

Semantic Errors Committed by Yemeni University Learners: Classifications and Sources

Mohammed Qassem Al-Shormani¹ & Yehia Ahmed Al-Sohbani¹

¹ University of Ibb, Yemen

Correspondence: Mohammed Qassem Al-Shormani, University of Ibb, Yemen. Tel: 967-7-7230-0815. E-mail: moshormani@gmail.com

Received: September 24, 2012 Accepted: October 15, 2012 Online Published: November 29, 2012

doi:10.5539/ijel.v2n6p120

URL: <http://dx.doi.org/10.5539/ijel.v2n6p120>

Abstract

Examining and studying semantic errors in second language acquisition (SLA) is an interesting and challenging area of investigation which is still fertile requiring much more research. In spite of the considerable number of investigations into lexical errors, very few studies have tackled semantic errors committed by SL learners. Thus, this study aims at not only examining Arab learners' semantic errors in English, how they can be classified but also investigating their possible L1 and L2 sources. 30 essays written by Yemeni University third-level students majoring in English were analyzed. We have also employed a comprehensive error taxonomy based on James's (1998) and Al-Shormani's (2010). The semantic errors identified in this study were classified into three broad categories, namely, *lexical*, *collocation* and *lexicogrammatical*. Each of these categories is classified into further categories and subcategories depending on the errors identified. There were 1388 semantic errors identified in our study. It has been found that *omission of letters* category scores the highest number of errors, viz. 251, i.e. 18.08% while *misselection of a prefix* category is the lowest where only 12, i.e. (0.68%) errors were committed. A hierarchy of difficulty is established where *formal misformations* category represents the highest extreme and *formal misselection* the lowest one. It has also been found that L1 sources include translating concepts, words and phrases literally from L1, i.e. Arabic into L2, i.e. English and applying Arabic linguistic rules to English. However, L2 sources include having false concepts about English, insufficient knowledge of English semantic system and confusion about English vocabulary.

Keywords: SLA, Yemeni learners, semantic errors, classifications, sources

1. Introduction

Compared to phonological and syntactic errors, there are relatively few studies which have tackled semantic errors. This view, in fact, has been held by several researchers (e.g. Blum-Kulka and Levenston, 1979; Levenston, 1979; Channell, 1981; James 1998; Obeidat, 1986; Stieglitz, 1983; Al-Shormani, 2010; Zughoul & Abdul-Fattah, 2003; Zughoul, 1991; Laufer, 1997; Wray, 2000) who ascertain that unlike syntactic, morphological and phonological errors, investigation into semantic errors has been neglected until recently. This neglect has been due to the fact that more interesting problems do not emerge until relatively late in L2 development, whereas many studies on L2 acquisition have been concerned with beginners and intermediate learners (Obeidat, 1986). The difficulty in tackling semantic errors lies in the fact that analyzing "language beyond sentence level gets a prominent role in the study of language with the basic tenet that the study of language in context will offer a deeper insight into how meaning is attached to utterances than the study of language in isolated sentences" (Obeidat, 1986: 74). Another reason according to some researchers (e.g. James, 1998) is that semantic errors present the greatest difficulty for descriptive analysis. For instance, meaning discriminations are difficult to make in sets of words when their semantic ranges cross. Moreover, Obeidat (1986: 74) points out that "research in discourse analysis has focused primarily on one aspect of the context, i.e. the incorporation of text(s) in language analysis. This incorporation has proven to be ineffective in the interpretation of semantics." Thus, in the course of analyzing and examining semantic errors and their sources involved in this study, such errors will be the result of relating L2 words and/or patterns to words and/or patterns of L1 and/or relating L2 words or patterns to familiar L2 words and/or patterns.

In this study, we tried to provide an in-depth analysis of the semantic errors in 30 essays written by Yemeni Arabic-speaking learners of English at the University Level, classifying them and examining their L1 and L2

sources. We analyzed 30 essays written by third-year students, English Department, Ibb University, Yemen. 1388 frequent semantic errors were identified. Regarding their sources, it was found that L1, i.e. Arabic sources include transfer from L1 into L2, translating concepts, words and phrases literally from L1 into L2, i.e. English, and applying Arabic linguistic rules to English. However, L2 sources include having false concepts about English, insufficient knowledge of English semantic system and confusion about English vocabulary.

2. Defining Semantic Error

Lennon's (1991:182) defines "error" in general as "a linguistic form or combination of forms which, *in the same context and under similar conditions* of production, would, in all likelihood, not be *produced by the speakers' native speaker counterparts*" (emphasis ours). However, a semantic error as used in this study can be defined as a violation of the rules of semantic system particular to English language.

3. Methodology

3.1 Participants

This study involves 30 Yemeni Arabic-speaking learners of English selected at random from the students of third-year, English Department, Ibb University, Yemen, in the academic year 2011-2012. They aged between 23-26 years though some of them may be older but not younger than that. They are male and female, viz. 17 female and 13 male but factors such as age and sex were not considered in this study. They have studied English for about nine years (six years at school and three at university). At the university, they have studied, among other courses, four courses of academic writing, namely, Writing I, Writing II, Writing III and Advanced Composition. To differentiate between "error" and "mistake," only those recurrent errors were deemed errors. Those which were not recurrent were not considered errors but mistakes.

3.2 Procedure

30 essays about different topics such as *internet, my family, my father, my ambition, my first day at college, Yemeni Unification, marriage* were analyzed. The 30 manuscripts were corrected by the researchers (native speakers of Arabic). In fact, the researchers consulted an experienced teacher of English (an Indian Professor of Applied linguistics) when there was a difficulty in identifying an error. Only those errors agreed upon by the three were counted. If there was a problem of identifying the source of an error, the researchers asked the subjects to say what they mean by a particular sentence, phrase and/or word as the study took almost a complete semester. The researchers used a simple frequency method in counting the errors and their percentages. In the manuscripts, there were other types of errors identified such as syntactic, spelling etc. but such errors were neglected because they were not relevant to the scope of this study. Thus, the number of the essays, i.e. 30 allows for an in-depth analysis of these 30 essays so as to identify and classify the semantic errors and diagnose their sources.

3.3 Error Classifications

Some researchers (e.g. Obeidat, 1986; Khalil, 1985; James, 1998; Jiang, 2004) classify semantic errors into three broad categories, viz. **lexical**, **collocation** and **lexicogrammatical**. As far as lexical errors are concerned, several researchers have ascertained that a lexical error is the use of a wrong lexis (word) which has been inappropriately used in an utterance. To take Llach's (2005: 49) definition, a lexical error is "the wrong word use of a lexical item in a particular context by comparison with what a native speaker of similar characteristics as the L2 learner (age, educational level, professional and social status) would have produced in the same circumstances." As for the classification of **lexical errors**, researchers (e.g. Engber, 1995; Duskova, 1979; Zimmermann, 1986; Zimmermann and Schmitt, 2002; Laufer, 1991; 1997; Lennon, 1991; Zughoul, 1991; Llach, 2005; Al-Shormani, 2010, (in press); Shalaby *et al*, 2007; Sonaiya, 1991; Mahmoud, 2011) have investigated different types of errors in choosing lexis committed by SL learners. However, it should be noted here that the type of classification followed is taxonomies and not lists of errors as done in dictionaries of errors. Duskova (1979), for instance, has examined the writing of fifty Czech postgraduate learners using only four categories of lexical errors: **confusion of words with formal similarity**, **similar meaning**, **misuse of words** and **distortions among lexical nonce mistakes** (e.g. throw: through). The categories used by Duskova are obviously too broad. For example, the category **misuse of words** can, in principle, incorporate many subcategories. However, we have employed a comprehensive taxonomy based on James' (1998) and Al-Shormani's (2010). A summary of our taxonomy is presented in Table1 below:

Table 1. Error Classifications, Frequency and Percentage

Category	F	%
I. Formal Misselection		
1. Misselection of a Prefix	12	0.86%
2. Misselection of a Suffix	51	3.67%
II. Formal Misformations		
1. Direct Translation from L1	149	10.73%
2. Borrowing	33	2.37%
3. Coinage	17	1.22%
4. Distortion due to Spelling		
4.1. Omission of letters	251	18.08%
4.2. Misordering of letters	62	4.46%
4.3. Overinclusion of Letters	41	2.95%
4.4. L1-Based Spelling Errors	24	1.72%
4.5. Misselection of Letters	37	2.66%
III. Lexical Choice		
1. Assumed Synonymy	103	7.42%
2. Derivativeness	66	4.75%
3. Confusion of Binary Terms	47	3.38%
4. Paraphrase	127	9.14%
5. Idiomaticity	42	3.02%
6. Similar Forms	29	2.08%
IV. Collocations		
1. Collocate Choice		
1.1. One Collocate Incorrect	72	5.18%
1.2. Both Collocates Incorrect	51	3.67%
2. Contextualization	44	3.17%
3. Wrong Forms	23	1.65%
V. Lexicogrammatical Choice		
1. Adjectives in place of Nouns	33	2.37%
2. Nouns in place of Adjectives	16	1.66%
3. Adjectives in place Adverbs	37	2.66%
4. Adverbs in place of Adjectives	21	1.51%
Total	1388	100%

4. Results and Discussion

4.1 Formal Misselection

James (1998) has classified formal misselection errors into misselection of a prefix, misselection of a suffix and false friends, however, the errors identified in our study fall into the former two categories.

