formal Errors: Formal Misselections, Misformations and Spelling

Formal Errors are sub-divided into three main categories: formal misselection; misformations and spelling errors. The formal misselection category then contains three subcategories: misselection of suffixes, misselection of prefixes and false friends. The researchers did not adopt the subcategory of vowel-based and the consonant-based formal misselection errors from Hemchua and Schmitt’s taxonomy because they thought these categories would be more appropriately placed in a category of spelling errors. The misformations category consists of two subcategories: borrowing and coinage. The term “claque” was replaced by a more transparent term — translation from L1, and placed under semantic errors as a subcategory of confusion of sense relations.

The spelling errors category only includes three subcategories: errors impeding comprehension, errors resulting in inappropriate meaning and errors due to L1 transfer. When analyzing the students’ papers, we found that spelling errors were so prevalent that including them in the study would skew the results; in fact, we felt that the spelling errors warranted a separate study. The subcategories chosen for spelling errors are the ones the researchers believed were most significant.

4.3.2. Semantic Errors: Confusion of sense relations, Collocation, Connotation, and Stylistic errors

The confusion of sense relations category includes 8 subcategories: the use of a general term for a specific one; overly specific terms, inappropriate co-hyponyms, near synonyms, translation from L1, binary terms, inappropriate meaning and distortion of meaning. The first four categories were adopted from Hemchua and Schmitt’s (2006) taxonomy. The confusion of binary terms category was borrowed from Zughoul’s (1991) taxonomy and two other categories, inappropriate meaning and distortion of meaning, were added by the researchers. Refer to the taxonomy in table 2 below and the results and discussion section for examples and explanations.

As for the collocation errors, we decided to collapse all their sub-types under one major heading “collocation errors”, as we believed that a detailed analysis of the various types of collocation errors was beyond the
scope of this paper and should be studied separately. Finally, the stylistic error category has 3 subcategories namely verbosity, misuse of compounds and circumlocution. Verbosity is used in Hemchua and Schmitt’s taxonomy and circumlocution is used in Zughoul’s. Unlike the taxonomies of these two researchers which only included either one or the other of these two subcategories, our taxonomy included both, thus making a distinction between verbosity and circumlocution.

4.4. L1 Transfer
Most studies on lexical errors classify errors due to L1 transfer as a separate category, which would often-times include 2 or 3 subcategories. This is the case for example in Hemchua and Schmitt (2006) who list L1 transfer as a type of formal error, and consider it as a misformation. In their taxonomy they include three subcategories: borrowing from L1, coinage based on L1, and direct translation from L1. Likewise, Sheshsha (1993) classifies lexical errors according to five categories divided into two main types: the first, intra-lingual (developmental) errors which include three subcategories, collocation errors, confusion of words with formal similarities and confusion of words with similar meaning; and the second type, interlingual errors (due to L1 transfer) which include only two subcategories, translation from L1 and divergence. Al Jabri (1998) like Sheshsha, divides inter-lingual errors into two subcategories: translation from L1 and divergence. However, in the process of analyzing the writing samples for the current study, we found a much more complicated picture; errors due to L1 transfer could not simply be classified under two or three categories. Many of the errors often seen as intra-lingual errors were in fact also due to L1 transfer (i.e inter-lingual errors). Thus a wrong use of collocation, a near synonym, stylistic errors or even formal misslection (as for example spelling errors) could also be due to L1 transfer, as will be discussed below. This supports the findings of other researchers who saw L1 transfer operating in many types of lexical errors (see for example Zughoul (1991), Mahmoud (2000), Chebchoub (2005), Mohammed (2005)). For this reason we provided on our taxonomy sheet a column for L1 transfer that would allow us to provide a dual classification for any error; in other words, an error could be classified as both developmental and yet at the same time influenced by the mother tongue. In addition to this, direct translation from L1 remains a distinct category listed under formal misselection. (See table 2. below.)

4.5 Error Count
The lexical errors counted were either a word, or two words, as for
example in the case of collocations, or whole phrases and sentences as in the case of stylistic errors. Using the taxonomy, the errors were identified and tallied on a taxonomy sheet for each student’s writing sample. Then, the total number of each type of lexical error in all of the papers was recorded, as was the total number of lexical errors in general. In addition, the average number of errors in the essays was calculated as was the percentage of each type of error. The total number of words in each paper was counted and the total number of words in all of the essays was recorded. The mean length of the essays was also calculated. (See table 1 below.)

