

Collocability Difficulty: A UG-based Model for Stable Acquisition

Mohammed Q. Shormani

Assistant Professor of Linguistics, College of Arts, Ibb University, P.O. Box: 70270, Ibb, Yemen

Email: dshormani@gmail.com

Abstract

An L2 learner is distinguished from another inasmuch as he/she has a good command on formulaic language like collocations which are deemed to be very specific to native speakers of a language being part of their competence and intuition. If this is true, it follows that English collocation, for instance, will be difficult for those learning it as an L2. Thus, this study is set to provide empirical evidence for collocability difficulties encountered by Yemeni Arabic-speaking advanced learners of English through investigating collocation errors deeply probing their sources, consequences and the cognitive strategies employed while acquiring such collocations. Sixty essays selected randomly from 107 given to Yemeni advanced learners majoring in English as home assignments. After identifying, classifying and tabulating errors, sources were classified into four categories, viz. *L1-transfer*, *L2-influence*, *mutual* and *unrecognized*. (30%) of the errors committed were ascribed to *L1-transfer*, (54%) to *L2-influence*, (13%) to *mutual* and (3%) to *unrecognized*. Based on such findings, I have proposed a UG-Model for learning collocations based on *subcategorization* and *selectional* restrictions. Though Yemeni learners are involved in this study, this model could be utilized for learning collocations in similar ES/FL contexts.

Keywords: Collocability Difficulties, Error Sources, UG-Model, Subcategorization and Selectional Restrictions

1. Introduction

The fact that what distinguishes an L2 learner from another is his/her acquisition of a larger amount of lexis which enables him/her to express him/herself proficiently, appropriately and accurately according to the situation he/she is in can never be denied. It is widely held that vocabularies are considered the building blocks of coherent and cohesive (con)text, meaning and hence, language as a whole. If this is true, it follows that it is the more words an L2 learner has and the more he/she is able to access such an aspect of knowledge what makes such a learner effectively incorporate new concepts into existing conceptual schemata and hence, producing pieces of language¹ suitable for a particular context surpassing his/her peers (Hennings, 2000; Sonaiya, 1991; Llach, 2005, Shormani, 2012a, 2012b, 2013a; 2014; Zughouli, 1991; James, 1998; Khalil, 1985; Han, 2004; Daneman, 1988; McCarthy, 1990; Saville-Troike, 2006; Laufer, 1986, 1990, 1997; Wang and Wen, 2002; Bahns, 1993, among others). This is true especially if communication impediment is to be taken into consideration for the fact that lexis and lexis alone are what results in such a problem. Takac (2008, p. 10) adds that the “knowledge of conceptual foundations that determine the position of the lexical item in our conceptual system” is so crucial and so is such knowledge manifested in different contexts through the use of lexis.

There are two types of vocabulary to learn in an SLA context: single lexis and words in company the latter of which is the main concern of this study, yet, excluding fixed expressions, viz. idioms and/or proverbs. In fact, the first to introduce the term “collocation” was Firth defining it as “the habitual co-occurrence of lexical items” (Firth, 1957, p.196). In addition, it has been claimed that their knowledge is confined to a native speaker of a particular language (McCarthy, 1990). To me, as it stands, the best definition for what a collocation is is provided by O’dell&McCarthy (2008, p. 3) as “a natural combination of words; it refers to the way English words are closely associated with each other.” If this is true, it follows that as far as second language acquisition (SLA) process is concerned, learners find collocations rather difficult constituting an obstacle to fluency and accuracy. In this regard, Lewis (1997, p. 15) states that “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.” What is more is that very good L2 learners are deemed to lose marks not because they fail to express themselves in a particular situation but because they find themselves unable to appropriately use the most common collocations in English (Hill, 1999; James, 1998).

If our above argument is true, it follows that L2 learners of whatever language backgrounds encounter a considerable difficulty not only in acquiring English collocations but a severe one in using them and consequently it follows that they certainly commit errors. It is also true that there are relatively few studies in this area of SLA research. This is due to the fact that lexical error studies in general and collocational in particular are difficult to handle for several reasons the most important of which is that lexical knowledge is difficult to assess (Shormani, 2012a) and there are no rules that could be followed in assessing it. Fewer, if any at all, are those studies deeply probing their sources and consequences, however.

Thus, this study is set to seek answers to questions such as why Arabic-speaking learners including Yemenis commit collocation errors, is collocation knowledge actually confined to native speakers as claimed by several scholars and linguists, does either L1, i.e. Arabic or L2, i.e. English have a role to play in that, and if so, to what extent each, are there any other sources, what roles do learning strategies play in committing such errors, among

others? However, most importantly, and based on the results concluded with, the main purpose is to seek an essential solution to such difficulties helping Arabic-speaking learners overcome such errors or at least reduce their occurrence. For this, 60 essays selected randomly from 107 were analyzed; the errors identified were classified into two major categories, namely, lexical and grammatical. Each category is in turn divided into subcategories. The sources were classified into four categories, viz. *L1-transfer*, *L2-influence*, *mutual* and *unrecognized* sources.²The learning strategies identified include 1) *L1 interference* represented by negative transfer of Arabic collocability into English, hypothesized one-to-one correspondence between L1 and L2, false equivalence, bilingual dictionaries, 2) *L2 influence* represented by misconception, insufficient knowledge in English collocates, internalizing L2 lexicon system, lack of exposure to English, overgeneralization which are both cognitive, and 3) compensative like guessing (Richards, 1974; Oxford, 1990; O' Malley and Chamot, 1990).

2. Previous studies

It goes without saying that L2 learners of whatever L2s employ different cognitive strategies in their acquisition. These strategies may or may not result in errors. To figure out what learning strategies made use of when acquiring English collocations, one should look at the nature of such collocations. In fact, English is characterized by many linguists, researchers and SLA scholars (e.g. Keshavarz&Salimi, 2007; Hill, 1999; Ellis, 1997;Lien, 2003; Li, 2005; Shormani, 2012a; O'Dell & McCarthy, 2008; Lewis, 1997; Mahmoud, 2005; Firth, 1957; Griffiths, 2006; Naba'h, 2011, among many others) to be very rich in collocations of several types, viz. verbal, nominal, adjectival, etc. They have also ascertained that it is a characteristic of native speakers *per se* to master such collocations. If this is true, it follows that learning collocations by L2 learners of English would be expected to be of a considerable difficulty. Such a difficulty lies in the fact that their nature is so complicated for such learners including Yemenis. This has also been proved true by a considerable number of SLA scholars, applied linguists (Naba'h, 2011; Keshavarz&Salimi, 2007; Martynska, 2004; Tang, 2004; Zughoul& Abdul-Fattah, 2003; Mahmoud, 2005; Shormani, 2012a; Shormani &Sohbani, 2012, among others).