4.1.1 Misselection of a Prefix

Arabic speakers learning English commit errors in *misselection of a prefix*. As can be seen in Table 1, this category includes 12 frequent errors, i.e. (0.86%) of the total number of errors committed. The following examples exemplify such errors:

- (1)**I am nonhappy in my study.*
- (2)**Don't talk with him. He is unsane*
- (3)**This verb is unregular so you must keep it by heart.*
- (4)**We should inlarge our minds and not marry young.*
- (5)**This question is disclear for me.*

Here, the learners committing the errors in (1) through (5) fail to select the appropriate prefix to form the correct word to express themselves in such contexts. For instance, in (1), the learner uses the prefix *non-* and adds it to the adjective *happy* instead of *un-*. In (2), the wrongly used prefix, namely, *un-* which is added to the adjective *sane* instead of *in-* renders the sentence semantically deviant. In (3) through (5), the wrongly prefixed prefixes are *un-*, *in-*, and *dis-* added to the words, *regular*, *large*, and *clear* instead of *ir-*, *en-* and *un-* respectively. In fact, the addition of such wrong prefixes to such words renders such sentences semantically deviant.

A cursory look at (1) through (5) gives us a clue that the wrongly used prefixes consist of two types: (i) *verbal prefixes* as in (4) where *in-* is prefixed inappropriately to form a verb from the adjective *large*. (ii) *negative prefixes* as in the rest of the examples where such prefixes have been used to negate the adjectives in question. Now, the question worth addressing here is that what makes Arab learners commit such errors. In other words, is it L1, i.e. Arabic, transfer or L2, i.e. English influence which makes Arab learners commit such errors? Let's first examine the Arabic *words* which are used for negating adjectives. Here, the term *words* is used instead of *prefixes* due to the fact that in Arabic and specifically negating elements are considered *words* and not *prefixes*. Bearing this in mind, there are two words in Arabic which are used for negating adjectives. These are *laisa* and *yairu* (meaning not) which are not verbal negative particles.

As for how verbs are derived from adjectives in Arabic, there is only one prefix in Arabic which when prefixed to an adjective or even a noun, the resultant word is a verb, namely, *yu-* as in *yu- + jami:l* (beautiful) → *yujamil* (beautify) with some changes as it is clear in this example where the long vowel /i:/ has been changed into /i/. According to this brief description of negating and verbal derivative *words* in Arabic, Arabic has nothing to do with committing such errors as in (1) through (5). Accordingly, it is only L2, i.e. English influence which causes such errors. In that, the learner, having no sufficient competence as to which prefix to be used with which adjective, is confused and this confusion comes from English itself and hence, while internalizing the English lexicon, he/she uses these prefixes wrongly rendering such sentences semantically erroneous.

4.1.2 Misselection of a Suffix

Arab learners also commit semantic errors in selecting a suffix to be added to a word in English. This category includes 51, i.e. (3.67%) errors. The following examples illustrate the issue in question.

- (6)**Our President is helpable to his people. I thank him.*
- (7)**This indicates Dr. Mahmoud's honestness.*
- (8)**I am an ambitionable person in my life.*
- (9)**I am interesting in reading English books.*
- (10)**Marriage in Taiz is espenssiver than in Aden.*

(6) through (10) show how suffixes have been wrongly chosen. For instance, in (6), the learner committing such an error uses the suffix *-able* and attaches it to *help* instead of *-ful*. In (7), the error lies in using the suffix *-ness* attached to *honest* instead of *-ty*. (8), in fact, is more interesting where the suffix *-able* has been added to the noun *ambition*. The interesting aspect in such an error lies in the fact that there is an independent adjective of the noun *ambition* which is *ambitious* needing no suffix. In (9), the learner adds the suffix *-ing* to the verb *interest* instead of *-ed*. The difference between *interesting* and *interested* is interesting. In that, Quirk *et al.* (1985) state that *interesting* can be used to describe things, places and objects which cause interest while *interested* is used to describe people's feelings. In fact, this error has been noted to be more committed by the subjects of this study than others in the *misselection of a suffix* category where learners were not able to distinguish between the suffixes *-ing* and *-ed* as in **the book was writing by Shakespeare* or **I am visited my friends tomorrow* where the suffixes *-ing* and *-ed* are wrongly misselected, respectively. Likewise, the comparative degree suffix *-er* is

among the many errors committed in this category as in (10) above. Thus, this misselection renders such sentences semantically erroneous.

As (McCarthy, 1981; Holes, 2004) among others have pointed out, Arabic word formation/derivation belongs to what is called *nonconcatinative morphology* which depends on *root* and *pattern* mechanism while English word formation belongs to what is called *concatinative morphology* which depends on concatenating (combining) morphemes to form new words. In fact, going deeply in such a phenomenon is beyond the scope of this study. What concerns us here is just stating that L1, i.e. Arabic has nothing to do with the committing of these errors because the mechanism of word formation/derivation is different in both languages. They can be called developmental and specifically overgeneralization errors where learners overgeneralize the use of a suffix to more than one stem and this actually happens when learners are in the process of internalizing the lexicon of L2, i.e. English.

4.2 Formal Misformations

James (1998: 149) defines *formal misformations* errors as those “that produce ‘words’ that are non-existent in FL.” He adds that these words “can originate in the MT or be created by the learner from the sources of the TL itself.” In addition, according to James’s taxonomy, *formal misformations* errors are subdivided into three categories, namely, (i) *borrowing*, i.e., L1 words used in L2 without any change (for example, *I shoot him with gun in kopf* <In German *kopf* = head>). (ii) *coinage*. What James means by *coinage* is “inventing a word from L1 (for example, *Smoking can be very nocive to health* <In Portuguese *nocivo* = harmful>). (iii) *calque*. Here, James means that “if the L2 word created is the result of literal translation of an L1 word, we have a calque” (James, 1998: 150). For instance, **We have to find a car to bring us go to* <bring us to>the hospital). However, it has been found that *formal misformations* errors identified in our study fall into three categories, viz. direct translation from L1, borrowing and coinage.

4.2.1 Direct Translation from L1

Errors in this category are committed when the Arabic-speaking learners directly translate a word, phrase and/or a sentence from Arabic into English. As can be seen in Table1, this category includes 149 frequent errors, i.e. (10.73%) of the total number of errors identified in our study. Thus, such errors are exemplified in the following sentences.

(11)**In our marriage imitations, men and woman do not sit in mix.* (traditions/together)

(12)**In winter, trees had yellow papers.* (yellow leaves)

(13)**The life in the village is good.* (life)

(14)**It is my wish to go to Mareb to see its old heritages.* (remains)

(15)**Can you bring me that book? Yes, from my eyes.* (with my pleasure)

(16)**In University, people do not refuse coeducation.* (reject)

(17)**My father has many adjectives which make people love him.* (qualities)

(18)**In our department, the doctors are the best in the world.* (professors)

The underlined words/phrases in (11) through (18) are literally translated from Arabic and have been used instead of the English words/phrases in brackets. It is, in fact, the literal translation that causes these errors. In that, the learner does not have that rich repertoire of L2 lexis and what he/she does is just translate from his/her L1, i.e. Arabic what he/she thinks will express him/herself through. This is not confined to lexis but it includes also transferring complete phrases as in (12, 15&18) where the phrases, *yellow papers*, *from my eyes* and *doctors* are used respectively. In (11), for instance, the terms *imitations* and *in mix* which are direct translations of the Arabic term *staq:ali:d* and *muxtalit* have been used instead of *traditions* and *together*. In (12) the phrase *yellowpapers*, which is a direct translation of the Arabic phrase *?awra:qsa:fra:*, instead of *yellow leaves* which is the correct phrase to be used here. The term *the life*, which is a direct translation of the Arabic term *?alhia:ah* has been used instead of the English *life* in (13).

In addition, the Arabic term *?a:θa:r* meaning *remains* is used in (14) instead of the English word *remains*. In Arabic, the phrase *from my eyes* is used when someone pleasingly responds to someone asking him/her to do him/her a favor and this actually what has been done in (15) where the phrase in question has been used instead of the English *with my pleasure*. In (16 & 17) the Arabic literally translated words *refuse* and *adjectives* have been used instead of *reject* and *qualities* respectively. In (18), the term *doctors* has been used instead of *professors* which is not suitable here. The learner fails to realize that the word *doctor* is used for those who work

at hospitals. Zughoul (1991) mentions a funny phrase, namely, *Cairo envelops*, like that of (15), which has been used instead of the English *compelling circumstances* where *Cairo* (which is the capital city of Egypt) means *compelling* and *envelops* means *circumstances*. In fact, such errors result not only in semantic deviant but also in funny utterances. In this regard, Zughoul (1991: 50) states that “the choice is an equivalent for an Arabic word or an expression on the literal level, but does not convey the meaning intended in the target language. Sentences produced with this kind of error sound odd and “funny,” and sometimes such sentences are coined by proficient speakers of English as jokes and puns.” Thus, what an Arabic-speaking learner does here is actually formulate his/her sentence in Arabic, translate it and then transfer it into English. Thus, we can conclude that such errors are interlingual, i.e. the source of which is L1, viz. Arabic and that L2, i.e. English has nothing to do with the committing of such errors.

4.2.2 Borrowing

Borrowing is a strategy used by L2 learners when they feel unable to find the exact word and/or phrase in L2 to use in an L2 utterance. This category comprises 33 frequent errors, i.e. (2.37%) of the errors committed. Thus, such errors are exemplified in the following sentences.