5 Results and Discussion

The 96 writing samples analyzed for this study yielded a total of 718 lexical errors, with an average number of 7.48 errors in each essay. The total number of words in all of the writing samples was 16,143 words, while the mean length of each writing sample was 168. Table 1 below shows the length and mean length of the writing samples as well as the number and mean number of errors in the students’ essays.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Total number of writing samples</td>
</tr>
<tr>
<td>2.</td>
<td>Total number of words in all writing samples</td>
</tr>
<tr>
<td>3.</td>
<td>Mean length of writing sample</td>
</tr>
<tr>
<td>4.</td>
<td>Total number of lexical errors in the writing samples</td>
</tr>
<tr>
<td>5.</td>
<td>Mean number of lexical errors in writing samples</td>
</tr>
</tbody>
</table>

Table 1. Length of the writing samples and number of lexical errors

The semantic lexical errors occurred with much higher frequency than the formal lexical errors, 434 errors (60.45%) compared with 284 (39.55%). (See figure 1 below.)

Figure 1. Formal versus semantic lexical errors
The major categories of both the formal and the lexical errors varied in their frequency of occurrence as displayed in figure 2 below, with one major category, connotation errors, not occurring at all. The subcategories within each major category of formal and semantic lexical errors showed even greater variation, with some types of errors not occurring at all or occurring only once, as for example the misuse of prefixes and using an overly specific term; others, on the other hand occurred quite frequently, such as the misuse of suffixes which occurred 128 times (17.83%). This type of error which is a sub-category of formal misselection was the most frequent. It was followed by lexical errors due to direct translation from L1, which occurred 113 times (15.74%). Errors due to direct translation from L1 were classified under a subcategory of semantic lexical errors, the confusion of sense relations. This type of error was followed in frequency by two other types of semantic lexical errors, collocation errors and the use of words with inappropriate meaning, both of which occurred 87 times (12.12%). Table 2 below shows the frequency of occurrence of the subcategories of formal and semantic errors. The errors are discussed in order of their frequency of occurrence beginning with the semantic errors.
Figure 3. Subcategories of formal lexical errors

Figure 4. Subcategories of semantic lexical errors
### Examples

<table>
<thead>
<tr>
<th>Examples</th>
<th>%</th>
<th>No.</th>
<th>A. Formal errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Formal misselection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saudi Arabia has a lot of change and <em>different</em> between now and 30 years ago.</td>
<td>17.83%</td>
<td>128</td>
<td>Suffix type</td>
</tr>
<tr>
<td>Also, living in medinah is <em>unexpensive.</em></td>
<td>0.14%</td>
<td>1</td>
<td>Prefix type</td>
</tr>
<tr>
<td></td>
<td>0.0%</td>
<td>0</td>
<td>False friends</td>
</tr>
<tr>
<td><strong>2. Misformations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In school, there are a lot of hard roles for example: no red shoes, no mobile, no <em>khol</em>, (eye liner)</td>
<td>4.31%</td>
<td>31</td>
<td>Borrowing</td>
</tr>
<tr>
<td>We <em>couble</em> stop any time in any were.</td>
<td>0.42%</td>
<td>3</td>
<td>Coinage</td>
</tr>
<tr>
<td><strong>3. Spelling errors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The people in Madina know asch other and visit asch.</td>
<td>5.43%</td>
<td>39</td>
<td>Impeding comprehension</td>
</tr>
<tr>
<td>Al-Medina Al-Mnawara the pest countery in the world.</td>
<td>6.27%</td>
<td>45</td>
<td>Inappropriate meaning</td>
</tr>
<tr>
<td>Now the live is better than the bast.</td>
<td>5.15%</td>
<td>37</td>
<td>L1 Transfer</td>
</tr>
<tr>
<td><strong>B. Semantic errors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1. Confusion of sense relations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The most people adapt with this big change in the life <em>ways</em> and still use the ancestor language. (style)</td>
<td>3.48%</td>
<td>25</td>
<td>General term for specific one</td>
</tr>
<tr>
<td>Advertisers use many <em>methods</em> to make us buy it. (ways)</td>
<td>0.14%</td>
<td>1</td>
<td>Overly specific term</td>
</tr>
<tr>
<td>Fast food is derty because we don't know the clean of <em>equipment.</em></td>
<td>0.7%</td>
<td>5</td>
<td>Inappropriate co-hyponyms</td>
</tr>
</tbody>
</table>
The fast food is very dangerous for people life. Similarly, such as: give the person power, strong, you can do hard working and gives some *activity*.