As far as the unintelligibility imposed by collocation errors is concerned, some scholars (e.g. Tang, 2004; Shormani, 2012a; Khalil, 1985; Hang, 2005; Zughoul& Abdul-Fattah, 2003) ascertain that lexical errors in general compared to other types of errors are considered the most serious as far as native speakers' judgment is concerned. For instance, Shormani (2012a) stresses that it is by the choice of words, communication is more hindered than it is by any other type of errors. For some others (e.g. Martynska, 2004; Tang, 2004), meaning of lexis does not matter as manifested by the habitual company this lexis occurs with and how contextualization plays a crucial role in such a use. What contextualization simply implies is collocability (the underlying ability of a particular word to co-occur with others). There are also those (e.g. Shormani, 2013a; Han, 2004) who see collocations as one of the obstacles behind L2 learners stopping short (fossilized) of a native-like proficiency for the fact that there is no rule, logic and/or even a particular way/technique to follow to understand and master collocation use and because of this, collocability is considered part of native speaker competence and intuition as maintained above.

Regarding their difficulty for L2 learners, several scholars (e.g. Farghal and Obiedat, 1995; Channell, 1981; Bahns&Eldaw, 1993; Gitsaki, 1997) maintain that an L2 learner may use a word in its improper context taking into consideration meaning alone without paying attention to the collocation range of a particular lexis which actually results in a piece of language being erroneous. Several researchers and SLA scholars (Liu, 1999; Yuan & Lin, 2001; Lien, 2003; Shormani, 2012a; Shormani &Sohbani, 2012) argue that collocation knowledge helps L2 learners have a good command on communication and develop a native-like proficiency.

However, lexical error studies are relative few compared to those done on other types of errors such as syntactic, phonological, morphological, etc. (Tomasello, 2007; James, 1998; Han, 2000, Cook, 2003; Taylor, 1976;Gass&Selinker, 2008; Laufer, 1986, 1990, 1997), fewer are those studies tackling collocation errors (Shormani, 2012a, 2012b, Shormani &Sohbani, 2012; Rababah, 2003; Zughoul, 1979; Mukattash, 1979, 1981; Mahmoud, 2005, 2011;Wray, 2000; Taylor, 1986) and fewest, if any at all, are those deeply probing the sources of such errors and their consequences (Shormani &Sohbani, 2012; Mahmoud, 2005). This is due to two major reasons: the first is the fact that studying collocation errors and probing deeply their sources is much challenging and complicated for their nature and what information they involve in addition to the difficulty in assessing them. The second is that studying and investigating what causes an error or a group of errors, deeply probing their consequences is a psycholinguistic task and hence, requiring a linguistic, psycholinguistic, pragmatic, both L1 and L2 acquisition knowledge and knowledge of the cognitive strategies L2 learners make use of consciously and unconsciously (Shormani, 2012b, in press). Thus, if scholars, SLA researchersand/or linguistsare fully aware of all these requirements, they will have the ability enabling them to deeply probe what goes wrong with what that makes such learners commit such errors and provide deep-rooted solutions to them.

Thus, studies concerning providing psycholinguistic investigations of the sources of collocation errors committed by Arab and/or Yemeni learners are relatively rare, if any at all. The available studies on the issue merely allude to the sources of such errors providing a hazy picture ascribing such errors to two sources, viz. L1 and L2 (cf. Shormani &Sohbani, 2012; Ridha& Al-Riyahi, 2011; Mahmoud, 2005). Those studies include, for

instance, Shormani &Sohbani (2012) who have done a study on Yemeni learners very briefly alluding to the sources of collocation errors for their major purpose was something else.³ They ascribe such errors to L1 and L2. Moreover, a study has been done by Ridha& Al-Riyahi (2011) whose main concern is classifying lexical collocation errors. They again ascribe such errors as they classify them, viz. interlingual and intralingual. Another study has been done by (Mahmoud, 2011) in which the researcher has classified the errors of vocabulary involved in his study and ascribed them to L1 and L2 transfer. Shormani(2014)criticizes the use of the term 'L2 transfer' claiming that it is unclear as to what type of knowledge that could be transferred from English into English and as the learner gets influenced by L2 and hence, the term L2 influence should have been used. Another study tackling lexical errors has been done by Zughoul (1991), in which he identifies collocation errors as in **efficient money* meaning *sufficient money*,claiming that it is L2 whichmakes students commit such an error. However, his classification was not satisfactory due to the fact that he ascribes almost all lexical errors including collocation to L1 *per se* which is due to his influence by Duskova (1969) whose results were based on a study done on Czech, viz. an Indo-European language, i.e. of the same language family like English and hence, different from Semitic languages including Arabic (Shormani, 2014). Mahmoud (2005) has done a study on lexical errors including collocation ones committed by Arab learners of English attributing them to Arabic and English where the former surpasses the latter, viz. scoring (61%) of the total. Khalil (1985) has done a study examining the issue of how communication is hindered by lexical errors including collocational trying to evaluate such hindrance. However, a very crucial conclusion is his affirmation that collocation errors hinder communication most.

3. The Present Study

As has been stated earlier, the study at hand aims at deeply probing the psycholinguistic sources of collocation errors and their consequences, investigating the cognitive strategies Arab learners employ in their acquisition of English collocability and seeking appropriate solutions to such difficulties.

