



# A Minimalist Approach to Collocation Acquisition: Integrating Syntax and Semantics

Mohammed Q. Shormani

Department of English Studies, College of Arts, Ibb University, Yemen

Received 09 July 2014, Revised 04 August 2014, Accepted 08 September 2014, Published 1 January 2015

**Abstract:** How language, be it first (L1) or second (L2), is acquired has been the concern of much research and investigation, and perhaps no other phenomenon has witnessed such interest (Shormani, 2014b). One such area within language acquisition is collocability and how it presents a difficult aspect to L2 acquirers. This is so due to the fact that collocability involves formulaic language the mastering of which has been considered specific to native speakers of the language being learned. Thus, in this article, I propose a minimalist approach based on integrating syntax and semantics. The former concerns combining (collocating) two lexical items by means of *Select* and *Merge* operations, and the latter concerns what goes with what in a collocation based on feature specifications encoded on each lexis (i.e. word). My proposal is based on substantial evidence proving the availability of syntax and semantics in collocability, hence, abstracting from usage-based approaches. Thus, each collocation produced by *Select* and *Merge* (syntax) has to “pass” the semantic constraints manifested in the Collocating Feature Specification Rule (=CFSR). However, if this produced collocation fails to “pass” CFSR, it has to undergo acquisition once more in which parameters are reset and retrigged through *acquisition reorientation*. The proposal places much emphasis on mental properties of Universal Grammar (UG), and the same is true concerning L2 acquisition settings in relation to providing L2 acquirers with “equal” linguistic input native speakers have. The minimalist approach developed in this article stems from the fact that minimalism is seen as a theory, *primarily* concerned with language acquisition in its two spheres (i.e. L1 and L2). As far as L1 acquisition is concerned, minimalism sees it as acquiring only *feature specifications* manifested in parameters. L2 acquisition, however, is seen by minimalism, as merely acquiring *features*, peculiar to L2 being learned, which are different from and/or similar to those of the acquirer’s L1, where the learner’s only task is to reset and retrigger the UG parameters specific to L2, simply because UG principles are universal. Hence, our major task as linguists, language educators and teachers alike is how to make use of and benefit from the minimalist assumptions and hypotheses concerned with language acquisition in general, and those concerned with collocability in particular.

**Keywords:** Minimalism, Select, Merge, Feature Specifications, Collocation Acquisition, Language Educators

## 1. INTRODUCTION

It is a common belief of L2 language acquisition that the most difficult area is how to master or be aware of which word goes with which in *a phrase* (i.e. collocation). Most of the time, our teachers were telling us when we produced a wrong collocated piece of language “no, we cannot say this, because so and so, you have to say this because so and so.” For instance, if we were addressing our teachers saying “excuse me, doctor?, for example, my Ph.D mentor used to tell me “no, we cannot address our teachers saying: ‘Doctor!’,’ because the term “doctor” is used with someone specialist in medicine and works at hospital. To address our university teacher, we say: “excuse me, Sir”, excuse me, Professor!” because the word “Sir,” for instance, shows some kind of respect, and the word “professor” shows or refers to their profession, i.e. specifying Ph.D holders who teach at university.” Another example concerns any other collocation, say, for

instance, “\**kind weather*.” The teacher hearing such an utterance (i.e. a collocation) will say: “no, the adjective *kind* cannot co-occur with the noun *weather*, the word *weather* can co-occur (collocate) with adjectives like *nice* as in “*nice weather*.” However, a linguist, unlike the teacher, would say that the word *Sir*, for instance, is restricted to be used in such context like addressing a teacher, professor, boss, etc. just to show some sort of respect, while the word *doctor* is used to address those who work at hospitals. Regarding the use of *kind* with *weather*, a linguist would argue that the word *weather* imposes some sort of constraints on the word (i.e. adjective, for instance) to collocate with. Now, the question is why is it that such restrictions are there? To put it simply, why can’t we say *kind weather*, but *nice weather*, *pay attention* but not *give attention*? In fact, the answer to this among other related questions is the major tenet of this article.