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
<th>Count</th>
<th>Issue</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near synonyms</td>
<td>5.15%</td>
<td>37</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Translation from L1</td>
<td>15.74%</td>
<td>113</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Binary terms</td>
<td>0.42%</td>
<td>3</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Inappropriate meaning</td>
<td>12.12%</td>
<td>87</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Distortion of meaning</td>
<td>2.09%</td>
<td>15</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>2. Collocation errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On the other hand, I wear in University different clothes like different skirt a lot of color.</td>
<td>6.27%</td>
<td>45</td>
<td>4.1</td>
<td>Verbosity</td>
</tr>
<tr>
<td>I like a style of life on there. (life style) Also, there is the field of Date. (date fields)</td>
<td>0.28%</td>
<td>2</td>
<td>4.2</td>
<td>Misuse of compounds</td>
</tr>
</tbody>
</table>
Also, both of them are existed in many kinds and choices. But in university not say that take out take of from my marks. (deduct)

<table>
<thead>
<tr>
<th></th>
<th>1.94%</th>
<th>14</th>
<th>Circumlocution</th>
<th>4.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100%</td>
<td>718</td>
<td>Total number of errors</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Lexical error taxonomy (Adapted from Hemchua & Schmidt, 2006 and James, 1998) and frequency of errors

5.1 Semantic Lexical Errors
Of the semantic errors, the major category with the highest frequency of occurrence was the confusion of sense relations. The total number of errors in this category was 285, with a percentage of 39.69% of the total number of errors. The types of errors included in the category listed in their order of frequency of occurrence were direct translation of lexemes from L1, using words with inappropriate meaning, near synonyms, a general term for a specific one, words that distorted meaning, inappropriate co-hyponyms, wrong binary terms, and finally overly specific terms. ‘Collocation errors,’ as a major category, were next in frequency after ‘confusion of sense relations’ and occurred with the same frequency as the sub-category of the ‘use of inappropriate words.’ Stylistic errors were the least frequent, with ‘verbosity’ the highest subcategory followed by ‘circumlocution.’ These types of semantic lexical errors are discussed below, followed by a discussion of the frequency of formal lexical errors.

5.1.1 Direct translation from L1
Direct translation from L1 was the most frequent semantic lexical error, occurring 113 times (15.74%). This coincides with findings in other research studies on writing errors. For example Mohammed (2005), in a study of collocation errors made by Arab students found that 61% of the collocation errors, regardless of their specific type, could be traced back to negative native language transfer. He cites a number of other researchers who also found inter-lingual transfer to be a key factor in students' errors at all levels of proficiency. (See for example Bhela, 1999; Mahmoud, 2000; Odlin, 1989; Ringbom, 1987; Sheen, 2001; Tang, J. 2002; cited in Mohammed, 2005, p.5). Similarly, in a study of lexical errors in the compositions of Arabic speaking college students, Zughoul (1991) observed that direct translation from L1 was second highest in frequency, after assumed synonymity. In his study, he reports that direct translation
occurred 82 times out of a total 691 errors (11.8%). Our findings, however, contrast with those of Sheshsha (1993) and Al-Jabri (1998) who studied a similar population to ours (Saudi college students in Makkah) but found intra-lingual errors to be more frequent than inter-lingual errors. One possible explanation for this is that our taxonomy is more comprehensive than theirs, and therefore breaks down the types of errors into many subcategories, which in turn results in lower frequency for each sub-type.