3.1. Method

3.1.1. Participants

The most suitable subjects for lexical errors in general and collocational in particular are advanced learners due to the fact that they are expected to have reached a considerable level of proficiency. Such learners are expected to have been exposed to knowledge-based courses like phonology, morphology, semantics, syntax, in addition to literature ones like novel, drama, poetry, etc. Thus, this study involves 60 essays selected at random from 107 essays given to would-be graduates majoring in English as home assignments, in their second semester, fourth-year, in the academic year 2012-2013. Their ages range from 24-28. They were male and female, viz. 40 female and 20 male. They have studied English for about 10 years and as far as university education is concerned, they have studied the courses mentioned above in addition to skill courses including *reading, spoken, writing* and courses like *grammar, vocabulary* among others.

3.1.2. Procedure

The participants of this study were left free to choose the topics to write about and hence, coming with such titles as (*Internet and Globalization, Yemeni Unity, My Ambition, My Family, Ibb City in Summer, Mareb Civilization, My English Study*, etc.) The total words were 18819 and the mean length of the 60 essays was 313.65 words (SD= 68.13, min= 193, Max= 479). The researcher (a native speaker of Arabic) corrected the participants' manuscripts. In fact, some errors imposed some kind of unintelligibility and in this case, the learner committing such errors was called and asked to state what he/she meant by the collocation involving the error(s). After error identification, the error(s) was classified and counted in the (sub)category it belonged to. The errors I had a doubt about were given to two consultants, viz. an Indian Professor of Applied linguistics and an American native. Thus, only those errors to which we all agree were counted and those being of some kind of doubt for the three of us were excluded from our corpus.

4. Classification of Collocation Errors

Collocation errors were classified into only two categories, viz. *interlingual* and *intralingual* (Mahmoud, 2005, 2011; Hemchua& Schmitt, 2006; Shormani, 2012a, 2012b; Shormani &Sohbani, 2012). Moreover, basing her classification on three criteria, namely, linguistics, psycholinguistic and pedagogical, Llach (2005) classifies them into four major categories, viz. semantic, syntactic, orthographic and pragmatic each of which is subclassified into subcategories. Her study's scope, however, was to investigate the relationship correlated between lexical errors and quality of composition. It is based on two criteria: linguistic and pragmatic. The former refers to Corder's (1973, p. 133) sense of the word having "semantic, syntactic and phonological properties." Llach (2005) adds pragmatics which deals with the use of the word, i.e. in Chomsky's (1968) sense that it does not suffice to only know "competence" but also the ability to use it in a context "performance" within which most of the lexical errors actually lie. What is more is that even those studies are not satisfactory due to the fact that "error categorization frameworks used... in [such] studies have addressed only a relatively limited number of lexical error categories" (Hemchua& Schmitt, 2006, p. 3).

Moreover, some kind of error category overlap in our error classification occurs here and there but we have done our best to make our error-bank as accurate as possible. Thus, defining what a collocation error underlying the study at hand is worth addressing here. In fact, the term “error” has been defined by many scholars and linguists (e.g. Lennon, 1990, 1991; Shormani, 2012a; James, 1998) trying to distinguish it from “mistake.” The latter is seen as a deviation belonging to “performance” and the former to “competence.” This makes it clear that there is some kind of lack of competence and/or knowledge underlying error committing being one of the concerns. The latter, however, is ascribed to “performance” (Chomsky, 1968; Corder, 1973; Shormani, 2012a, among many others). Thus, for the purpose of this study, a collocation error could be defined as a recurrent deviation committed at the level of collocability as a result of violating predictability and knowledge peculiar to native speakers of English and hence, all other errors of whatever type they are are excluded. Thus, Figure 1 below presents the only types of collocation errors involved in this study.

Figure 1: Summary of Collocation Errors

I Lexical Collocations

- 10. One Incorrect
- 11. Both Incorrect
- 12. Contextually Odd
- 13. Incorrect Category

II Syntactic collocations

- 1. Omission of a Prep.
- 2. Misuse of a Prep

5. Results and Discussion

As Table 1 below shows, different issues such as, error hierarchy, error seriousness, taxonomy and error source(s) could be utilized but due to being concerned merely with collocation errors, such issues are excluded.

Table 2: Lexical Errors: Category, Source, Frequency and Percentage

Category	L1-transfer		L2-influence		Mutual		Unrecognized		Total	
	F	%	F	%	F	%	F	%	F	%
lexical Collocations										
One Incorrect	37	29%	74	57%	15	12%	3	2%	129	18%
Both Incorrect	31	26%	62	53%	20	17%	4	3%	117	16%
Contextually Odd	33	30%	68	62%	8	7%	0.00	0%	109	15%
Incorrect Category	21	24%	54	61%	13	15%	0.00	0%	88	12%
Syntactic collocations										
Misuse of a Prep.	51	35%	66	45%	21	14%	8	45%	146	20%
Omission of a Prep.	46	32%	74	52%	17	12%	5	4%	142	19%
Total	219	--	398	--	94	--	20	--	731	100%
%	30%	--	54%	--	13%	--	3%	--	100%	--

5.1. Lexical Collocations

In this study, the concept of lexical collocation is one where two words belonging to different grammatical classes, and such ones where the lexically grammatical output collocate is made of two or more words but of the same grammatical class will be excluded as in **In our village, we have a powerful station* where the resultant grammatical collocate is *power station* being of N+N type. Thus, as mentioned earlier, “cf. Figure 1” collocation errors are classified into four categories, namely, *one incorrect*, *both incorrect*, *contextually odd* and *incorrect category*. Every collocation given in the course of this study is also confined to only two collocates.

5.1.1. One Incorrect

What is meant by this category is that one word of the collocation is not correct. This category includes errors ascribed to different sources. For instance, *L2-influence* includes 74 errors, i.e. (58%) which is followed by *L1-transfer* with 37 errors, i.e. (28%) which is followed by *mutual* with 15 errors, i.e. (12%) and the last one is *unrecognized* with 3 errors. i.e. (2%). Examples in (1-4) illustrate such an issue.