From a semantic point of view, some linguists (e.g. Martynska, 2004; Tang, 2004) argue that meaning of a word be determined most by the habitual company this word occurs with, hence, contextualization plays a crucial role in what a word means. What contextualization simply implies is collocability (the underlying ability of a particular word to co-occur with other(s)). There are also those (e.g. Shormani, 2012a; Han, 2004; Valette, 1991) who see lexical errors in general and collocational in particular as obstacles behind L2 learners stopping-short (fossilized) of a native-like proficiency, and hence, unintelligibility will be a by-product of such communication impediment (e.g. Shormani, 2013a; Tang, 2004; Shormani, 2012a&b; Khalil, 1985; Hang, 2005; Zughoul & Abdul-Fattah, 2003; Valette, 1991, among others). Along similar lines, Lewis (1997, p. 15) states that “fluency is based on the acquisition of a large store of fixed or semi-fixed prefabricated items that are available as the foundation for any linguistic novelty or creativity.”

Now, the question is why is it that such a difficulty accompanies collocation acquisition? Answering this question has been formalized by Shormani (2014c). He maintains that such a difficulty is due to there being no rule, logic and/or even a particular way/technique to follow in understanding and mastering collocations and their use. The nature of collocability involves some sort of arbitrariness and idiosyncrasies. Along similar lines, Takac (2008, p. 10) writes that collocations and collocability represent “knowledge of conceptual foundations that determine the position of the lexical item in our conceptual system” This, more or less, lends us support to assume that collocability is part of the native speaker’s competence, creativity and intuition (Cowie, 1998; Wray, 2000), and hence, mastering collocability is specific to such native speakers (of a language) alone. If this is true, it certainly follows that it has only to do with innate properties, viz. UG, that determines language acquisition in general and acquiring collocations in particular (see e.g. Shormani, 2014a&b; Kreidler, 2002; Griffiths, 2006; Keshavarz & Salimi, 2007; Martynska, 2004; Tang, 2004; Mahmoud, 2005; Shormani, 2012a, 2014c; Shormani & Sohmani, 2012). Thus, it is our duty to make our students able to produce and use collocations because, I assume, mastering such formulaic language enables them to effectively incorporate new concepts into existing conceptual schemata, and hence, producing pieces of language suitable for a particular context (Shormani, 2014c).

Thus, as far as collocations of English in the context of L2 acquisition are concerned, there are numerous studies that have tackled collocations or word collocates applying several approaches like corpus linguistics (see

Seretan, 2013), lexical approach (see Firth, 1957), the meaning-text theory (see Mel’čuk, 1998), cohesion theory (see Halliday & Hasan, 1976), grammar-lexis theory (see Hunston & Francis, 2000), fluency and accuracy (see e.g. Lewis, 1997). However, collocations have been almost neglected in syntactic theory, specifically generative grammar. To my best knowledge, there is a study done by (Rögnvaldsson, 1993) in which he investigates collocations and their formation in Old Icelandic in generative syntax. Collocations as maintained by Chomsky (1965) are semantically based, i.e. the semantic factor is manifested in *selectional restrictions*. According to Magnúsdóttir (1990), neither syntax nor semantics *per se* could handle collocations and their formation since the meaning of a collocation is compositional and idiosyncratic, and involve arbitrariness, and in this regard, Magnúsdóttir (1990, p. 204) argues that collocations are best defined in terms of being pieces of language containing lexes which collocate under specific constraints “not definable by syntax nor selectional restrictions [i.e. semantics] alone.” Such constraints are best “referred to as lexical restrictions since the selection of the lexical unit is not conceptual, thus synonyms cannot replace the collocate.” Thus, I claim following Rögnvaldsson (1993), that collocations and collocation acquisition be best handled in the generative grammar, specifically, *Minimalism*. Minimalism, in fact, provides a straightforward account of how collocations work and how their acquisition is accounted for (I return to this in section 5).