(19)**The Arabic Nahw is easier than that of English.* (syntax)

(20)**I cannot understand the aswat of English.* (sounds)

(21)**Alsayda, Khadija was the first wife of the prophet Mohammed.* (honorable lady)

(22)**Girls in Yemen use kuhul to look beautiful.* (eye-liner)

(23)**My ambition is to go to Makka to do Hajj.* (pilgrimage)

(19) through (23) present borrowing words that do not exist in L2, i.e. English. They are, in fact, L1, viz. Arabic words though their English equivalents do exist in English. An explanation of this use could be the lack of such English words in the learners' repertoire and hence, what the learner does is just borrow Arabic words, instead. The Arabic word, *Nahw* meaning *syntax* is used in (19) because the learner lacks the word *syntax* in his/her English repertoire. Similarly, the Arabic word *aswat* meaning *sounds* is used in (20). In (21) through (23), the Arabic words *alsayda*, *kuhul* and *Hajj* have been borrowed from Arabic instead of using the bracketed words, namely, *honorable*, *eye-liner* and *pilgrimage* respectively. Thus, James (1998: 149-150) describes such a tendency stating that such errors originate in L1 of the learners. He adds that such errors are a result of “those [words] created for TL from MT resources.”

4.2.3 Coinage

This category scores (1.22%), i.e. 17 frequent errors. *Coinage* errors are committed by Arabic-speaking learners when their English repertoire does not help them express themselves when they try to communicate in English. The coined words can be of Arabic or English origin. However, the source of such errors is Arabic rule application in English and hence, L1 *per se*. These are exemplified in the following.

(24) **My sister wifed very small.*

(25) **...and foded us eggs in morning..*

(26) **...but we cand to stop here.*

(27) **...and I salyed to Allah to pass the exam.*

(28) **...then my father swaised the car.*

Examples (24) through (28) are very interesting. In that, they reflect the very low level of the learners' proficiency in English. Their source is actually the learners' L1, i.e. Arabic *per se*. For instance, *wifed* in (24) is a derived verb from the English noun *wife*. Here, the learner coins it depending on Arabic because in Arabic the verb *tazawjat* (married) is derived from the Arabic noun *zawja(h)* (wife) and as the learner does not know the English word *marry*, he/she thinks that the Arabic rule of verb derivation can be applied to English *wife*. The same thing can be said about *foded* in (25) where the error lies in coining the verb *foded* from the English *food* by applying the same Arabic analogy. In (26), the coined word *cand* is to be seen as the past form of *can* and there is an evidence which is *?an*. That is, in Arabic, the word for *could* is *?istata:ʕ* which must be followed by *?an* (to) and that exactly what happens here. In both (27) and (28), the errors are more interesting. In that, the coined words behave differently from those just discussed. In that, while the coined words in (24) through (26) are originated from English, those coined in (27) and (28) are originated from Arabic. Thus, in (27) the coined word *salyed* is the past form of the Arabic verb *yusali* (pray) in Arabic and what the learner does is coin the word and then add the *-ed* suffix to it. Similarly, the coined word *swaised* in (28) is the colloquial Arabic noun *swais*

(key) and here the learner coins it as a verb and adds **-ed** also to it. What the learner means to say in (28) is *then my father started the car with the key* something literally equivalent to the Arabic sentence *θuma ?abišayal ?alsaiarata*.

4.2.4 Distortion due to Spelling

When spelling a word wrongly, such a word will be semantically distorted and this affects the production of an utterance in which such a word is used. Thus, the third category of lexical errors according to our classification is **distortion due to spelling** errors. It has been found that **distortion due to spelling** errors committed by the subjects of this study fall into five categories, viz. **omission, overinclusion, misselection, misorderin** and **L1-based**. These categories will be discussed and exemplified as follows.

4.2.4.1 Omission of Letters

Arab learners commit spelling errors by omitting one or more letters from a word and such omission results in semantically deviant utterances. This category includes the largest number of errors committed in this study not only in **distortion to spelling** category but in the whole study scoring (18.08%) of the errors committed. The following examples demonstrate the issue in question.

(29)*My ambtion in life is to be a doctor.

(30)*In Ymenmariag is expnsiv.

(31)*Intrntconcton in the wrld.

(32)*My Englsh is not god.

(33)*I com from my vilage.

(34)*Yemeni unification has been achied in 1990.

The **omission of letters** errors in (29) through (34) vary between omitting vowel and consonant letters. In addition, this omission can be one letter as in (29) and (33) where the vowel letter *i* is omitted from the word **ambition** and the consonant letter *l* has been omitted from the word **village** or more than one as in the rest of the examples. In (30), the words **Yemen, marriage** and **expensive** have been spelt incorrectly by omitting one or more letters from them. The same thing can be said about the words in (31) through (34) where more than one letter has been omitted.

As James (1998) has pointed out, these errors are L2 based. That is, the source of such errors is not L1 rather L2 itself. In fact, the data on which he has based his conclusions are taken from Indo-European languages like German, French, Spanish etc. on whose NSs he has conducted his studies in addition to the studies he has referred to like Laufer (1990, 1991). These languages in a way or another are similar to English in their scripts, viz. Roman. This seems true, too, in the case of Semitic languages, viz. Amharic, Hebrew and Arabic. For instance, Semitic languages do not allow more than two-letter/sound consonant clusters. Accordingly, had it been a transfer from Arabic, the words **ambtion, intrnt, Englsh** which are spelt wrongly would not have been the way they are. For instance, the words **ambition** and **intrnt** contain three and four consonants in a cluster respectively and according to Arabic clustering, there must have been some vowel letters inserted in between these clusters irrespective of the vowels to be inserted to meet the Arabic clustering. Thus, the word, **ambtion**, would have been spelt as **amebtion, ambetion**, or whatever. Thus, it is untenable to say that such errors are caused by L1, i.e. Arabic interference and hence, this goes in line with what James (1998) has gone to.

4.2.4.2 Misordering of Letters

The committing of errors in **misordering of letters** consists in misordering the letters in words used in the sentence. This category includes 62 frequent errors, i.e. (4.46%). Such errors are exemplified below.

(35)*My frist at college was very nice becuase of with mny friends.

(36)*He deos not like English grammar.

(37)*Learners of English improve spet by spet

(38)*Yemeni poeple is very generous

(39)*And water in Yemen becomes littel and littel.

(40)*I recievedyuor letter yesterday alos.

(41)*I have tow sisters and tow brothers nothing eles.

Now, looking at (35) through (41), there are many words that are ‘misordered’ in spelling. These words are *frist becuase*, *mnyaydeos*, *spet*, *poepple*, *littwl*, *receivedyuor*, *alos*, *tow* and *eles* respectively. Whether these errors can be attributed to L1 or L2, James (1998: 150) ascertains that such errors are a result of “intralingual [influence] created without recourse to L1 resources. The outcomes are forms non-existent in the TL.” In this regard, Mohanty (2006: 127) supports James’s views arguing as to what makes Arab learners at all levels of English learning process, viz. “school-going children, freshmen, sophomores, senior university learners including [even] those working for their doctoral degree,” commit *misordering* errors in the cardinal number *two* as *tow*. He states that the source of such an error is not L1. In fact, Mohanty emphasizes that the source of the *misordering* of *two* as *tow* is L2, i.e. English itself. He further adds that the source of *misordering* errors lies in the fact that “the English words the Arabic speaking learners are exposed to during their courses of study show that ‘consonant-w-o’ is an extremely rare sequence of letters vis-à-vis ‘consonant-o-w’, which is very common specially at the end of words” providing a list of such words as “arrow, bow, cow, how, low, mow, now, row...” etc and because of this, Arab learners get confused in the spelling of the word *two* and consequently they spell it *tow* Mohanty (2006: 128). Accordingly, such *misordering* errors are purely intralingual caused by the developmental strategies the learners are undergone. In fact, this category includes a considerable number of words where letters are misordered such as *acheive*, *recieve*, *grwo*, *sue*, *biuld*, *gose*, *swo*, etc.

4.2.4.3 Overinclusion of Letters

This category involves 41, i.e. (2.95%) of the errors committed. *Overinclusion* errors are commonly committed by Arab learners learning English not only in the beginning stages of their acquisition but also in advanced levels. In fact, these errors could be seen as a result of overgeneralizing an existing rule in L2. The following examples illustrate this issue.

(42)**My grandmother is visitting us in our new house.*

(43)**In our new house, we have a large dinning room.*

(44)**Aziza and me sleept together yesterday.*

(45)**Ali’s writting is very nice.*

(46)**I am not listennig to English songs.*

As has been noted earlier, James (1998) states that *distortions due to spelling* errors are L2 based. In fact, Arabic language has nothing to do with the committing of errors in (42) through (46) simply because there is no rule in Arabic stating that *when a verb ends in a consonant preceded by a vowel, the last consonant is doubled* when *-ing* or *-ed* suffix is added as in *cut* → *cutting* which is why such errors occur, for instance. However, what, in fact, makes learners commit such errors is overgeneralization. For instance, in (42), the learner overgeneralizes the *doubling rule* mentioned above to the verb *visit* and thus doubles the letter *t*. The same thing can be said of the rest examples of this group except (44) where the learner overincludes the letter *e* in the past form of the verb *sleep*, i.e. *sleept*. In fact, this category includes a considerable number of errors of such kind as in *felt*, *leaf*, *saiyd*, *travel* and so on.