- (1) **...and our village has kind weather.* (nice)⁴
- (2) **I feel happy when the teacher gave questions to us.* (asked)
- (3) **Yemen has many well-known castles in Mareb, in Sana’a, in Ibb and in Haja.* (famous)
- (4) **...and separation will make us lack unity.* (lose)

In (1) through (4), the error lies in the fact that *kind* is collocated with *weather* where *kind* is the wrong collocate being not possible in English.⁵ This error is ascribed to *L1-transfer* simply because the equivalent of the word *kind*, i.e. *latiif* is possible in Arabic. The learner here influenced by Arabic-English dictionaries hypothesizes the

existence of one-to-one correspondence between English and Arabic and hence, committing such an error. The error in (2) is ascribed to L2-influence *per se* for the fact that the verb *yustii*(give) cannot be collocated with *su?aal* (question) in Arabic. One could argue here that the learner committing such an error is hypothesizing that as in using the verb *ask* in *I gave him what he asked*, he/she could use the verb *give* in such a context which is not possible in English. However, in (3), the error can be ascribed to *L1-transfer* on the basis that in Arabic, the equivalent to *well-known* *ma?ruuf/ma?shuur* has the same sense of *famous*. However, it could be ascribed to English for the fact that the learner committing such an error influenced by bilingual Arabic-English dictionaries hypothesizes that *well-known* can be used for *famous* and can be collocated with *castles* resulting in committing the error in the above sentence.

The error in (4), however, consists in using *lack* for *lose*. It cannot be ascribed either to Arabic or English. It cannot be ascribed to the former for the fact that the equivalent to such a sentence is not acceptable in Arabic. It cannot also be ascribed to the latter because again it is not acceptable in English. The error here lies in the difference between *lack* and *lose* where the former cannot be used in such a context, i.e. *lack* cannot be collocated with *unity*. It could be ascribed to guessing or misconception and hence ascribed to neither language.

5.1.2. Both Incorrect

As the name suggests, the error in this category involves two words collocated wrongly. This category includes errors ascribed to different sources. For instance, *L2-influence* involves 62 errors, i.e. (54%) which is followed by *L1-transfer* with 31 errors, i.e. (26%) followed by *mutual* with 20 errors, i.e. (17%) and the last one is *unrecognized* with 4 errors, i.e. (3%). This is exemplified in (5-8).

(5) *...and there are many pretty areas in Bada'n. (nice places)

(6) *.we established two homes; one in village and another in Ibb city. (built two houses)

(7) Ibb has beautiful persons. (kind people)

(8) *and I like him because he is an interested human. (kind person)

The ungrammaticality of (5) lies in the fact that *pretty areas* is not a possible collocation in English. It is ascribed to Arabic alone because the learner committing such an error, being influenced by bilingual dictionaries where the meaning of *pretty* is *jamiil*, and *areas*, *manaatiq*, which in turn is possible in Arabic, hypothesizes a one-to-one correspondence between both languages thinking that such a collocation is possible in English, as it is in Arabic, which is not and hence, resulting in a lexically deviant expression. The error in (6) is ascribed to *L2-influence* for the fact that the verb *?asasna* (established) and *baytain* (two homes) is not a possible collocation in Arabic. However, it could be ascribed to English because the learner committing such an error is misled by English meaning overlap where one of the meanings of *establish* is *build* and what he/she does here is just use one for the other and hence, rendering an ill-formed collocation.

The error in (7) can be ascribed to *L1-transfer* on the basis that in Arabic, the equivalent to *beautiful persons*, i.e. *?a?xaa?taibuun*, to a great extent has the same meaning of *kind people*, and hence, there is some kind of *false equivalence*. However, it could be ascribed to English for the fact that the learner committing such an error wrongly hypothesizes that *beautiful persons* can be used for *kind people* and hence, resulting in committing such an error. The error in (8) lies in substituting *interested human* for *kind person*. It has no source either in English or in Arabic. It cannot be ascribed to the former simply because such a collocation does not exist in Arabic. It cannot also be ascribed to English as such a collocation is not possible either. Had he/she been influenced by English, he/she would have used *interesting man/person*. Thus, here, there is some kind of guessing.

5.1.3. Contextually Odd

This category includes 109 errors, i.e. (15%) of the total errors committed which are ascribed to three sources. In this category, there are no errors attributed to an *unrecognized* source. For instance, *L2-influence* involves 68 errors, i.e. (63%) which is followed by *L1-transfer* with 31 errors, i.e. (30%) followed by *mutual* with 8 errors, i.e. (7%). As maintained by Shormani & Sohmani (2012), the errors in this type of collocation are not caused by something wrong either grammatically or semantically. However, the context is the only factor that renders such collocations lexically impossible. (9-11) illustrate the issue in question.

(9) *...because he has a big mind when he speaks. (reasonable thinking)

(10) *...and I love syntax because it has a good plan. (convincing/easy syllabus)

(11) *...but there are many voices in English. (many sounds)

What makes the collocation in (9) ill-formed is the context where it is used. The learner means here that the person he/she describes has *a reasonable thinking when he speaks*. In fact, this error is ascribed to Arabic simply because such an expression is possible and acceptable in Arabic. The error in (10) lies in using the collocation *good plan* in an incorrect context. The source of such an error is vividly *L2-influence* simply because the learner is hypothesizing that as he/she can use this collocation in a different situation as in *I have a good plan for tomorrow*, he/she can use it in the above context which renders the sentence ungrammatical. The error in (11) has *mutual* nature, i.e. it can be ascribed to Arabic and English both. It can be ascribed to the former simply because in Arabic, there is only one term used to describe both *sounds* and *voices* which is *?a?swaat* without any distinction either for sounds of speech or for voices of people. The learner might be misled by bilingual dictionaries. It could also be ascribed to English on the basis that the learner wrongly hypothesizes that *voice* and

sound can be used interchangeably without being aware of the difference between both expressions as to when and where to use either.

5.1.4. Incorrect Category

The errors in this category lie in the fact that the learner tries to collocate lexes of different grammatical classes where the “errors consist in using one or more collocates whose grammatical form/category is not correct” (Shormani & Sohmani, 2012, p. 134). This category includes 88 errors, i.e. (12%) of the total number of errors committed in this study. They are ascribed to three sources. Like the previous category, this category does not include errors attributed to an *unrecognized* source. For instance, *L2-influence* involves 54 errors, i.e. (63%) which is followed by *L1-transfer* with 21 errors, i.e. (24%) followed by *mutual* with 13 errors, i.e. (15%). Each source along with an example is provided in (12-14).