Examples of lexical errors due to L1 transfer that we found in the writing samples of our students occurred on the word level, phrase level and even sentence level as can be seen in the following examples:

Translation from L1 on the word level:

1. *Jeddah is very important and Al Madinah because found the Holy Mosqe*
   
   *Jeddah hamma jidan wa (kaddalik) al madi:nah haiθ yujad al haram*

2. *They use this chance to put on their ads and they make it all about football.*
   
   *Yastağ ilu:n haðihi ilfursa liwaq3 ?i3lana:tihim*

On the level of the noun phrase, a common error was the use of the definite article “the” with abstract nouns, such as work. This is an obvious transfer from Arabic where all nouns accept the definite article ?al (the). Hence the word 3amal (work) would occur with the definite articles (?al) in the following contexts:

3. *Then his father comes from the work.*
   
   *θum 3ad walidahu min al 3amal*

4. *But now everything has change. The women now can go to the work.*
   

5. *When I finish school we have exam in the both.*
   
   *Wa hin antahi min al dirasa, 3indana ixtibar fil ?i3nain*

Other examples of L1 transfer on the phrase level are:

6. *It has two mosques which are the most important mousqes in all over the world.*
   
   *Wa biha masjidain huma ?aham masjidain fi djami:3 ?anhaa Il 3a:lam*

In this case, there is a transfer of the use of the preposition, “in” in
the expression “all over the world”, which is correct in Arabic, but erroneous in English. The same applies to example 7 below, where the phrase “go out from the house” is a direct translation from Arabic.

7. * For example women in the past were not able to go out from the house.
   fa maθalan al nisα fil maθi lam yastaθi3u al xurudj min al manzil

8 * Next, jibril came to prophet Muhammad and he said read but prophet Mohammed can’t read and then he gave him prophethood and he carried it at age 40.
   θumm ataa jibril lilrasool Muhammad waqαal iqraθ walakin al rasool Muhammad laa yastaθi3an yaqraθ θumm maθ habahu annbuwah wa hamalaha min sinn al arba3i:n

L1 transfer also occurred on the sentential level where the entire sentence is a direct translation from Arabic as in the following example:

9. * Saudi Arabia is a country has a high place between the countries in all over the world.
   Al mamlaka al 3arabiya θal saudiya dawla laha makaana 3aliya bain al duwal fi kul al 3alam

As mentioned above, L1 transfer as an error type was coded in our taxonomy both as an independent semantic error and also as an intervening factor that would account for errors in other subcategories, whether formal or semantic. Often we found an erroneous lexical choice that would be the result of two overlapping factors, as for example a spelling error, or wrong use of collocation due to L1 transfer, or a near synonym due to L1 transfer. Similarly verbosity, a stylistic error, could also be due to the influence of the students’ native language, Arabic, where the stringing together of synonyms is considered a stylistic embellishment of a text, and a display of the writers verbal ability and her/his mastery of vocabulary. As such, by the standards of Arabic discourse, the repetition of synonyms or near synonyms would be favorably regarded. An example of this is the use of lexical couplets discussed below.

5.1.2 Lexical Choices with Inappropriate Meaning

Next in frequency after L1 transfer was the use of a word that gave an inappropriate meaning. This type of error occurred 87 times and represented almost a third (30.52%) of the confusion of sense relations category, and 12.12% of the total number of errors. Examples of this error
are:

1. * He is job with sheep. (referring to the Prophet Mohammad who for a while in his early youth worked as a shepherd.)
2. * Then he back in Makkah to Hajj. (Then he returned to Makkah for Hajj.)
3. * He organized Qubba Musqe in Madinah.
4. * Likewise Riyadh, Madina is a modern city.
5. * I will rememper the lovely for Al Madina.
6. * I can join my hobby of drawing with my work.
7. * I found the prep year is very hard and the work is very long.
8. * But they (University and school) are quit different in other points for example timing and uniform.
9. * It has a lot of calores and may be get for you some of dangers decise
10. * Teenagers like it because it is saucy and falvourid

As stated above, some of the erroneous lexical choices classified under the subcategory of ‘giving an inappropriate meaning’ could, at the same time, be traced back to L1 transfer. An example of this is the use of the word ‘join’ in sentence number 6:

6. * I can join my hobby of drawing with my work.