4.2.4.4 L1-Based Spelling Errors

This category includes 42, i.e. (1.72%) of the errors committed. Errors in this category, however, are L1-based, i.e. the source of such errors is Arabic. For the time being, consider the following examples which illustrate the issue in question.

(47)**And my friend is apsent today.*

(48)**My apmition is to be a translator.*

(49)**But there are no barks in our city.*

(50)**Yemeni people become save after unification.*

(51)**I have many friends but Sameera is the pest.*

(47) through (51) exemplify the *L1-based* errors in *distortion due to spelling* committed by Arabs learning English. Thus, opposing James’s (1998) considerations that *distortions due to spelling* errors are L2 based, the committing of such errors lies in the fact that unlike Roman languages (languages having Roman scripts such as Spanish, German, French, Czech...etc.), for Semitic languages like Amharic, Hebrew...etc., to which Arabic belongs, the issue is somehow different at least for this type of errors. As far as Arabic language is concerned, there are no sounds corresponding to English /p/ and /v/ and thus Arabic speakers learning English find considerable difficulty in learning these sounds. Accordingly, one can argue that Arabic speakers tend to

pronounce the English /p/ and /v/ as /b/ and /f/ respectively because of the L1 interference. In fact, researchers (e.g. Al-Shabbi, 1994; Mukattash, 1986; Noor, 1996; Mohanty 2006; Shalaby *et al*, 2007) hold that the absence of /p/ and /v/ in Arabic affects not only Arab learners' spoken English but extends to their writing as well. Mohanty (2006), for instance, maintains that such **spelling** errors are a result of interlingual transfer stating that "Arabic does not have /p/; so the Arabs substitute /b/ for /p/ while speaking English. This habit of speech is also extended to writing and, as a result, we get 'bark' for 'park', 'bush' for 'push.'" Mohanty (2006: 127). In addition, Shalaby *et al* (2007), state that the main reason behind such errors is **hypercorrection**. In that, the learners try to hypercorrect their **spelling** errors in writing and hence, leading to committing such lexical errors as a result of **hypercorrection**. This is exactly what happens in (47) through (51) where the words **apsent**<**absent**>, **ampition**<**ambition**>, **barks**<**parks**>, **save**<**safe**>, **pest**<**best**> are hypercorrected and hence, resulting in such lexical errors. Thus, unlike other types of errors in **distortion due to spelling**, the source of the errors in (47) through (51) is L1.

4.2.4.5 Misselection of Letters

Errors in this category are committed in such a way that the learner selects an incorrect letter for a correct one. The main reason behind such errors could be the inability to read properly. Thus, this category includes 37 frequent errors scoring (2.66%). Now, consider the following examples.

(52)**Present continuse is difficult for me.* (continuous)

(53)**Our president is very genruse.* (generous)

(54)**Salta is delituous.* (delicious)

(55)**Yemeni unification makes Yemen statase high.* (status)

(56)**English is problmatik for Yemeni learners.* (problematic)

(57)**So we have to be constuose in the class.* (conscious)

In (52) through (57), the errors exemplify the **misselection** of one or more letters in which the learners fail to select the correct ones. The correct spelling of the words intended are bracketed against each sentence above. The misselected-spelt words are **continuse**<**continuous**>, **genruse**<**generous**>, **delituous**<**delicious**>, **statase**<**status**>, **problmatik**<**problematic**> and **constuose**<**conscious**> respectively. Now, examining the source of such errors and as far as Arabic is concerned, Arabic has nothing to do here. In fact, there is no any comparison between these spelling errors and Arabic writing system. Therefore and as James (1998) has rightly pointed out, the source of such errors is English. In addition, Raimes (1985: 247) suggests that such **misspellings** errors occur because unskilled L2 writers "concentrate on the challenge of finding the right words and sentences to express their meaning instead of editing." Thus, these errors result in some words which are non-existent in L2 and hence, leading to semantically deviant sentences. Thus, the only source of such errors is L2, i.e. English.

4.3 Lexical Choice

It goes with saying that "sufficient lexical knowledge plays an important role" in the easiness or difficulty of SLA process (Al-Jarf, 2011:161) and lexical choice errors have been paid much attention to by many researchers. Thus, Lexical choice has been seen by researchers (e.g. James, 1998; Zughoul, 1991) to be an area that may be classified under IL semantics which has not received as much emphasis as the other two components of IL, namely, phonology and syntax in language learning/teaching research. Thus, Arabic speakers learning English encounter difficulties in learning English semantics and so they usually commit errors which can be attributed to the transfer from their L1, Arabic, and also to their transitional development or overgeneralization. In addition, researchers have dealt with the classification of the **lexical choice** errors. For instance, Zimmermann (1983) has classified **lexical choice** errors into four categories, namely, **paraphrase**, **wrong choice**, **verbosity** and **sense relation**. In addition, Hang (2005) has classified **lexical choice** errors into three categories, viz. **individual lexical items**, **combined lexical items** and **derivational errors**. However and as far as Arab learners who commit such errors are concerned, an attempt will be made here to classify **lexical choice** errors. As shown in Figure 1 above, Lexical Choice categories will be discussed as follows:

4.3.1 Assumed Synonymity

This category includes considerable number of errors where 103 frequent errors, i.e. (7.42%) of the errors were committed and hence, scoring the fourth rank. In fact, many researchers and applied linguists believe that a word in any language may have more than one meaning. Thus, synonymy depends on how many words that have the same meaning in a language. English is full of synonyms due to the influence of other languages and the borrowing phenomenon. In that, Zughoul (1991: 48) holds that "English is said to be very rich in synonyms

because of the French, Latin, and Greek influences on the language and because of the vast number of borrowings from different languages.” On the other hand, it has been widely argued in semantics that there is no real and exact synonymy between two or more words. In this regard, Griffiths (2006: 26-28) argues that two words may be synonymous by having the same or related meaning but when one considers *entailment* the matter differs. For instance, if one says that the words **huge** and **big** have the same or related meaning, the entailment they should represent is not true. That is, “while the sentence *The bridge is huge* entails *The bridge is big*, we do not get entailment going the other way; when *The bridge is big* is true, it does not have to be true that *The bridge is huge* (it might be huge, but it could be big without being huge)” (emphasis in the original). Thus, the following sentences exemplify the issue in question.

- (58)*Your sound is nice like that of singers. (voice)
 (59)*In Yemen ice does not fall in winter. (snow)
 (60)*After Reunification, many Yemenis go to cities to get better careers. (jobs)
 (61)*Coeducation establishes friendship between the two racess. (sexes)
 (62)*The teacher asked us to meet him when he is empty. (free)
 (63)*There are not many occupations in the village. (jobs)
 (64)*When her father comes back from his job, she looks happy. (work)
 (65)*Yemeni TV contains a lot of educational programs. (offers)
 (66)*Ali is a long man. (tall)
 (67)*I will communicate with you through email. (contact)

Words which are considered synonyms, particularly those listed in dictionaries, to which SL learners are often exposed, are in fact different in meaning in some way. As Griffiths (2006: 28-39) holds, there is a difference in style, in geographical distribution, in formality, in vulgarity, in attitude of the speaker, in collocation, in connotation, and possibly some other ways. In a great number of cases, these differences can be specified in terms of features which tend to be more language specific than universal as a foreign language learner might assume. In (66), for instance, the words **tall** and **long** are different though they have the same meaning. In that, **tall** might have the feature [+human]. A man can be **tall**, and a distance **long** in English, but both **man** and **distance** when described in terms of ‘length’ in Arabic have one word, namely, **tawi:l**, encompassing both **long** and **tall**. This actually depends on “selectional restrictions” placed on the choice of lexical items which seem so fine to the extent that the distinction between two synonyms becomes so hard to realize and hence, a SL learner thinks that he/she would use the words in the set of synonyms in an interchangeable way and thus rendering such utterances semantically deviant as presented in (58) through (67).

Thus, the *assumed synonymy* in (58) of **sound** and **voice** makes the learner committing such an error think that he/she could use one of them to express him/herself in such a context and hence, such an error can be said to be of an intralingual nature. On the other hand, considering both terms **sound** and **voice**, in English, they both have the same equivalent in Arabic, viz. **ṣawt** and here such an error could be said to be of an interlingual nature. The same thing can be said of (61), (64), (65) and (67) where the errors committed are of an interlingual nature. In (60), (59), (61), (62), (63), (66), the errors are purely of an intralingual nature due to the fact that had it been a transfer from Arabic, the errors would not have been committed because Arabic has exact equivalents of the terms mentioned. Thus, such errors can be attributed to the false conception Arab learners have about L2, i.e. English.