(12) *When my grandfather died, my father became the power man in the village. (powerful man)

(13) *...and it makes a big different in my life. (a big difference)

(14) *I am the responsibility man in my family because my father is dead. (the responsible person)

The error in (12) consists in collocating the nouns *power* and *man* which is not possible in English especially in this context. The *incorrect category* here lies in the fact that the first collocate has to be an adjective, i.e. *powerful* constituting *powerful man*. What is more is that at the NP level, there is some kind of oddity when saying *power man* and thus, the *incorrect category* here is the use of a noun in place of an adjective. This error is ascribed to *L1-transfer* on the basis that such a collocation, i.e. *rajul-u l-quwat-i* (the power man) in Arabic is lexically correct. What the learner does here is just transfer this very structure from Arabic into English and hence resulting in committing the error in question. In (13), the wrong category here is in the second collocate, namely *different* which has to be *difference*. Having not efficient knowledge to distinguish the adjective from a noun in English lexis, the learner fails to use the correct grammatical category and hence this error is ascribed to English simply because in Arabic such a structure is impossible. The error in (14) consists in using the noun *responsibility* in the place of *responsible*, i.e. an adjective. In fact, this error can be ascribed to L1, i.e. Arabic for the fact that the equivalent, i.e. *rajul-u l-mas'uliat-i* (the responsibility man) is possible and hence committed due to false equivalence. It could also be ascribed to English due to the lack of sufficient collocability competence the learner committing such an error suffers from.

5.2. Syntactic Collocations

Keshavarz & Salimi (2007) refer to syntactic collocations as grammatical collocations classifying them into *noun+ preposition*, *adjective+ preposition* and *verb+ preposition*. Consequently, I have attempted to classify the errors in *syntactic collocations* into *omission of a preposition* and *misuse of a preposition* as can be seen in Figure 1 above. Why this has been done is due to the fact that the *variable* is the preposition with respect to which the errors are committed while the first part, viz. a noun, adjective or verb remains a *constant*. This category involves 288 errors distributed as follows: *misuse of a preposition* includes 146 errors and *omission of a preposition* includes 142.

5.2.1. Misuse of a Preposition

This category includes 142 errors, i.e. (19%) of the total errors committed which are ascribed to four sources. For instance, *L2-influence* involves 74 errors, i.e. (52%) which is followed by *L1-transfer* with 46 errors, i.e. (32%) followed by *mutual* with 17 errors, i.e. (12%) and *unrecognized* comes last with 5 errors, i.e. (4%). Thus, the sentences in (15) through (19) exemplify the syntactic collocation errors in misusing a preposition.

(15) * I am not interested with syntax. (interested in)

(16) * Our country depends in Oil. (depends on)

(17) * our college consists from. (consist of)

(18) * ...but I am afraid at syntax exam.

In (15) through (18), the prepositions *with*, *in*, *from* and *at* have been used for *in*, *on*, *of* and *of*, respectively. In (15), the error can be attributed to L1, i.e. Arabic *per se* where *interested with* is a direct translation of the Arabic preposition phrase *muhtam-un bi*. However, (16) is ascribed to English where *depends in* has been used instead of *depends on* with which Arabic has nothing to do simply because *yastamid'ala* which is lexically grammatical in Arabic means exactly *depends on*. (17) could be ascribed to Arabic simply because *yatakawanmin* (consist from) is lexically possible. It could also be ascribed to English simply because both *of* and *from* have the same meaning in some contexts and hence, the learner seems to overgeneralize the use of *of* over *from*. (18) could not be ascribed to Arabic nor English. Had it been a transfer from Arabic, it would have been *afraid from*. It could not be ascribed to English simply because the learner is advanced, and it is impossible that he/she would not have come across *afraid of* being a very common expression in English.

5.2.2. Omission of a Preposition

146 errors, i.e. (20%) of the total errors committed occur in this category. As far as the source is concerned, these errors, like the above category, are ascribed to four sources. For instance, *L2-influence* involves 66 errors, i.e. (45%) which is followed by *L1-transfer* with 51 errors, i.e. (32%) followed by *mutual* with 21 errors, i.e. (14%) and *unrecognized* comes last with 8 errors, i.e. (6%). Such sources are exemplified in (19) through (22) as follows.

- (19)* *He went # his house.* (went to)⁶
(20)* *...and that was my answer your question.* (answer to)
(21)* *... everyone should wait his friend.* (wait for)
(22)* *I am proud my father.* (proud of)

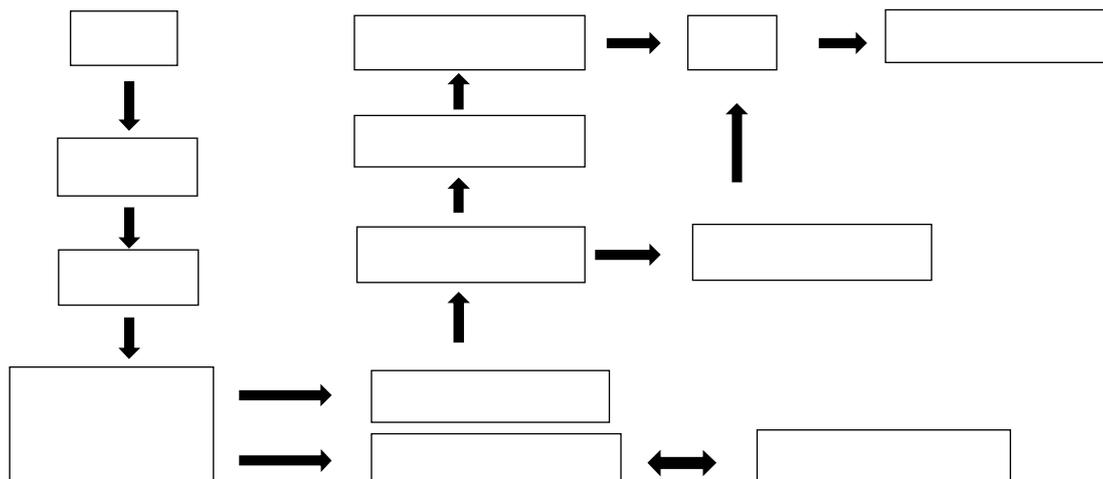
In (19) through (22), the omitted prepositions are *to, to, for* and *of*, respectively. The error in (19) is purely of *L1-transfer* nature. However, the variety of Arabic underlying such errors is not Modern Standard Arabic but rather Yemeni one where *raahbait-uh* (exactly: *went his house*) is the main cause. (20) is purely of *L2-influence* simply because had it been a transfer from Arabic, it would have been *my answer to* as it is not possible in Arabic to say *?ijaabat-isu?aal-ka* (*my answer your question*) but we can say *?ijaabat-u su?aali-ka* (*answer your question*) as a *Construct State* where *-i*(my) cannot accompany a *Construct State* (Fassi Fehri, 1993; Shormani, in press). What could be said regarding this very error is that the learner generalizes the use of *answer* as a verb over that of the noun where the preposition *to* is a must for the latter but not for the former. (21) is *mutual*, i.e. it can be ascribed to Arabic because the learner transfers the noun *?intidaar* meaning *waiting* and uses it for the verb and to English because the learner might have been influenced by the English *await* which does not require a preposition hypothesizing the correctness of such a use. (22) is *unrecognized* for it cannot be ascribed either to Arabic or English. It could not be ascribed to Arabic because had it been, it would have been *?anafaxu:r-un bi ?abi* (literally: I am proud *with* my father), it could not be ascribed to English simply because the learner is advanced and it cannot be assumed that he/she would not have been come across *proud of*.