The word join here is used with the sense of “combine.” This is a direct translation of the word ajma3u used in both classical and colloquial Arabic with the sense of either put together or combine, as in the Arabic phrase ʔagma3u [ma bain] hiwayati wa 3amali. Another example of inappropriate meaning due to L1 transfer is in example no. 9:

9. * it has a lot of calores and may be get for you some of dangers decise.

Fiilha su3rat katir wi mumkin tidjib lak ʔamrad khatira

The phrase get for you, while being an inappropriate choice, is also a direct translation from the colloquial Arabic phrase tidjib lak. The majority of errors belonging to this category, however, could not be traced to L1 transfer, and can simply be explained by the students’ incomplete knowledge of L2. For example in example no. 10, the adjectives saucy and falvourid were probably meant to be ‘spicy and tasty,’ and the student’s incomplete knowledge of the language accounts for the incorrect choice of these two adjectives.

Some errors could be accounted for by hyper-correction. For example, the misuse of the phrase may be in example 9 in place of ‘may’ is a frequent error among Arab speakers who often as a form of hypercorrection overuse
verb ‘to be’, in imitation of the copula ‘be’.

5.1.3 Near Synonyms
The subcategory of the ‘near synonyms’ was the next in frequency of occurrence. However, it occurred at a much lower rate than either translation from L1 or inappropriate meaning, occurring only 37 times at a rate of 5.15% of the total number of errors. The following are examples of this type of error, with the presumably intended word in brackets following the sentence.

1. * Also, there are many places to tour. (visit)
2. * So my relationship with this city is very deep. (strong)
3. * Medina is the place where the prophet Mohammed lived after travelling from Mecca. (migrating).
4. Later her traveled to many cities and rural. (villages)

Example 2, besides being an example of a near synonym, is also an example of L1 transfer, where a relationship may be described as “deep” (?amiqa). Interestingly, though, the Arabic word for “strong” (qawiyya) is also frequently used to describe relationships. Moreover the word “strong” is a word that the student probably knows, but wrongly assumes to be synonymous with “deep.” The case is not the same with the misuse of the word traveling for migrating, as in example number 3, where the latter does require more sophisticated knowledge of English vocabulary.

5.1.4 Collocation Errors
We found the collocation errors hardest to classify, as many could also be classified as simply inappropriate meaning, or near synonyms. However, we agreed upon 87 collocation errors, 12.12% of the total number of errors. Some of the examples are listed below with the correct collocation in brackets following each sentence.

1. *In addition there were no places to get fun time. (to have fun)
2. *Because the most people adapt with this big change in the life ways and still use their ancestor language. (in their life-style)
3. *For example, you can’t practice any exercise except sitting. (do any exercise)

The frequency of collocation errors found in the students’ essays we analyzed is much lower than that found by Hemchua and Schmitt (2006). In their study, collocation amounted to 26.05% of the total number of errors. The discrepancy between our results and theirs could be attributed to the difficulty of classifying collocation errors as they overlap with other types of lexical errors, as stated above. However, despite the fact that the
collocation errors found in the essays we analyzed were fewer than those found by other researchers, they were still relatively frequent in their occurrence compared to other errors in our study, coming fourth in place (together with words yielding inappropriate meaning) after misuse of suffixes, distortion of lexical items due to spelling errors, and translation from L1. The results indicate that there needs to be more direct teaching of collocations in English. Furthermore, it points to the importance of teaching vocabulary items in context rather than in isolated lists.

5.1.5 Stylistic Errors
Stylistic errors in general occurred 61 times (8.49%). This category included subcategories such as verbosity, circumlocution, and misuse of compounds, which occurred 45 times, 14 times and twice respectively.

5.1.5.1 Verbosity
Examples of verbosity found in the students’ writings are:

1. * I don’t know more about studying in university but I think it will be good and nice.
2. * The travel between the cities is very easy and simple.
3. * Al Madinah have a lot of garden and park.
4. * It gives the person power and strong.
5. * He was very intelligent and smart.
6. * We can visit our family and friends by easy way.