## 2. USAGE-BASED APPROACHES

As has been stated above, collocations have been tackled in different approaches, but there has been no reliable theory that has tackled them appropriately, and no theory has related them to language acquisition. Thus, as has been noted by Gitsaki (1999), there are three main approaches that have tackled collocations from a usage-based perspective: lexical, semantic and structural.

### 2.1. The Lexical Approach

The lexical approach to the study of collocation is based on the assumption that a word’s meaning is identified according to the habitual company such a word occurs with. In fact, collocations in this approach are seen as a lexical phenomenon which is independent of grammar. Firth (1957) views a collocation as a “mode of meaning [in which] the lexical meaning of any given word is achieved by multiple statements of meaning at different levels.” These levels are the orthographic level, phonological level, grammatical level, and collocational level (Firth, 1957, p. 192). Take the word *peer* from Firth as an example to illustrate how lexical approach deals with collocation. He describes the behavior of the word *peer* at the orthographic level by considering the letters



which constitute it making it different from those of *pier*, though they sound the same at the pronunciation level. These two levels are then followed by the grammatical level. In the latter, the word *peer* as a noun is different from that when it is a verb. In addition to these three levels, formal and etymological meaning could also be made use of, and even social aspects (Firth, 1957, p. 192). He also exemplifies the word *peer* in relation to the collocational level. One of the meanings of the word *peer*, for instance, is its collocation with the *school* as in *school peers*.

In addition, to Firth, there is a “general rule” pinpointing that contextualization plays a role in determining the nature of a word, i.e. two occurrences of a word depend on the context. Every word has a different identity depending on the context it occurs in. For instance, the word *bank* has two different identities: one in *I have recently opened an account in the bank* and another in *we used to spend our happiest moments at the bank*. In the former, the word *bank* means a financial institution, and in the latter, it means the edge of a river. According to Firth, there is a difference between contextual meaning and collocation meaning. For instance, the word *hour* in *I see you in an hour* has a meaning different from that in *I cannot go now because it is a rush hour*. The former is contextual and the latter is collocational. That is, in the former the word *hour* has an ordinary meaning, i.e. related to time while in the latter, the whole collocation *a rush hour* means *crowdedness*. He also classifies collocations into “general or usual collocations and more restricted technical or personal collocations” (Firth, 1957, p. 195). However, Firth’s theory of lexical approach has been modified by some linguists (e.g. Halliday, 1966; Sinclair, 1966) who pay more attention to the importance of lexical collocations, i.e. collocations that consist of lexical items, in an integrated lexical theory. In fact, though grammar is integrated in Halliday’s extension of Firth’s lexical approach, it is not as could be expected (Halliday, 1966, p.148). In this regard, Halliday proposes that grammar could be considered a tool that makes language an organized system of choices, if there are items that fail to “resolve themselves into systems,” (Sinclair, 1966, p. 411). This theory enables linguists to divide words into sets based on their collocational environment and similar collocation restrictions. The words *bright*, *shine* and *light* are, for instance, members of the same lexical set, because they are frequent collocates of the word *moon* (Halliday, 1966, p.156). The very slight notion of grammar made use of by Halliday is manifested in the notions syntagmatic and paradigmatic relations. In the former, he distinguishes between the words in relation to a specific lexical item (i.e. its collocation with a specific word). For instance, lexical items like *strong* and *powerful* are considered members

of the same lexical set because they collocate with lexical items like *argument* as in *strong argument* and *powerful argument* and *strong tea* but not *powerful tea*.

However, one major problem encountered by this approach is “the circularity of the definition of the basic unit of description, the lexical item” (Sinclair, 1966, p. 412). That is, every lexical item is described in terms of its environment, i.e. the context in which it is used. For instance, one of the meanings of *night* is its collocability (i.e. ability to collocate) with *dark*, and vice versa (Firth 1957, p. 196). In addition, the lexical approach to collocation has been found (see Lyons, 1966) to be basically based on the assumption that the meaning of a word is a “complex of contextual relations” which is puzzling. Indeed, Lyons criticizes the apparent lack of principles by means of which “lexical groups by association” can be established and “lexical sets” can be defined (Lyons, 1966, p. 287ff). In fact, criticizing the lexical approach, Lyons proposes abandoning of Firth’s theory altogether.