6. UG-Model for Stable Acquisition: Proposal

In generative grammar, what collocability implies is that what goes with what is a matter of *selectional restrictions* as these restrictions are semantically based (Chomsky, 1965). As has been stated so far, knowledge of collocability is part of the native speaker's communicative competence simply because a particular "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" (Keshavarz & Salimi, 2007, p.83). If this is true, it follows that a native speaker of a language, say, English, does not have to learn English collocations simply because these restrictions come at no cost. Another very crucial issue is that syntax also plays a fundamental role in structuring these collocations coming up with what is so-called *subcategorization restrictions* though a minor role is attributed to such restrictions as far as lexical collocations are concerned, a major role is played by them in syntactic collocations, however (Magnúsdóttir, 1990). As far as this study is concerned, collocations are different from fixed expressions, viz. idiomatic expressions like idioms or proverbs. The difference lies in the fact that while the meaning of the former is compositional, that of the latter is not. If this is true, it follows that idiomatic expressions violate selectional restrictions but not subcategorization ones which is not the case with collocations for both restrictions. In other words, lexical collocations involve some kind of 'idiosyncratic tendencies' while syntactic ones are almost obligatory (Magnúsdóttir, 1990).

Now, if collocability is part of the native speaker's competence, i.e. he/she is equipped or genetically predisposed with such knowledge in terms of *Universal Grammar* (UG), and since the native speaker of English does not have to learn them, it follows that they constitute a considerable difficulty for learners of English and as far as this study is concerned, this seems to be true. In other words, looking at Table 1 above, this study vividly shows such a difficulty where 731 errors have been committed though the learners involved are advanced. Lexical errors score the first rank with 443 errors, i.e. (61%), Syntactic collocations occupy the second rank with 288 errors, i.e. (39%). This is a very serious issue and has to be rethought and reconsidered. In fact, this considerable issue puts us as linguists, applied linguists and language teachers vis-à-vis a complicated phenomenon that has to be sought appropriate solutions to. Consequently, I propose here a UG-Model based on *subcategorization* and *selectional restrictions* for acquiring English collocations as presented in Figure 2 below.

Figure 2: UG Model for Learning English Collocations



The proposed UG-Model shows how collocations are learned. The lexicon is part of UG consisting of lexes. After lexes come up from the lexicon, via a process called *Select*, they are merged via a syntactic operation called *Merge* (Chomsky, 1995). Lexes are **Lexical Arrays** numerated via the computation system (H_{CL}) via two syntactic operations (*Select* and *Merge*) (Chomsky, 1998). To put it simply, when lexes are selected and merged (i.e. when *collocating* takes place), the result is either lexically or syntactically correct, and here learning commences where cognitive strategies come to play. If the learner employs the rightly required strategy, the process continues to **subcategorization restrictions**. Subcategorization restrictions represent the underlying abilities of a lexical category to select (subcategorize for) another lexis. For instance, a transitive verb like *write* subcategorizes for another lexis as an object. Here, if the collocation is syntactic and if it passes such restrictions, the output is a syntactically correct collocation. If, however, it is a lexical collocation, it has to continue to pass the **selectional restrictions** and if it does, the output is a lexically correct collocation which meets the UG requirements and the final stage is a correct collocation. The same thing also applies to syntactic collocations. Now, if the strategies made use of by the learner are not the right ones, the resultant collocation(s) will be incorrect. Such a resultant collocation will have to return to cognitive strategies and here lies the core of this model. In other words, the role of learning lies in correcting the incorrect strategies made use of by (or introduced via teaching to) the learner, showing him/her that he/she should do such and such to avoid committing errors. For instance, learners could be made clear that one-to-one correspondence between English and Arabic, not only in collocations but also in many other linguistic aspects, is almost misleading and cannot be considered an appropriate strategy to learn English collocations either lexical or syntactic for the fact that divergence is more probably to occur than convergence in the issue in question.

As stated above, all the collocation errors committed by the subjects of this study are ascribed to different sources which are in turn due to employing incorrect cognitive strategies. As far as **subcategorization restrictions** are concerned, syntax classes could be utilized. For instance and as far as verbs are concerned, there is a class of verbs subcategorizing for no constituent because subcategorizing for a constituent is beyond their underlying ability. If this is true, it follows that they are listed in our lexicon as entries without being able to have any object(s) as in **Ali died a dog* where the NP *a dog* has to be removed. There are also verbs subcategorizing for a PP as in **She relies her father* where the verb *rely* has to have the preposition *on* as a syntactic collocate. Some verbs subcategorize for two NPs like *give* and so on. More importantly, there are also verbs with multiple subcategorizations like *grow* in *Ali is growing (this year)*, *Ali is growing carrots (this year)* and *Ali is growing his field Indian carrots (this year)* subcategorizing for no, one and two constituents, respectively. **Subcategorization frames** for each class of verbs could be presented in syntax classes (Shormani, 2013b)

Regarding other categories like adjectives and nouns, in vocabulary and grammar classes, learners could be given lists of the common adjectives and nouns along with the prepositions they are only to collocate with. Writing classes could also be utilized for such a task. In these classes, learners should be given drills where such collocates are recognized and emphasized in their writings. For instance, learners could be given writing assignments and asked to write paragraphs and essays where collocations are paid much attention to. Critical reading of novels, poems and plays is emphasized, too.