4.3.2 Derivativeness

There is a very fundamental difference between English and Arabic derivation. While, for instance, English adjectives can be derived from nouns, adverbs from adjectives and so on, derivation in Arabic takes a different direction. In that, Arabic depends heavily on nonconcatenative morphology in deriving words while English on concatenative morphology. In concatenative morphology, words are simply formed by combining morphemes together while nonconcatenative depends on the **root** and **pattern** mechanism (Holes, 2004). That is, from the trilateral **root** of a word, which is usually considered to be the perfect or the past tense form, words can be derived from such a **root** and some of the derivations may seem irrelevant to someone who doesn’t know Arabic. Let’s take the trilateral **root** of the verb to write **K T B**, for instance, the following paradigm is formed, **maktab** (office), **kita:b** (book or letter), **ka:tib** (clerk), **maktaba(h)** (library or bookstore), among the many others. Therefore, an Arabic-speaking learner of English hypothesizes that he/she can apply derivation in the same way it works in Arabic to English. In fact, researchers (e.g. Zughoul, 1991) stresses the Arabic derivativeness as a

peculiar feature of Arabic that has been shown to have a bearing on *lexical choice* errors made by Arabic-speaking learners of English. Thus, this category includes 66, i.e. (4.75%) of the errors committed as seen in Table 1. Now, let's consider the following examples:

(68)* *Ibb is very beautiful so I call everyone to visit it. (invite)*

(69)* *Salta is the national cookin Yemen. (dish)*

(70)* *Yemeni government executed many great developing projects. (did)*

(71)* *I always travel to college by bus. (go)*

(72)* *Spoken English is my lovely lecture. (favorite)*

(73)* *His writing is good because of his satisfaction. (persuasion)*

(74)* *I think, English is easy to educate it in Yemen. (learn)*

(75)* *I hope to save a good living for my family. (secure)*

(68) through (73) show that Arabic *derivativeness* has a great role to play in committing such errors. Thus, the learner committing the error in (68), substitutes the verb *call* for *invite* because in Arabic, the equivalent of *call* is *yad'u* which is also used in the sense of *invite*. *Cook* in (69) has been used instead of *dish* because the Arabic word of dish is *tabxa* derived from *FBX* meaning (to cook). In (70), the learner might know the verb *do* in English but under the influence of Arabic, he/she chooses the word *executed* thinking that *did* cannot be used here as projects *cannot be done* in Arabic but *executed*. *Derivativeness* is also clear in (71) where the learner uses *travel* instead of *go* because in Arabic the sense of *yaðhab* (to go) is derived from the sense of *travel*. The adjective *lovely* in (72) has been substituted for *favorite*. In Arabic the word *mahbu:b* (lovely) is derived from the two-letter root *HB* (to love) and the derived *mahbu:b* (lovely) can be used where, *lovely*, *dear*, *darling*, *favorite* etc. are used. Likewise, the word *satisfaction* is not haphazardly used in (73) but according to a hypothesis criterion. In that, the Arabic *?iqna:ʔ* (satisfaction) is derived from the root *Q N ʔ* which means both *to satisfy* and *to persuade* in English so what the learner does here is hypothesize that Arabic and English are similar in their derivation ways and that *satisfy* and *persuade* can be used interchangeably and thus such a use renders the sentence in question semantically erroneous. The same thing can be said about (74) and (75) where the words *educate* and *save* are substituted for *learn* and *secure* respectively. All in all, the errors in this category are thus L2 based resulting from hypothesis testing criterion the learner is making use of thinking that Arabic and English behave similarly in *derivativeness* aspects.

4.3.3 Confusion of Binary Terms

What is meant by this category is those errors which are made when the learners get confused about lexical items that are usually categorized as "relational opposites" (Laufer, 1997). For instance, the use of *look* and *feel* in **I look happy* and **He feels happy* respectively where the correct counterparts of such examples are *I feel happy* and *He looks happy* respectively (Al-Shormani, 2010). This category includes 47, i.e. (3.38%). Now, consider the following examples.

(76)* *Dr. Hudalearns us Spoken English. (teaches)*

(77)* *I looked happy yesterday. (felt)*

(78)* *Oh, she feels happy today. (looks)*

(79)* *But I have to return the books I lent from the library. (borrowed)*

(80)* *Can you borrow me 50 Rials. (lend)*

(81)* *Everyday I come to school on foot. (go)*

(82)* *...and these contain my typical day. (make up)*

Errors in (76) through (82) exhibit semantically erroneous utterances due to the learners' confusion in the use of *binary terms*. In (76), for instance, the error results from the confusion between *learn* and *teach* and hence, the learner substitutes *learns* for *teaches*. The other substitutes are *looked* for *felt* in (77), *feels* for *looks* in (78), *lent* for *borrowed* in (79), *borrow* for *lend* in (80), *come* for *go* in (81) and *contain* for *make up* in (82). It should be noted here that these *binary terms* reflect a relationship of oppositeness between terms in a complementary distribution sense. That is, where one of these *binary terms* is used, the other is not. As for the source of such errors, in fact, there are those which are caused as a result of L1, i.e. Arabic interference as in (76) and (82) where the learner depends on his/her L1 strategy in using *learn* instead of *teach* and *consist* instead of *makeup* respectively. This is actually due to the fact that the *binary terms*, viz. *learn* and *teach* are used in Arabic

interchangeably and thus such errors are interlingual. The same thing can be said about *consist* and *make up*. Thus, the learner committing such errors hypothesizes that this Arabic peculiar feature is shared by English as well which renders such sentences semantically erroneous. However, in (77) through (81), the errors are purely intralingual. Had it been a transfer from L1, i.e. Arabic, there would not have been errors committed and these sentences would have been semantically correct as Arabic has its equivalent terms of such English ones. Thus, such errors are a result of inadequate acquisition of such terms and their uses in the learners' developmental semantic competence in the learning process.

4.3.4 Paraphrase

Zughoul (1991) calls this category *circumlocution*. In fact, this is one of the most problematic areas for Arabic-speaking learners learning English. This is very clear from the number of errors committed in this category, viz. 127, i.e. (9.14%) constituting the highest percentage among the other subcategories in *Lexical choice*. The following sentences exemplify the issue in question.

(83)* *The women who are carrying babies should stay at home.* (pregnant)

(84)* *...and they (women) should go to doctors to make check on their bodies.* (checkup)

(85)* *Yemeni Unity is the small and big people's right.* (generations)

(86)* *Two rooms in our house are for sleeping.* (bedrooms)

(87)* *When I went abroad, I took my big bag of clothes.* (suitcase)

(88)* *My ambition is to change English language to Arabic.* (translation)

(89)* *Tomorrow, I have a party of my day I was born.* (birthday)

What makes Arab learners commit such errors is the lack of the exact word in their lexicon and thus they try to "exploit" the "paraphrase strategy" to express themselves in such contexts. In (83), for instance, the learner committing such an error does not know the word *pregnant* and instead he/she paraphrases its meaning, viz. *carrying babies* which leads to such a semantically deviant utterance. Similarly, in (84) the learner wrongly substitutes the phrase *small and big people's* for *generations*. In (85) through (89) the learners have used the phrases *for sleeping*, *big bag of clothes*, *change English to Arabic* and *day I was born* for *bedrooms*, *suitcase*, *translation* and *birthday* respectively. In fact, *paraphrase* as used here differs from frequently used types of lexical paraphrase as those of lexical corporation or decomposition (Nilsen and Nilsen, 1975) (e.g. covered with cement – cemented, learn – become knowledgeable) normally used in English. However, paraphrase as used by Arab learners in these examples is generally of the decomposition type but not normally allowed in English. It strikes anyone to think that they are "unEnglish." Accordingly, the source of such errors is L2, i.e. English due to applying such type of paraphrase and thus L1, i.e. Arabic has nothing to do with the committing of such errors. These errors may be ascribed to the learners' lack of sufficient lexical competence that enables them to form semantically well-formed sentences.

4.3.5 Idiomaticity

Idiomatic expressions are difficult to learn by all SL learners in general and Arab learners of English in particular due to their complexity. The meaning they convey is first contextualized and second cannot be composed of the sum of the meanings of each word isolatedly but from the whole as one constituent. Thus, this subcategory involves 42 errors, i.e. (3.02%) Now, let's consider the following examples.

(90)* *I always get up at 6 o'clock.* (wake up)

(91)* *...and sleep at 9 o'clock.* (go to bed)

(92)* *She heard English music.* (listened to)

(93)* *...but he was not in house.* (at home)

(94)* *Before going to college, I used to repeat my lessons.* (revise)

(95)* *When I get up I change my clothes.* (get dressed)

Zughoul (1991) holds that *idiomaticity* is related to translation. Here, the learner, unaware of the idiomatic expression in L2, tries to come up with an equivalent to the expression in his/her L1. The errors in (90) through (95) represent this tendency. For instance, in (90) the learner committing the error means to say *wake up* but instead he/she uses *get up*. In (91), *sleep* has been used for *go to bed*, *heard* for *listened to* in (92), *in house* for *at home* in (93), *repeat* for *revise* in (94) and *changed my clothes* for *get dressed* in (95). As has been stated above, these errors are related in a way or another to translation from L1, i.e. Arabic and thus such errors are of

an interlingual nature, i.e. they are of L1 source. In Arabic, there is no difference in (90) between *getup* and *wake up*, for instance, and the Arabic term used for this is *?istaiqið*. Similarly, in (91) *sleep* is used for *go to bed*. The same thing can be said about the rest of the examples in (92) through (95).

4.3.6 Similar Forms

As the name of this category implies, the errors in this category result from confusion between the incorrect and correct words to be used but in a binary relation. Confused, the learner, thus, chooses the incorrect word and substitutes it for the correct one. This category includes 29 frequent errors, i.e. (2.08%). The following sentences exemplify the issue in question.