## 2.2. The Semantic Approach

After pinpointing the shortcomings of Firth’s approach to the study of collocations, there have been several attempts to handle such an area. **The Semantic Approach** is one of these attempts. This approach to the study of collocation, as a linguistic phenomenon, is based on the idea of rejecting the equation of “one word, one meaning” paying much attention to the semantics of language, i.e. “word meanings do not exist in isolation, and they may differ according to the collocation in which they are used” (Robins, 1967, p. 21). In fact, the semantic approach investigates collocations from a semantic perspective. It seems that the underlying idea of this approach is based on Chomsky’s (1965) subcategorization restrictions which are strict rules, i.e. rules that “analyze a symbol in terms of its categorial context” and selectional restrictions which are also strict rules, i.e. rules which “analyze a symbol in terms of syntactic features of the frames in which it appears” (Chomsky, 1965, p.95). These rules assist the generation of grammatical strings. The failure of strict subcategorization rules will, for instance, result in such strings as *\*Ali liked* and *\*Ali seems Alia to go* while the failure of selectional rules results in utterances like Chomsky’s famous phrase *Colorless green ideas sleep furiously* (Chomsky, 1965, p. 149).

What is crucial to this approach is the assumption that the meaning of a word is a sum of the semantic properties of such a word, i.e. the semantic features of a word are what determines the other words it can occur with, though there was some attention attributed to grammar in this approach as semantics *per se* cannot be relied on for the



study of collocations. However, the portion attributed to grammar was so less to be considered at all. In addition, this approach fails to account for collocation restrictions and presuppositions of a given lexical item. For instance, there is a difference in the ability of the word *customers* and *clients*. Regarding the former, we say that *customers* buy something with money while *clients* get less tangible professional or technical service. Accordingly, *customers* can be used in relation to *bakers*, *butchers*, *grocers*, etc. while the word *clients* can be used in relation to *solicitors*, *architects*, and based on this, the word *customers* seems to be more applicable to *banks* rather than *clients* (Cruse, 1986, p. 281). In fact, this was one of the main points raised against the semantic approach for the fact that there being several idiosyncratic co-occurrences or combinations that this approach fails to account for. Such idiosyncratic co-occurrences cause severe problems for the study of collocations.

### 2.3. The Structural Approach

Thus, the problems and the shortcomings raised against the above two approaches have led linguists to seek another approach to the study of collocations, an approach that can adequately account for the nature of collocations, describing them in a proper way. As has been stated earlier, the lexical and semantic approaches have almost discarded the role of grammar (i.e. syntax) from the study of collocations. The structural approach, however, incorporates the syntax, however alone, i.e. it discards semantics. Thus, it has been proposed that an adequate approach to the study of collocation should incorporate syntax because, as has been seen regarding the two approaches discussed above, grammar cannot be dispensed with. For instance, Mitchell (1971) proposes that collocations should be studied within grammatical matrices. Mitchell bases his ideas on the fact that word forms like *writes*, *writer* and *writing* are forms of the same lexeme. Thus, instead, he prefers such forms to be abstracted and maintain only the root, *write*. In addition, collocations like *heavy drinker* and *drinks heavily*, Mitchell maintains, should be dealt with similarly for the fact that both give a similar (though not identical) meaning. This could be based on their roots, viz. *drink-* and *heav-* for both the above collocations can be accounted in such a way that while the former can be used as a noun when the suffix *-er* is attached to it or a verb when the suffix *-s* is attached to it. The same thing can be said about the root *heav-* having *-y* and *-ily* when it is used as an adjective and an adverb, respectively, in relation to the root *drink-*.