Unlike subcategorization restrictions, selectional ones are concerned with the type of constituents a word to collocate with. Students' attention should be drawn to the fact that violating selectional restrictions goes against our senses and our world and encyclopedic knowledge. Selectional restrictions seem to be more concerned with lexical collocations and hence, semantics classes should be utilized for learning such collocations. For instance, learners could be introduced to the nature of selectional restrictions in terms of (-/+) features of collocates such

as (-/+ human), (-/+ animate), (-/+ abstract), (-/+ edible), etc. For instance, the verb *eat* has to have (+ edible) object and (+ animate) subject. In fact, this is encoded in the lexis as an underlying ability of such a verb. Such restrictions are also somewhat implied in such collocations as *pay attention* but not *give attention* and *to save time* but *not to keep time*, *to make progress* but not *to do progress* and so on.

Since L1, i.e. Arabic has a considerable portion in committing collocation errors, learners' attention should be drawn and directed to the crucial differences between English and Arabic and that a one-to-one correspondence or equivalence, for instance, between both languages is almost misleading and cannot be considered an adequate strategy to learn English collocations either lexical or syntactic and that lexical divergence is more expected than lexical convergence (Shormani, 2014). This also leads to a very crucial conclusion being that the use of Arabic in classes should be avoided or at least reduced to a minimal level.

The fact that the learner using such cognitive strategies in his/her learning indicates that he/she is an active interlocutor/participant in the learning process but unfortunately such a use is not often error-free. However, the learner should be made clear that he/she could make use of such cognitive strategies but very cautiously. Bilingual-dictionary influence has been identified as a cognitive strategy learners unconsciously make use of and hence, such dictionaries should be avoided and instead, learners could make use of monolingual, viz. English-English dictionaries especially in the case of advanced learners as it is the case of the subjects of this study.

8. Conclusion

Thus, this study has been concerned with collocability difficulties Arabic-speaking advanced learners of English are facing in their learning of collocations both lexical and syntactic. Different types of errors have been identified and so have sources. Different cognitive strategies such as *false conceptions*, *wrong hypothesizing*, *L1 interference*, *hypothesizing a one-to-one correspondence between English and Arabic*, *internalizing English linguistic system* among others have also been identified. Based on these results and findings, a UG-Model for learning such pieces of language based on *subcategorization restrictions* and *selectional restrictions* has been proposed. In fact, the model proposed here could be developed in a more detailed way to include all types of learning. I hope it could be of some value for applied linguists, SLA scholars, syllabus designers, teachers and learners alike if utilized for understanding the nature of collocations and how they work at both levels, viz. syntax and lexicon and how they could be learned. Thus, though the subjects of this study are Arabic-speaking learners of English, its results and findings could be generalized and applied in similar ES/EL contexts.

References

- Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21 (1), 104-114.
- Bahns, J. (1993). Lexical collocations: a contrastive view. *ELT Journal*, 47 (1), 56-63.
- Channell, J. (1981). Applying Semantic Theory to Vocabulary Teaching. *ELT*, 35(2).115-122.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1968). *Language and Mind*. New York: Harcourt Brace Jovanovich.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. (1998). Derivation by phase. In *MIT Occasional Papers in Linguistics*. Cambridge: The MIT Press.
- Cook, V. (2003). (ed.). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Corder, S. (1973). *Introducing Applied Linguistics*. Harmondsworth: Penguin Books Ltd.
- Daneman, M. (1991). Individual Differences in Reading Skills. In *Handbook of Reading Research*, (eds) Barr, R., Kamil, M., Mosenthal, P. and Pearson, P. White Plains: Longman.
- Duskova, L. (1979). On Sources of Errors in Foreign Language Learning. *IRAL*, 7(1), 11-36.
- Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University press.
- Farghal, M. and Obidat, H. (1995). Collocations: A Neglected Variable in EFL. *IRAL*, 33, 315-331.
- Fassi-Fehri, A. (1993). *Issues in the Structure of Arabic Clauses and Words*. Dordrecht: Kluwer Academic Publishers.
- Gass, S. and Selinker, L. (2008). *Second Language Acquisition: An Introductory Course*. (3rd ed.). New York: Routledge.
- Gitsaki, C. (1997). *Second lexical language acquisition: A study of development of Collocation knowledge*. Maryland: International Scholars Publications.
- Griffiths, P. (2006). *An Introduction to English Semantics and Pragmatics*. Edinburgh: Edinburgh University Press.
- Han, Z.-H. (2000). Persistence of the Implicit Influence of NL: The Case of the Pseudo-Passive. *Applied Linguistics*, 21(1), 55-82.
- Han, Z.-H. (2004). *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters.
- Hang, J. (2005). A Study of Lexical Errors in Cantonese ESL Students' Writing. University of Hong Kong MA Thesis.
- Hemchua, S. and Schmitt, N. (2006). An Analysis of Lexical Errors in the English Compositions of Thai Learners. *Prospect*, 21, 3, 3-25.