(96)* *And it was the reason of the lost of my ambition.* (loss)

(97)* *Yemeni Unity weps discrimination between south and north.* (wipes)

(98)* *In the city, more facilities are capable.* (available)

(99)* *This will boarden the gap of difference.* (broaden)

(100)* *She will spring it for you.* (bring)

(101)* *Children through the stone on the window.* (throw)

(102)* *...but he cannot desolve any problem in his class.* (solve)

(103)* *...and she has her own internet sight.* (site)

Errors in (96) through (103) exhibit a tendency of wrongly choosing a word similar to that intended and thus resulting in a semantically deviant utterance. The learners get confused due to approximate similarity which is either graphic or phonetic. Duskova (1979) identifies this source in her study of Czech learners of English, and several other studies of learners from various language backgrounds corroborated Duskova's taxonomy. In (96) through (103), the wrongly chosen substitutes are *lost*, *weapes*, *capable*, *boarden*, *spring*, *through*, *desolve* and *sight* for *loss*, *wipes*, *available*, *bring*, *throw*, *solve* and *site* respectively. Now, looking at the nature of these errors, Arabic language has nothing to do with the committing of such errors. This also goes in line with some studies done by researchers and applied linguists (e.g. Duskova, 1979; Zughoul, 1991; Shalaby *et al*, 2007; Zimmermann and Schmitt, 2002; Llach, 2005; Hemchua and Schmitt, 2006) who ascertain that *similar forms* errors are L2, i.e. English based. According to such studies, these errors are developmental caused by the lack of the sufficient knowledge that enables the learners to differentiate between forms and their uses. Thus, the source of such errors is L2, i.e. English itself.

4.4 Collocation

Wray (2000) contends that collocational knowledge is considered a fundamental part of native speaker communicative competence. In addition, Keshavarz and Salimi (2007: 83) have ascertained that "any speech community establishes a set of idiomatic ways of expressing ideas by favoring, purely through repeated use, certain complete phrases and a great many partly filled phrase-frames" holding that collocations belong to formulaic language. Thus, it can be held that non-native speakers including Arab learners with "insufficient communicative competence find formulaic language extremely difficult." In fact, the term *collocation* was first used as a technical term by Firth saying that he "proposed it to bring forward as a technical term, meaning by collocation, and apply the test of collocability" (Firth, 1957: 194). Several researchers (e.g. McCarthy, 1981; Xiao and McEnery, 2006; Zughoul and Abdul-Fattah, 2001; Farghal and Obeidat, 1995; Mahmoud, 2005; Keshavarz and Salimi, 2007; Lewis, 1997) who define a collocation agree that it is a lexical unit consisting of a cluster of two or three words from different syntactic categories. These definitions, in fact, are just paraphrases of Firth's (1957: 183) definition that collocations are "words in habitual company." According to O'dell and McCarthy (2008: 3), "[c]ollocation means a natural combination of words; it refers to the way English words are closely associated with each other." For instance, *pay* and *attention* go together but not *give* and *attention*, *commit* and *crime* but not *do* and *crime* and *heavy* and *rain* but not *strong* and *rain* and so on. For the purpose of this study, a collocation will be defined as two words belonging to different grammatical categories to exclude binomials where the two words are from the same category and are connected implicitly or explicitly by a conjunction (e.g. and, or) as in *push and shove*, *sick and tired*, *here and there*, *in and out*, *life and death*, *dead or alive*.

Collocations, according to Keshavarz and Salimi, typically consist of nouns, adjectives, verbs, and adverbs which form "habitual company" with each other. However, the way such grammatical categories collocate fall into four categories: (i) *adjective-noun* as in *strong tea*, *heavy rain*, (ii) *verb-noun* as in *make an impression*, *make a decision*, *inflict a wound*, (iii) *verb-adverb* as in *affect deeply*, *amuse thoroughly*, (iv) *noun-verb* as in

alarms go off (ring, sound), **blood circulates** (clots, runs). As such, it has been found that the **collocation** errors committed in our study fall into three categories: 1) **Collocate choice**, 2) **contextualization** and 3) **wrong form**. Each of these categories will be classified into further subcategories.

4.4.1 Collocate Choice

In fact, this category is also classified into two subtypes, viz. (i) **one collocate incorrect** and (ii) **both collocates incorrect**. By the former is meant when one of the two collocates is incorrect, it results in an incorrect collocation. The latter means that the two collocates constituting the collocation are incorrect and hence resulting in an incorrect or semantically deviant collocation. A sample of the former is provided below.

4.4.1.1 One Collocate Incorrect

- (114)* *It is not too late to repair one's mistakes.* (correct)
 (115)* *After every lecture, I used to make my homework.* (do)
 (116)* *Everyone should work his or her work properly.* (do)
 (117)* *This car took my attention.* (attracted)
 (118)* *My brother got a killing accident yesterday.* (fatal)
 (119)* *The teacher gives us expensive advice.* (valuable)
 (120)* *We must produce our decision.* (make)

This category involves 72 frequent errors, i.e. (5.18%) which is the highest score among the other subcategories in **collocations**. In (114) through (120), the errors in the collocations, viz. **repair one's mistakes**, **make my homework**, **work his or her work**, **took my attention**, **killing accident**, **expensive advice** and **produce our decision** are a result of the first wrong collocate in each respectively which indicates that only one constituent of the collocations is incorrect. This is due to the fact that **repairs**, **make**, for instance, cannot collocate with **mistakes** and **homework** respectively. The same thing can be said about the rest of the examples. Thus, in (114) through (120), the collocates **repair**, **make**, **work**, **took**, **killing**, **expensive** and **produce** have been used instead of **correct**, **do**, **do**, **attracted**, **fatal**, **valuable** and **make** respectively. The sources of such errors vary, i.e. there are those errors which can be ascribed to L1, i.e. Arabic as in the case of (114), (115), (116) and (119). All in all, in (114), the learner transfers the Arabic word **yuslih** which means both **repair** and **correct** in English, in (115), (116), the learner transfers the Arabic term **yaʿmal** which means both **make** and **work** in English and in (119), the learner transfers the Arabic term **ya:lia(h)** (expensive) which can be used for **advice**, **clothes**, **friend** among several uses and thus these errors are of an interlingual nature where L1, i.e. Arabic, interference plays the main role in their production. However, there are those errors which can be attributed to L2, i.e. English as in the case of (117), (118) and (120). In (117), for instance, had it been a transfer from Arabic, it would have been semantically correct because in Arabic specifically, MSA, there is an equivalent of the English **attracted** which is **jaʿaba(t)** and thus such errors are of an intralingual nature where the learner has not acquired such collocational competence yet.

4.4.1.2 Both Collocates Incorrect

Unlike the errors in (114) through (120), the errors in this category are due to both words, i.e. both collocates are incorrect. This category includes 51 errors scoring (3.67%). The following sentences exemplify the issue in question.

- (121)* *We have to give concentration to our study.* (pay attention)
 (122)* *Internet is used through computers which are basic tools.* (important devices)
 (123)* *I will save my word to get my great ambition.* (keep my promise)
 (124)* *Through internet, you can easily connect persons.* (contact people)
 (125)* *...and also your related people.* (close friends)
 (126)* *In Yemen, there are tall places.* (high mountains)

(121) through (126), exemplify the phenomenon of **both collocates incorrect** in collocation formation. In (121), for instance, the learner committing such an error has wrongly selected **give concentration** instead of **pay attention**. In (122) through (126), the incorrect collocations are **basic tools**, **save my word**, **connect persons**, **related people** and **tall places** which have been used instead of **important devices**, **keep my promise**, **contact people**, **close friends** and **high mountains** respectively. These errors, in fact, result from the fact that in each collocation, both collocates cannot collocate with each other. As for what causes such errors, it should be noted

that it is L1, i.e. Arabic interference *per se* that causes such errors. In (121), for instance, it is semantically correct in Arabic to use *yuṭṭir tarki:z/?ihtima:m* (give concentration) for *pay attention*. In (122), *?a:la:tra?isia(h)* (basic tools) is used in Arabic for *important devices*. In addition, in Arabic, *yahfaḍkalimat-uh* (save my word) can be used for *keep my promise* and the same thing can be said about the rest of the collocations stated above. Thus, such errors, unlike those in *one collocates incorrect*, are L1-based, i.e. their source is L1, viz. Arabic.

4.4.2 Contextualization

Errors in this category have nothing wrong grammatically and semantically. It is only the context in which such collocations are used which results in such errors. This category includes 44 errors, i.e. 3.17%. Now, consider the following examples.

(127)* *My sister brings a boy this time.* (gives birth to a boy).

(128)* *It is very bad to lose dignity before marriage.* (lose virginity)

(129)* *Womencarrying their babies should be careful.* (pregnant women with their babies)

(130)* *I like Dr. Mahmoud because he was a very educated person.* (knowledgeable person)

(131)* *Many innocent people die because of Al-Qaida management.* (Al-Qaida organization)

(132)* *He has had economical circumstances these days.* (financial conditions)

(127) through (132) exemplify *contextualization* errors. If one closely examines these collocations, they are in themselves linguistically correct but it is only the context in which these collocations are used that makes them semantically deviant. For instance, in (127), the collocation *brings a baby* is grammatically correct but the context requires *gives birth to a boy* to be used. The same thing can be said of the other collocations wrongly used. In (128), while the learner means *lose virginity*, he/she wrongly uses *lose dignity*. When examining the source of such errors, there are those which are caused by L1, i.e. Arabic interference as in the case of (128), (129) and (131) where *lose dignity*, *carrying their babies* and *Al-Qaida management* can be used to mean *lose virginity*, *pregnant women with their babies* and *Al-Qaida organization* in English respectively. However, there are those errors whose source is L2, i.e. English as in the case of (127), (130) and (132) where learners committing such errors were not able to distinguish between *brings a boy* and *gives birth to a boy* in (127), for instance. Had it been a transfer from Arabic, the learner would have used *waladatwalad-an* (beget a boy) which is a frequent expression used in Arabic.