One of the most striking points of this approach is that it attributes something (even very much less than expected) to language learning by L2 learners by introducing the notion of language blocks and lexicalized sentence stems.

For instance, Pawley and Syder (1983) have introduced this notion suggesting that in addition to grammatical rules (subcategorization rules), second language learners need to “learn a means for knowing which of the well-formed sentences are native-like-- a way of distinguishing those sentences that are normal or unmarked from those that are unnatural or highly marked” (Pawley & Syder, 1983, p. 194). However, instead of making use of selectional restrictions, they propose that L2 learners have to learn language in blocks, and a big portion of a native speaker’s lexicon consists of lexicalized stems of sentences.

Thus, to me, as it stands, this approach is syntactic in nature, i.e. as its name suggests, structure is paid much attention to in studying collocation. In fact, this approach has been based on Chomsky’s ideas of syntax, specifically, subcategorization rules. However, it ignores selectional rules. In other words, as the semantic approach, discussed above, discards subcategorization rules, the structural approach discards selection restriction rules. Thus, it seems that this does not always hold true. In that, there are certain roots that cannot behave similarly as in *drink-* and *heav-*. For instance, the roots *faint-* and *praise-* can collocate resulting in *faint praise* when the former is used as an adjective and the latter as a noun. However, this seems not to hold true of the other choice, i.e. while the former collocation is acceptable, collocations like *praised faintly* are not grammatical.

Thus, after this sketchy discussion of the well-known usage-based approaches which have tackled collocations, it seems that they all seem inadequate. *Inadequate* in the sense that they each either make use of meaning alone like the lexical and semantic approaches or structure alone as in the case of the structural approach. In addition, they hardly pay attention to language acquisition. Hence, the study of collocation has been abstracted from the relation to SLA, though very slightly mentioned by the structure approach. Even this slight portion was based on memorization of lexicalized sentence stems in relation to native speakers. In fact, the structural approach in general was based on behaviorism which views language learning as *habit formation* got by memorization of a large corpora of stimuli (for a comprehensive critique of behaviorism as a theory of language acquisition, see Shormani, 2014b). Thus, we need an approach that integrates syntax and semantics manifested in subcategorization restrictions and selection restrictions, where the latter are manifested in feature specifications encoded on lexical items. In other words, we need an approach that tackles collocation in such a way that the co-occurrence of a lexical item with another is controlled by syntax and restricted by this item’s



underlying selectional restrictions, i.e. the tendency for a lexical item to co-occur with another based on their mutual subcategorization and feature specification. We need an approach to the study of collocation that attributes not only “slight” portion to SLA but an equal one to both L1 and L2 acquisition. In fact, this issue has been questioned by several linguists, though having different views regarding such a portion. For instance Wray (2000) has argued that collocational competence is closely tied to the native speakers as being an integral part of the native language (Fan, 2009). In addition, Cook (1999) argues that since native speakers’ language is the norm, it is difficult for L2 learners to reach the former’s competence whatever language is sought for, i.e. either collocation or any other aspect. Further Fan proposes that L2 collocation awareness is very much related to that of their L1 and suggests that teachers should raise their students’ awareness of collocations of their L1 before raising their awareness to L2’s collocations.

Thus, in this article, I diverge from usage-based approaches to the study of collocations for the fact that they each fail to adequately account for the nature of collocability in general and collocations in particular. One more essential aspect is that such approaches leave open a fundamental aspect related to the study of collocations. They each tackle collocations from a native speaker’s side, and hence, leaving the issue of how they behave from an L2 acquisition perspective. It is also true that their theoretical bases are hazy still, they are all based on behaviorism manifested in the way they stress memorization of lexicalized stems of sentences, and it has been proved true that behaviorism is not a reliable theory, at least alone (see Shormani, 2014b, for how behaviorism alone is not a reliable theory for language acquisition).

### 3. Syntax

The *Principles and Parameters* (=P&P) approach is a generative approach based on the biolinguistic ontology to the study of language in general and language acquisition in particular. The main concern of P&P has been addressing the questions in (a-e) below.