- Hennings, D. (2000). Contextually Relevant Word Study. *Journal of Adolescent and Adult Literacy*, 44, 268-279.
- Hill, J. (1999). Collocational Competence. *English Teaching Professional*, 11, 3-6.
- James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis*. Hong Kong: Addison Wesley Longman Limited.
- Keshavarz, M., and Salimi, H. (2007). Collocation Competence and Cloze Test Performance: A Study of Iranian EFL Learners. *IRAL*, 17(1), 81-92.
- Khalil, A. (1985). Communicative Error Evaluation: Native Speakers' Evaluation and Interpretation of Written Errors of Arab EFL Learners. *TESOL Quarterly*, 19(2), 335-351.
- Laufer, B. (1986). Possible Changes in Attitude towards Vocabulary Acquisition Research. *IRAL*, 24(1), 69-75.
- 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. (1997). Some Properties of the Foreign Language Learner's Lexicon as Evidenced by Lexical Confusions. *IRAL*, 29(4), 317-330.
- Lennon, P. (1991). Error: Some Problems of Definition, Identification and Distinction. *Applied Linguistics*, 12(2), 180-195.
- Lennon, P. (1992). Error and the Very Advanced Learner. *IRAL*, 29, 1, 31-44.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Hove: Language Teaching Publications.
- Li, Chia-Chuan. (2005). A Study of Collocation Errors Types in ESL /EFL College Learners' Writing. Unpublished Master thesis. Ming Chuan University.
- Lien, H. Y. (2003). The effects of collocation instruction on the reading comprehension of Taiwanese college students. Indiana University of Pennsylvania Ph. D Dissertation,
- 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.
- Magnúsdóttir, G. (1990). Collocations in Knowledge Based Machine Translation. In Pind, J. and Rögnvaldsson, E. (eds.), *Papers from the Seventh Scandinavian Conference of Computational Linguistics*, Reykjavik, 204-207.
- Mahmoud, A. (2005). Collocation errors made by Arab learners of English. *Asian EFL Journal*. Retrieved February, 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.
- Martynska, M. (2004). Do English Language Learners Know Collocations? *InvestigationesLinguisticae*, 11, 1-12.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- Mukattash, L. (1976). Errors Made by Arab University Students in the Use of English Prepositions. *Glottodidactica, International Journal of Applied Linguistics*, 17, 47-64.
- Mukattash, L. (1981). Wh-Questions in English: A problem for Arab students. *IRAL*, 19(4), 317-332.
- Naba'h, A. (2011). Lexical Errors Made by In- Service English language Teachers in Jordan. *Damascus University Journal*, 27(1+2), 49, 75.
- O'Dell, F., and McCarthy, M. (2008). *English Collocations in Use. Advanced*. Cambridge: Cambridge University Press.
- Oxford, R. (1990). *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House Publishers.
- Rababah, G. (2003). Communication Problems Facing Arab Learners of English: A Personal Perspective. *Journal of Language and Learning*, 3(1), 180-197.
- Richards, J. (1974). *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman.
- Ridha, A. and Al-Riyahi, N. (2011). Lexical Collocation Errors in the Writings of Iraqi EFL Learners. *Journal of the College of Arts, University of Basrah*, 58, 24-51.
- Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. New York: Cambridge University press.
- Schmitt, N. and Zimmerman, C. (2002). Derivative Word Forms: What Do Learners Know? *TESOL Quarterly*, 36(2), 145-171.
- Shormani, M. (2012a). *L2 Acquisition in the Arab World: An Empirical Study of Yemeni Learners' English Syntactic and Semantic Error Gravity*. Saarbrücken: LAMBERT Academic Publishing, Germany.
- Shormani, M. (2012b). Sources of Syntactic Errors in Yemeni Learners' English Compositions: A Psycholinguistic Analysis. *AWEJ* 3(4), 275 -296.
- Shormani, M. (2013a). Fossilization and Plateau Effect in Second Language Acquisition. *Language in India*, 13, 763- 784.
- Shormani, M. (2013b). *An Introduction to English Syntax: A Generative Approach*. Saarbrücken: LAMBERT Academic Publishing, Germany.
- Shormani, M. (2014). Lexical Choice Difficulties: A Psycholinguistic Study towards a Solution. *IISTE*, 3, 66-76.
- Shormani, M. (in press). Free Genitives in Yemeni Arabic: A Minimalist N-to-Spec Approach. *Syntax*.
- Shormani, M. & Sohmani, Y. (2012). Semantic Errors Committed by Yemeni University Learners: Classifications

- and Sources. *International Journal of English Linguistics*, 2(6), 120-139.
- Sonaiya, R. (1991). Vocabulary Acquisition as a Process of Continuous Lexical Discrimination. *IRAL*, 14(4), 269-284.
- Takac, V. (2008). *Vocabulary Learning Strategies and Foreign Language Acquisition*, Multilingual Matters Ltd.
- Tang, B. (2004). Collocation Errors in the Intermediate EFL Learners' Interlanguage. *Sino- US English Teaching*, 1(12), 39- 47.
- Taylor, C. (1976). Sources of Error in Foreign Language Teaching. *ELT*, 30(3), 190-195.
- Taylor, G. (1986). Errors and Explanations. *Applied Linguistics*, 7(2), 144-166.
- Tomasello, M. (2007). Cognitive Linguistics and First Language Acquisition. In Geeraerts, D. and Cuyckens, H. (eds.), *Handbook of Cognitive Linguistics*, Oxford: Oxford University Press, 1092-1112.
- Wang, W. and Wen, Q. (2002). L1 Use in the L2 Composing Process: An Exploratory Study of 16 Chinese EFL Writers. *Journal of Second Language Writing*, 11(3), 225-246.
- Wray, A. (2000). Formulaic Sequences in Second Language Teaching: Principles and Practice. *Applied Linguistics*, 21(4), 463-489.
- Yuan, H. and Lin, C. (2001). Collocations: A case study of Ming Chuan University students. *The proceedings of Eighteenth Conference on English Teaching and Learning in the Republic of China*, 407-422.
- Zimmermann, R. (1986). Semantics and Lexical Error Analysis. *EnglischAmerikanische Studien*, 2(86), 294-305.
- Zughoul, M. (1991). Lexical Choice: Towards Writing Problematic Word Lists. *IRAL*, 29(1), 45-60.
- Zughoul, M., and Abdul-Fattah, H. (2003). Collocation Strategies of Arab Learners of English: A Study in Lexical Semantics. (ERIC Document No. ED 479746).

¹What is meant here by the term “piece of language” is a phrase and since the term *collocation* implies a phrase, I use it here and elsewhere in this paper to mean just a collocation without any theoretical implications.

²This classification has been adopted and adapted from Shormani (2012b, 2014).

³ Their main focus was how to classify semantic errors alluding to collocation ones.

⁴The Asterisk * stands for the ungrammaticality of the word/phrase/sentence it is put before.

⁵The use of the adjective *kind* but in a different context has been also identified by (Zughoul, 1991).

⁶# stands for the omitted preposition.