4.4.3 Wrong Forms

Wrong forms collocation errors consist in using one or more collocates whose grammatical form/category is not correct and hence, such collocating results in an incorrect collocation. 23, i.e. (1.65%) were found in this category. The following examples exemplify such a phenomenon.

(133)* *It is difficult to get marriage because of high dowry.* (get married)

(134)* *So he cannot overcome his familiar problems.* (family problems)

(135)* *She has finance problems.* (financial problems)

(136)* *This is a goodmusician band.* (musical band)

(137)* *Dr. Mahmoud is special person in syntax.* (specialist person)

(138)* *In the south of Yemen the socialism party was ruling before Unity.* (socialist party)

(133) through (138), exemplify the errors in *wrong forms* category where learners fail to use the correct grammatical form of one collocate to come up with a well-formed collocation. For instance, in (133), the learner fails to use *get married* instead he/she uses *get marriage*. In (134), the error lies in using *familiar problems* for *family problems* and so on as can be seen in the underlined incorrect collocation whose correct forms are bracketed against each. Now, while the errors in the previous categories have been ascribed either to L1, i.e. Arabic, L2, viz. English or both, these errors are, in fact, a result of L2 influence *per se*. In fact, Arabic has nothing to do with such collocation errors because such forms do not exist in Arabic and thus such errors can be said to be of an L2 source. As many researchers (Mahmoud, 2005; Odlin, 1989; Ringbom, 1987; Zughoul 1991; Zughoul and Abdul-Fattah, 2003; Lewis, 1997) among others have pointed out, these errors are caused by the learners' failure to comprehend such collocations and how to use the correct form required.

It has been widely stated that the most problematic issue for SL learners is that they are unable to understand how lexical collocations work in the language being learned in addition to hypothesizing that there is one-to-one correspondence between L1 and L2 James (1998: 152-54). In this regard, (Hill, 1999:5) holds that "[s]tudents with

good ideas often lose marks because they don't know the four or five most important collocations of a key word that is central to what they are writing about." As a result, they create longer, wordier ways of defining or discussing the issue, increasing the chance of further errors. He cites the example: *His disability will continue until he dies* rather than *He has a permanent disability*. This also hinders learners from being fluent in the language. Describing the importance of collocation and idiomacity in fluency of SL learners, Lewis (1997: 15) states "fluency is based on the acquisition of a large store of fixed or semi-fixed prefabricated items, which are available as the foundation for any linguistic novelty or creativity."

4.5 Lexicogrammatical Choice

Lexicogrammatical choice error can be defined as the error where a word of a particular grammatical category is used instead of another grammatical category as in *I like syntactic very much* where the word *syntactic*, an adjective is used instead of *syntax*, a noun. Thus, **lexicogrammatical choice** is a very fundamental skill that requires learners to be able to distinguish between what is a verb, noun, adjective, adverb etc. This choice of lexes also includes the ability to differentiate as to how, when and why to use a particular grammatical category in its proper position in a sentence. Obeidat (1986) holds that one feature of the **lexicogrammatical choice** errors is that they involve lexical items which are related by derivational morphology and, therefore, of different syntactic categories. The errors examined here have to do with the difference between adjectives and nouns and adjectives and adverbs. Arab learners especially those with low competence seem not to be able to distinguish between base adjectives and derived adjectives on the one hand and adjectives and nouns, on the other hand. Thus, it has been found that **Lexicogrammatical choice** errors identified in our study fall into four categories of substitution, namely, **nouns in place of adjectives**, **adjectives in place of nouns**, **adjectives in place of adverbs** and **adverbs in place of adjectives**.

4.5.1 Adjectives in Place of Nouns

Errors in this category involve placing adjectives instead of nouns. 33 errors, i.e. (2.37%) of the errors were committed in this category. The following examples illustrate the issue in question.

(139)* *He doesn't like the proud in spite of his intelligent in the class.* (pride, intelligence)

(140)* *Nowadays, there is no different between man and woman.* (difference)

(141)* *There is no happy in this life.* (happiness)

(142)* *In spite of its significant and necessary, marriage also has its disadvantages.* (significance, necessary)

(143)* *There is large distant between college and our house.* (distance)

Errors in (139) through (143) exemplify clearly how the learners committing such errors substitute the adjectives **proud**, **intelligent**, **different**, **happy**, **significant**, **necessary** and **distant** for the nouns **pride**, **intelligence**, **difference**, **happiness**, **significance**, **necessity** and **distance** respectively. Now, considering the source of such errors, Arabic has nothing to do with the committing of such errors because Arabic does not allow such constructions. Consequently, these errors are L2 based caused by the insufficient knowledge the learners have which does not enable them to differentiate between nouns and adjectives and how to use them properly.

4.5.2 Nouns in Place of Adjectives

Errors in this category are the opposite of those in the category just discussed. 16 errors, i.e. (1.66%) were committed in this category. Now, consider the following examples which demonstrate the issue in question.

(144)* *But my brother does not want to get marriage.* (married)

(145)* *Ali was absence yesterday.* (absent)

(146)* *Many Yemenis get divorce because of family problems.* (divorced)

(147)* *She is not very beauty.* (beautiful)

(148)* *After unification, many people feel happy and pleasure.* (pleased)

(149)* *Our grandparents are old but wisdom.* (wise)

(144) through (149) show how the learners committing them fail to use adjectives as required by the syntactic contexts in which they are used. The learners substitute the nouns **marriage**, **absence**, **divorce**, **beauty**, **pleasure** and **wisdom** for the adjectives **married**, **absent**, **divorced**, **beautiful**, **pleased** and **wise** respectively. Again, Arabic has nothing to do with the committing of such errors because Arabic, in fact, does not allow such semantically deviant structures. Therefore, these errors are intralingual made because of the confusion the learners get while being unable to distinguish between the forms of a word, viz. when it is a **noun** and when it is an **adjective** and

the reason of which might be having false conception about L2 lexical and grammatical systems. They are committed while the learner tries to internalize the L2 linguistic system which may be overcome in later stages if they are not to be fossilized.

4.5.3 Adjectives in Place of Adverbs

Errors where adjectives are used wrongly in place of adverbs are common among Arab learners of English. Obeidat (1986) holds that such errors are a result of not knowing where to use an adjective or an adverb though such constructions are not allowed in Arabic. This category includes 37 errors, i.e. (2.66%). Now, consider the following examples.

(150)* *I run to school quick.* (quickly)

(151)* *Only few learners in our college speak English fluent.* (fluently)

(152)* *...and this person talks very polite.* (politely)

(153)* *I hope the world lives peaceful.* (peacefully)

(154)* *We have to think careful before answering any question.* (carefully)

In (150) through (154), the learners committing such errors fail to distinguish an adjective from an adverb. Thus, the adjectives **quick**, **fluent**, **polite**, **peaceful** and **careful** have been substituted for the adverbs **quickly**, **fluently**, **politely**, **peacefully** and **carefully** respectively. Here, too, Arabic has nothing to do with the committing of these errors. In other words, transfer from L1, i.e. Arabic does not play any role in committing such errors because Arabic does not allow such semantically deviant structures and hence, these errors are L2 based which can be ascribed to the insufficient competence of the learners at this stage where the learners are still internalizing the L2 linguistic and semantic systems.

4.5.4 Adverbs in Place of Adjectives

Unlike the category just discussed, errors in this category present the difficulty Arab learners encounter when using adjectives where they use adverbs instead. There were 21, i.e. (1.51%) committed in this category. The following examples show such difficulty.

(155)* *He is very happily today.* (happy)

(156)* *... and we can find the person who is politely.* (polite)

(157)* *They have deeply feelings in their hearts.* (deep)

(158)* *You look sadly today.* (sad)

(159)* *Why are you tensely today?*

The learners committing the errors in (155) through (159) fail to use the correct grammatical categories, viz. adverbs instead of adjectives. Thus, the wrongly used adverbs are **happily**, **politely**, **deeply**, **sadly** and **tensely** which have been used for the adjectives **happy**, **polite**, **deep**, **sad** and **tense** respectively. Like other categories discussed above, these errors are L2 based, i.e. developmental resulting from the overgeneralization of using adverbs over adjectives within the L2 itself. In short, the **lexicogrammatical choice** errors represent the learners' confusion of how, when and why they have to use a certain grammatical category instead of another and being of an L2 nature, these errors also represent the insufficient lexicogrammatical competence which is not what is needed to overcome such difficulties.