In examples 1-5 above, the verbosity in the students’ writing is a transfer from their L1, where the use of couplets such as “easy and simple,” “big and large,” is a favorable rhetorical style, referred to as lexical couplets. A lexical couplet “is a phrase or sentence coordinating two or more words with shared semantic components and ‘a single referent’, like the idiomatic English ‘bits and pieces’” (Rieschild, 2006, p.6). In Arabic discourse, this is considered a stylistic embellishment. Synonymous parallelism may extend to adjacent clauses as coordinated phrases with a cognate verb and noun, lexical couplets may include rhyme (phonetic repetition) further intensifying the utterance (Rieschild, 2006). However, not all verbosity errors were examples of couplets; example 6 above is an example of using too many words to express a concept.

5.2 Formal Errors
In the writing samples we analyzed, the formal errors were fewer than the semantic; 39.55% compared with 60.45% of the total number of errors. The major category of formal errors includes three subcategories. The first is formal misselection, which includes errors in the choice of prefixes,
suffixes, and false friends. The second is the subcategory of misformation which includes lexical errors due to borrowing and coinage. And, the third category is distortions due to spelling. However, as stated above, we found the spelling errors in the students’ writings so pervasive as to warrant a separate study. Thus, we considered only three types of spelling errors: those that impeded comprehension, those that yielded an inappropriate meaning, and those that appeared to be due to L1 transfer. The results of the study showed that the three types of errors occurred with near frequency, 39 times (5.43%), 44 times (6.13%) and 37 times 5.15%, respectively. The different types of formal errors are discussed below, ordered in their frequency of occurrence.

5.2.1 Formal Misselection
The most prevalent type of error in this subcategory was that of suffix misselection. This error occurred 128 times out of the total of 284 formal errors, at the rate of 45.07% of the formal errors, and 17.83% of the total number of errors. There was only one error of prefix misselection in the 96 essays we analyzed, and no cases of the use of false friends.

5.2.1.1 Formal Misselection of Suffixes
As stated above, of the formal misselection errors, the misselection of suffixes was the most frequent type of error, occurring 128 times out of a total of 284 formal errors, 45.07% of the formal errors, and 17.83% of the total number of errors. The errors in the choice of suffixes reflect a gap in the students’ knowledge of the morphological structure of words and point to a need of direct teaching of word parts.

1. * But both of them are interesting and the both have its memorize.
2. * I became responsibility for myself.
3. * I found a lot of different between my scools and my university.

5.2.1.2 Distortion of Lexical Items Due to Spelling

Spelling errors due to L1 transfer
The spelling errors due to the influence of the L1 phonological system were both interesting and amusing. The transfer came mainly from the spoken Saudi variety. Interestingly, many were cases of hyper correction, were the phoneme /v/, which is generally not used in Saudi dialects, was over used in place of the phoneme /f/ in English words.¹ An example is the spelling of the word safe as *save. Another common type of spelling error was the use of the phoneme /b/ in place of the phoneme /p/, again because the latter is not found in classical Arabic, and only occurs in some spoken varieties.² Unlike the phoneme /v/ which is used in some Arabic...
vernaculars, /p/ is only scarcely used in Arabic vernaculars, and occurs mainly in borrowed words. The phoneme /p/ also occurred in place of the phoneme /b/ in English words as for example in the adjective _beautiful_, a hypercorrection error for the adjective beautiful.

The following are some of the spelling errors due to L1 transfer that we found in the students’ papers:

**Use of /f/ in place of /v/**
1. * The children should *foid* fast food. (avoid)

**Hypercorrection: use of /v/ in place of /f/**
2. And they became happy with the *givts* which come free inside the egg of the chocolate. (gifts)
3. Madinah is a city which *provent* Mohamed is built. (Prophet)

**Use of /b/ in place of /p/**
4. * He was very *bolit*. (polite)
5. * Then he went to Hajj in 10th Hijri with *gruwb* from Madina. (group)
6. * Rait at Jeddah *bleses* very modern and quiet. (places)

**Hypercorrection: use of /p/ in place of /b/**
7. * The home food is the *pest* one. (best)
8. * They don’t take their mothers’ sandwiches any more. They think it is *imparassing*. (embarassing)
9. * Finally, I well *rememper* the lovely for Al Madina. (remember)

As in the case of other types of lexical errors, sometimes a distortion error due to misspelling can be classified under more than one category, as in the case of the words “pest”, and “Al Nupy”. The former may also be categorized as an error that gives inappropriate meaning, while the latter is borrowed from Arabic, where the word _al Nabiy_, means the Prophet. In general, the spelling errors due to phonological differences between L1 and L2 all point to the fact that the phonological influence of L1 is not limited to oral performance but carries over to the written performance of the students too.