- a. What constitutes knowledge of language?
- b. How does this knowledge evolve?
- c. How is this knowledge acquired?
- d. What are the relevant brain mechanisms?
- e. How is this knowledge put to use?

In fact, all these questions and answering them revolutionize the study of language and how language is acquired and put to use, framed in the biolinguistic discipline. Question (a), for instance, represents what has been known as *Humboldt’s problem*. As has been

noted by Hornstein (2009), Humboldt was confined by what constitutes *knowledge of language*. In fact, this has been dealt with from a biolinguistic perspective. The core of (a) concerns the knowledge or the faculty of language (=FL) as noted in (Stroik & Putnam, 2013). Question (b) is referred to as *Darwin’s Problem*. This question is concerned with the basic notion of how language evolves in human beings alone (i.e. species-specific). Regarding (1c), the question addresses what has been known as *Plato’s Problem*. Shormani (2012, p. 54) argues that the latter refers to the difference between what a child knows and his/her lack of experience as input. In that, there are too complex linguistic structures that cannot be learned so quickly from the environment. To explain such a phenomenon, Chomsky (1987) holds that children have an innate faculty in their brain which is responsible for and guides them to master these complex linguistic rules in an early age. Question (d) has been referred to as *Broca’s Problem*, concerning the structure of the FL from a Neurolinguistic perspective. In other words, what is there in the brain that makes humans produce and perceive language. In fact, Broca studies the language disorder found in some children, and he has concluded that there is an area in the brain located in the frontal lobe of the left hemisphere, which is responsible for speech production and that if it is damaged, it causes a language disorder. This area has been given his name, i.e. Broca. Regarding (e), it has been referred as *Descartes’ problem*. The latter tackles language from a logical perspective. In other words, it is concerned with how we understand a piece of language differently in different contexts.

According to P&P, UG has four major modules, phonology, morphology, syntax and semantics, in addition to lexicon. However, the major concern of P&P is syntax and how it interacts with lexicon by means of the computational system ( $C_{HL}$ ). Chomsky (1981, p. 5) points out that the computational system has four major components: i) lexicon, ii) Deep Structure (=DS), i.e. a level at which categorial features (i.e. some properties of UG like Case and theta roles are assigned), Surface Structure (=SS), i.e. the transformational subcomponent, a level at which surface positions of lexical items are obtained by means of movements, iii) *Phonological Form* (=PF), i.e. a level where pronunciation is processed, and iv) *Logical Form* (=LF), i.e. a level where meaning is assigned. These components and subcomponents are interacting with one another when producing any piece of language, i.e. a phrase, clause or sentence, in terms of X-bar theory, *subcategorization restrictions* and *selectional restrictions*. As far as X-bar theory is concerned, it can be briefly simply viewed as a universal principle of UG which provides a hierarchical



three-level structure of every produced phrase. However, since it is not much relevant to our main goal, I will not discuss it any further here. Suffice here to pinpoint the main focus of P&P, which is the basic elements of the lexicon and their structure (i.e. syntax) and relations are the main objects for syntactic study. These are manifested in syntactic structures, i.e. phrases and their construction. However, before deeply probing this aspect, it is essential to first introduce what UG is and how it relates to lexicon.

### 3.1. UG and Lexicon

As has been alluded to above, UG constitutes part of the FL every human possesses. It can be simply defined as a set of mental rules (specifications) every human genetically, biologically and innately is endowed with. These rules are of two types: *principles* and *parameters*, whereby the latter regulate the former. Take, for instance, the *head principle* which states that *every phrase must have a head*. However, the position of this head in relation to its complement (i.e. either to the right or to the left) varies from one language to another. This variation, in terms of two choices, is called *head parameter*. In other words, *head parameter*, for instance, verifies/specifies *head principle*. In addition, while principles are language-universal (i.e. found in all languages), parameters are language-specific (i.e. every language chooses one choice of only two choices). Take the PP, as an example, we say that the head of PP is P but the position of its DP complement differs. For instance, in languages like English and Arabic, the head comes first as in *In the house* and *fi l-bayt-i* (in the house), respectively, while in languages like Hindi, head comes last as in *ghara mea* (house in). Parametric variation is also manifested at the level of VP. For instance, in English and Arabic head comes first as in *eat an apple* and *?akal-a tufaahat-an* (ate an apple), English and Arabic, respectively. However, in languages like Korean the head (i.e. the verb) comes last as in *muneul dadara* (door close).