5. Conclusions and Pedagogical Implications

In the course of this study, semantic errors committed by Yemeni learners are classified into several categories and subcategories. In general they are classified into three major categories, viz. **lexical**, **collocational** and **lexicogrammatical**, each of these categories has been classified into further categories and subcategories. There were 1388 semantic errors identified in this study. The number of errors committed in each category and/or subcategory varies from a category and/or subcategory to another. It has been found that **omission of letters** category scores the highest number of errors, viz. 251, i.e. 18.08% while **misselection of a prefix** category is the lowest where 12, i.e. (0.68%) errors were committed. A hierarchy of difficulty can be seen in Table 1 above where **formal misformation** is the most difficult area for Arab learners of English. **Lexical choice** is the second most difficult area for them and so on. The least difficult area for them is **formal misselection**. This could be taken into account and paid much attention to by Arab and non-Arab applied linguists, scholars, researchers and University teachers in choosing which semantic area they should pay attention to more than any other in designing their courses and even in presenting their such semantic units in classroom.

As for the sources of the semantic errors committed by Yemeni Arabic-speaking learners, there are two different sources of such errors, namely, L1, i.e. Arabic and L2, i.e. English. As for L1 based sources, it has been found that the sources of these errors vary between following different strategies such as translating from Arabic as in the case of some categories in **lexical choice** and **collocation** errors, applying Arabic rules to English as in **derivativeness**, the Arabic sound system as in the case of the absence of /p/ and /v/ in **distortion due to spelling** errors among others. As for the L2, i.e. English based sources, the main source is having false conceptions of the way meaning networks work in English as in the case of **paraphrasing**, **collocation** and **lexicogrammatical choice** errors. In addition, L2 based errors are obvious as in the case of **misordering**, **overinclusion in distortion due to spelling**, **formal misselection** and **lesicogrammatical** errors. It has also been found that the common L2 source is the insufficient knowledge Arab learners have about English semantic system which makes them commit such errors. However, it has been found that there is some kind of interrelated or interchangeable source which can be called **hypothesis testing** strategy followed by Arab learners of English. This strategy can be applied both to L1 as in the case of errors committed in **derivativeness** and L2 as in the case of errors committed in **synonymity**. Based on these findings, Arab and non-Arab applied linguists, scholars, researchers and University teachers should pay attention to how to make their students aware of the differences between L1 and L2 to minimize their roles in committing such errors.

Acknowledgments

The authors would like to gratefully thank the anonymous reviewer(s) for their valuable comments and suggestions that contribute to the development of this study. The study has benefited substantially from their suggestions and the IJEL editor's comments.

References

- Al-Jarf, R. (2011). Teaching English Word-Formation Processes to Translation Students. *kommunikacija i kultura online*, 2(2), 160-174.
- Al-Shabbi, A. (1994). Communicative Approach to Spelling for Arab Learners. *King Saud University Journal*, 6(2), 21-33.
- Al-Shormani, M. (2010). *L2 Acquisition and Syntactic and Semantic Error Gravity in University Learners' English of the Arab World* (Unpublished Ph.D. Dissertation). Hyderabad University, India.
- Al-Shormani, M. (in press). *Sources of Syntactic Errors in Yemeni Learners' English Compositions*.
- Blum-Kulka, S., & Levenston, E. (1979). Lexical Simplification in Second Language Acquisition. *Studies in Second Language Acquisition*, 2(2), 43-65. <http://dx.doi.org/10.1017/S0272263100000073>
- Channell, J. (1981). Applying Semantic Theory to Vocabulary Teaching. *ELT*, 35(2), 115-122.
- Corder, S. (1981). *Error Analysis and Interlanguage*, Oxford: Oxford University Press.
- Dulay, H., & Burt, M. (1974). Natural Sequences in Child Second Language Acquisition. *Language Learning*, 24, 235-243. <http://dx.doi.org/10.1111/j.1467-1770.1974.tb00234.x>
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. New York: Oxford University Press.
- Duskova, L. (1979). On Sources of Errors in Foreign Language Learning. *IRAL*, 7(1), 11-36.
- Engber, C. A. (1995). The Relationship of Lexical Proficiency to the Quality of ESL Compositions. *Journal of Second Language Writing*, 4(2), 139-155.
- Farghal, M., & Obeidat, H. (1995). Collocations: A Neglected Variable in EFL. *IRAL*, 33(4), 315-31. <http://dx.doi.org/10.1515/iral.1995.33.4.315>
- Firth, J. (Ed.). (1957). *Papers in linguistics*. Oxford: Oxford University Press.
- Griffiths, P. (2006). *An Introduction to English Semantics and Pragmatics*. Edinburgh: Edinburgh University Press.
- Hang, J. (2005). *A Study of Lexical Errors in Cantonese ESL Learners' Writing* (Unpublished Master's Dissertation). University of Hong Kong, China.
- Hemchua, S., & Schmitt, N. (2006). An Analysis of Lexical Errors in the English Compositions of Thai Learners. *Prospect*, 21(3), 2-24.
- Hill, J. (1999). Collocational competence. *English Teaching Professional*, 11, 3-6.
- Holes, C. (2004). *Modern Arabic: Structures, Functions, and Varieties*. Washington, DC: Georgetown University

Press.

- James, C. (1998). *Errors in Language Learning and Use Exploring Error Analysis*. Harlow: Longman.
- Keshavarz, M., & Salimi, H. (2007). Collocational Competence and Cloze Test Performance: A Study of Iranian EFL Learners. *IRAL*, 17(1), 81-92.
- Laufer, B. (1997). Some Properties of the Foreign Language Learner's Lexicon as Evidenced by Lexical Confusions. *IRAL*, 29(4), 317-330.
- Laufer, B. (1990). Why Are Some Words More Difficult than Others? Some Intralexical Factors that Affect the Learning of Words. *IRAL*, 28(4), 293-307.
- Laufer, B. (1991). Some Properties of the Foreign Language Learner's Lexicon as Evidenced by Lexical Confusions. *IRAL*, 29, 317-330.
- Lennon, P. (1991). Error: Some Problems of Definition, Identification, and Distinction. *Applied Linguistics*, 12(2), 180-195. <http://dx.doi.org/10.1093/applin/12.2.180>
- Levenston, E. (1979). Second Language Acquisition: Issues and Problems. *The Interlanguage Studies Bulletin*, 4, 147-160.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Hove: Language Teaching Publications.
- Llach, M. P. (2005). The Relationship of Lexical Error and their Types to the Quality of ESL Compositions: An Empirical Study. *PortaLinguarum*, 3, 45-57.
- Mahmoud, A. (2000). Modern Standard Arabic vs. Non-Standard Arabic: Where Do Arab Learners Transfer From? *Language, Culture and Curriculum*, 13, 126-136. <http://dx.doi.org/10.1080/07908310008666594>
- Mahmoud, A. (2005). Collocation errors made by Arab learners of English. *Asian EFL Journal*. Retrieved May 13, 2012, from <http://www.asian-efl-journal.com>
- Mahmoud, A. (2011). The Role of Interlingual and Intralingual Transfer in Learner-centered EFL Vocabulary Instruction. *AWEJ*, 2(3), 28-49.
- McCarthy, J. (1981). A Prosodic Theory of Nonconcatenative Morphology. *Linguistic Inquiry*, 12, 373-418.
- Mohanty, P. (2006). A Unique Error in the Arabic Speakers' Written English: A Search for Its Sources. *Indian Linguistics*, 67(1), 127-128.
- Mukattash, L. (1986). Persistence of Fossilization. *IRAL*, 24(3), 187-203.
- Nilsen, D., & Nilsen, A. (1975). *Semantic Theory: A Linguistic Perspective*. Rowley, Massachusetts: Newbury.
- Noor, H. (1996). *English Syntactic Errors Committed by Arabic Speaking Learners of English*. King Abdulaziz University, Educational Science. *Scientific Publishing Centre*, 1441-1465.
- Obeidat, H. (1986). *An Investigation of Syntactic and Semantic Errors by Arab Learners* (Doctoral dissertation). University of Illinois at Urbana-Champaign.
- Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Ringbom, H. (1987). *The Role of First Language in Foreign Language Acquisition*. Clevedon: Multilingual Matters.
- Schmitt, N., & Zimmerman, C. (2002). Derivative Word Forms: What Do Learners Know?. *TESOL Quarterly*, 36, 145-171. <http://dx.doi.org/10.2307/3588328>
- Shalaby, N., Yahia, N., & El-Komi, M. (2007). Analysis of Lexical Errors in Saudi College Learners' Compositions. King Saud University. *Journal of Applied Linguistics*, 21, 17-32.
- Sonaiya, R. (1991). Vocabulary Acquisition as a Process of Continuous Lexical Discrimination. *IRAL*, 14, 269-284.
- Stieglitz, E. (1983). A Practical Approach to Vocabulary Reinforcement. *ELT*, 37(1), 71-75. <http://dx.doi.org/10.1093/elt/37.1.71>
- Wray, A. (2000). Formulaic Sequences in Second Language Teaching: Principles and Practice. *Applied Linguistics*, 21(4), 463-489. <http://dx.doi.org/10.1093/applin/21.4.463>

- Xiao, Z., & McEnery, A. (2006). Collocation, Semantic Prosody and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics*, 27(1), 103-129. <http://dx.doi.org/10.1093/applin/ami045>
- Zimmermann, R. (1986). Semantics and Lexical Error Analysis. *EnglischAmerikanischeStudien*, 2(86), 294-305.
- Zughoul, M. (1991). Lexical Choice: Towards Writing Problematic Word Lists. *IRAL*, 29(1), 45-60.
- Zughoul, M., & Abdul-Fattah, H. (2003). *Collocational Strategies of Arab Learners of English: A Study in Lexical Semantics* (ERIC Document Reproduction Service No. ED 479746).