### 5.2.3 Misformation Due to Borrowing from L1

In the 96 essays we analyzed, we found 31 examples of borrowing from Arabic. Unlike in the case of coined words (discussed below), borrowed words did not necessarily occur in the writing samples of students with low proficiency. Examples of borrowing from L1 can be seen in the following:

1. Then he married _al sayeda_ kahdija when he was 25. (Referring to the
Prophet Mohammad’s (PBUH) wife “the honourable lady” Khadija.)
2. Next coming all from the gaswaths, like Badr. (The Prophet’s military excursions).
3. Then he became said. (Referring to the Prophet Mohamed (PBUH)).
4. All the Sahaba lived in it. (The Companions of the Prophet Mohammad (PBUH)).

We noticed that the majority of borrowed words occurred in the writing samples of students who chose the topic about the life of the Prophet Mohammad (PBUH). Thus it appears that it is the nature of the topic that motivated the use of words borrowed from Arabic. Other examples of words related to religion that occurred in many of the students’ writings were Hajj, Ghar Hiraa, Haram (pilgrimage, Hiraa cave, and Holy Mosque, respectively). Actually, one would see these three words occurring as they are (transliterations of the Arabic words) in most books on Islam, and in most biographies of the Prophet. As such then, one must admit that counting these words as lexical errors is questionable; they should rather be considered as code switching. Students who wrote about college life compared to school life used very few borrowed words.

5.2.4. Coinage of New Words

Although this type of error occurred only three times out of the total number of errors (718), it was actually quite significant as it rendered the students’ writing quite humorous, yet also incomprehensible, as in the following examples,

1. * When his age a 40 jebrel coming his in the clomepn.

Based on our knowledge of the biography of the prophet Mohamed, we understood the student’s sentence to mean:

When he was 40 years old, the angel Gabriel came to him in the cave.

It is apparent from the student’s sentence (and the rest of the paragraph that she wrote for the assignment) that her proficiency in English is quite low, lower than most of the other students. She probably coined the word out of desperation and inability to come up with another word for cave or mountain. The fact that the word she coins “clomepn” is not even remotely related phonologically or morphologically to the words ‘cave’, or ‘mountain’, demonstrates the total lack of such words in her repertoire of English vocabulary.

The second example of a coined word is slightly different in nature as it is a coined verb based on an actually existing noun, and makes use of a
verb forming suffix. The following sentence refers to Prophet Mohammad (PBUH):

2. * And this year wifed Al sayedah Khadija.

What the student wants to say is:

And that year he married (the honourable lady) Khadijah.

The coined word is also influenced by L1, where the word for "married" is derived from the noun for wife “zawj.”

Other errors in the taxonomy occurred with quite low frequency, and due to constraints of space, we felt the statistics and examples provided in table 2 above would give a fair idea of the nature of these errors and their frequency of occurrence.

7.0 Implications of the Study

The findings of the study suggest that L1 plays an important role in the acquisition of L2 lexemes, and in the students’ production and choice of lexical items in their writings. Since this is the case, then it may prove productive to capitalize on this strategy that is used by the students. Rather than leaving students to make their own comparisons and associations with their native language, perhaps teachers should be encouraged to refer back to students’ L1 in FL classes where students share the same native language. They could elicit from the students’ comparisons that would help them learn about the English language. It may also prove worthwhile to encourage native English speaking EFL teachers to learn about the language of the host country where they happen to be teaching, so that they can understand the reasons behind some of the errors their students make.

In addition, students should be trained in how to effectively use an English monolingual dictionary and a thesaurus. Furthermore, all new lexical items should be taught in context. Students could also practice the use of superordinates in vocabulary learning as findings on confusion of binary terms and near synonyms point to the significance of this particular training. Pre-service teachers should not only be equipped with linguistic knowledge such as morphology, phonology, syntax and pragmatics of the target language, but should also be trained to apply this knowledge in a classroom setting. For instance, more direct teaching is needed of the morphological structure of words, and of word associations and collocations. Or as suggested by Zughoul (1991), a problematic word list can be generated and given to students, however, in order for such a list to be fruitful, the problematic words should be taught in context and students
should be encouraged to use the new words in their speaking/writing classes.