Now, the question is that is it UG *per se* that constitutes human FL? In fact, *not*. In other words, from a neurologic perspective, the FL is divided into subfaculties. One of these subfaculties is occupied by the human lexicon. Lexicon is a dictionary-like organ where all words we know about a language are stored. However, the words stored in the lexicon have nothing indicating any kind of features encoded on such items,

i.e. it is not clear which lexical item can co-occur with which lexical item. Thus, sentences like *\*Ali wrote* and *\*Alia gave a book* are expected to be encountered which are both ungrammatical. The former is ungrammatical due to the fact that the verb *write* has to have a DP object. The same thing can be said about the ungrammaticality of the latter though some kind of difference does exist. The difference lies in the fact that the verb *write* is monotransitive and *gave* is ditransitive. In other words, the verb *write* has to have one DP as an object, but *gave* has to have two DPs. However, a question arises here, i.e. is it possible to put any DP(s) in the complement slot of both verbs? Before answering this question, it is necessary to examine sentences having the same verbs seen above as in *??Ali wrote a camel* and *??Alia gave a mountain a book*. Now, looking at both sentences and though they are syntactically grammatical, they still sound odd, or otherwise semantically ungrammatical. If this is true, it follows that we need particular DPs to function as objects of both verbs. For instance, the verb *write selects* a DP from a set of lexical items including *letter, book, paper, research*, etc. The same thing can be said regarding the verb *give* where the possible words to be selected include *friend, Ali, Ahmed, teacher, student*, etc. as in *Ali wrote a book* and *Alia gave Ali a book*, respectively. In fact, syntax does maintain that every verb, for instance, has the ability of having a constituent(s) as object(s), but it does not specify which or what type of this object. The underlying ability of a lexical item to have an object is called *subcategorization restrictions* of that item.

### 3.2. Subcategorization Restrictions

In P&P framework, subcategorization *Restrictions* are said to be properties encoded in every lexis, be it a verb, noun, adjective, etc. As stated above, regarding verbs, for example, subcategorization restrictions differ from a verb to another. In other words, there are verbs which subcategorize for no constituent like *laugh, die, smile, sneeze*, etc. There are also verbs subcategorizing for one constituent which might be a DP, PP, VP, or a clause like *cook, rely, try or believe*, respectively. There are also verbs subcategorizing for two constituents like *give, consider, send*, etc. There are also verbs with multiple subcategorizations like the verb *grow* in *Ali is growing (this year), Ali is growing carrots (this year) and Ali is growing his field Indian carrots (this year)* (Shormani, 2014c). The following subcategorization frame is provided for verbs and similar ones can be analogized for the other lexical items, like nouns, adjectives, prepositions, etc.

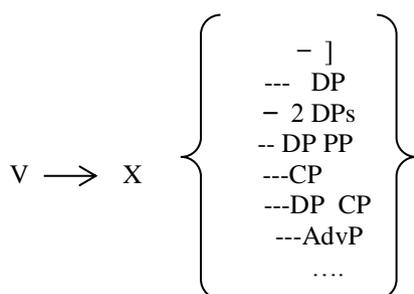


Figure 1: Verb Subcategorization Frame (cf. Ouhalla, 1999)

The frame represented in Figure 1 above is a general one, where V stands for a verb, the arrow ( $\longrightarrow$ ) stands for the context, and X is a variable standing for the syntactic category functioning as a complement. The context [---], for instance, represents the *subcategorization frame* for verbs like *laugh*. The frame CP represents verbs like *think* as in *I think (that) you are a good student*.