As found in the study, students tend to spell words according to their pronunciation resulting in phonological interference in choosing between /b/ vs. /p/, or hypercorrection in cases where the sound is absent in the Arabic language such as using the word /pest/ for/best/. For this, we suggest that teachers directly point out to students that a change in a phoneme in a word will yield a completely different meaning in English. Teachers can use exercises to help students distinguish between minimal pairs and also try to enhance their phonemic awareness when teaching vocabulary and spelling. As for dealing with collocation errors, students may be informed of the value of corpora and may be encouraged to access such corpora online and to utilize this facility when learning collocations. Also, beginners and low intermediate level students can initially memorize chunks of words in learning collocations.

The findings of our study concur with Zughoul’s (1991), more so than Hemchua and Schmitt (2006), probably due to the similarity of the students’ language background. While Hemchua and Schmitt found that the majority of lexical errors were not L1-specific, we found that the underlying source for most lexical errors in our study can be traced back to L1 influence, though some may be caused at the same time by minimal or incomplete knowledge of the English diction. With this in mind, we suggest that teachers use lexical error taxonomies or develop their own in their vocabulary teaching. We strongly believe that these taxonomies not only serve as research tools but, more importantly, as instructional tools that should be employed by teachers. When used effectively, these lexical error taxonomies can assist students in raising their meta-cognitive skill of recognizing and possibly even self-correcting their errors. This could be a way to help minimize the fossilization of lexical errors.
Footnotes:
1. The phoneme /v/, however, is not absent from all Arabic vernaculars. It occurs in Egyptian dialects where it is quite commonly used both in proper names, such as “Mervat,” (a girl’s name) and in borrowed words, such as “villa”. Both words are also common in Saudi dialects, but are pronounced with a /f/ in place of the /v/.
2. In the Egyptian vernacular, the use of /p/ is generally regarded as a feature of the speech of upper middle and upper classes. It occurs for example in borrowed words such as “panatalon” (pants). The phonological error of substituting /b/ for /p/ in English words, however, is persistent and is common even among the most highly educated, particularly those who did not receive bilingual English/Arabic education.

References


تهدف هذه الدراسة إلى تقديم نموذج شامل يمكن استخدامه في تحليل أخطاء المفردات اللغوية الإنجليزية في كتابة الطلاب السعوديين وقد طُبِّق هذا على الطلاب اللاتي يدرسن في برنامج السنة التحضيرية بجامعة طيبة بالمدينة المنورة بالمملكة العربية السعودية. كما تهدف الدراسة الحالية إلى الإجابة على الأسئلة التالية: أي- ماهي الأخطاء الشائعة في استخدام الطلاب للمفردات اللغوية عند كتابتهم لموضوعات الإنشاء باللغة الإنجليزية؟ ب- أي من هذه الأخطاء أكثر شيوعاً من غيرها؟ ج- أي من هذه الأخطاء يحدث نتيجة لتأثير اللغة الأم (العربية) على اللغة الإنجليزية؟ وقد قام الباحثون بتحليل ست وسعين مقالًا والتي شكلت العينة لهذه الدراسة، وقد احتوت المقالات على سبعون وثمانية عشر خطأ. وقد احتل المرتبة الأولى من حيث عدد الأخطاء تلك الأخطاء التي ارتكبها الطلاب في استخدام اللواحق (suffixes) التي تضاف في نهاية بعض المفردات (17.83% من إجمالي عدد الأخطاء)، ويليها الأخطاء المرتبة على الدرجة الحرفية من العربية للإنجليزية لبعض المفردات (15.74%). ولكن بشكل عام كانت الأخطاء التي تتبع بالمعنى أكثر من الأخطاء في التركيب اللغوي (60.45%) مقارنة ب (39.55%). وقد قام الباحثون في نهاية البحث بمناقشة التطبيقات التربوية لتدريس المفردات اللغوية في فصول اللغة الإنجليزية لغير الناطقين بها.