#### 4. SEMANTICS

As has been stated above, semantics is a component of UG. Now, recall from the section above that subcategorization restrictions result in sentences like *??Ali wrote a camel* and *??Alia gave a mountain a book*. Though these sentences are syntactically grammatical (by virtue of having the required arguments (i.e. subjects and one or more complements), they still sound odd, but then the question is why is it so? In fact, the oddity of such utterances gives us a clue that it is not any DP(s) that can function as a complement but rather a specific one. The specific type of such DPs to function as a complement in a particular context is called *selectional restrictions*. This is sketched in the following section.

##### 4.1. Selectional Restrictions

As has been stated above, subcategorization restrictions do not state anything regarding the specific type of lexical items that can function as complements of verbs, for instance. We have also argued for there being some kind of semantic constraints (restrictions) to be imposed on the lexical item to be able to function as a complement, or otherwise as an argument in general. These constraints are called *selectional restrictions* which can simply be defined as the semantic restrictions a predicate imposes on its arguments (i.e. the external and internal arguments), for instance, in the case of verbs, the subject and the object, respectively. In other words, a lexical item, say, a verb has, in addition to subcategorization restrictions, some semantic restrictions which are essential “for computing semantic relationships between elements in a sentence” (Myers&Blumstein, 2005, p. 279). If those restrictions are violated, such relationship will not be computed (i.e. structured). Thus,

while subcategorization restrictions are concerned with the number of and the categorial types or c-selection (i.e. Category selection, i.e. DP, AP, PP, CP) of the constituents that occur to the right of a lexical head, selectional restrictions are concerned with the S-selection (semantic selection) of such constituents in both sides of the head (sometimes called predicate). The former are referred to as the ability of the head to restrict the number and type of the internal arguments and the latter as the ability of the head to restrict the type of both internal and external arguments which can be assigned  $\theta$ -roles. In other words, selectional restrictions restrict the categorial features of the constituent that can occur not only to the right of a particular lexical head but also the one(s) occurring to its left. However, a difference is to be noted here between both restrictions.

Unlike the subcategorization properties which are listed in the dictionary, selectional restrictions are not listed with the lexical item as entries. Some linguists (e.g. Ouhalla, 1999) argue that violation of selectional restrictions goes against our senses and our world and encyclopedic knowledge. In fact, it was Chomsky (1965, *et seq*) who first describes selectional restrictions as a phenomenon, maintaining that such a phenomenon exists in any language. He adds that there is some kind of semantic ungrammaticality or oddity arising from violation of selectional restrictions as in *\*Ali eats a mountain* where the verb *eats* needs an object (subcategorization restrictions) and this object must be edible, i.e. *eaten* (selectional restrictions). However, as argued for above, these restrictions apply to both arguments of a verb. For instance, sentences like *\*The book laughed* where the subject of the verb *laughed* is the DP *the book* violate the selectional restrictions imposed by the verb *laugh*. However, these restrictions are not applied when the arguments of the verb are sentential as *He eats what the mountain gave him* where the nominal clause *what the mountain gave him* functions as the complement of the verb *eats*. There are also some verbs which are used incorrectly by L2 learners of English like *drive* and *ride* as in *\*Ali drives a bike* and *\*Ali rides a car*.

Thus, from a P&P perspective, the above argument gives us a clue that the relation between UG and lexicon is more or less manifested in subcategorization restrictions where a lexical category, be it a verb, noun, adjective, etc. subcategorizes for a complement. However, subcategorization restrictions do not specify what goes with what in a phrase. As noted above, this could be compensated by selectional restrictions which are part of semantics. These selectional restrictions are represented by categorial features encoded on the lexical items. These categorial features are manifested in terms of inflections.



Now, assume that both subcategorization and selectional restrictions are manifested in a collocability rule like (1) (see Rögnavaldsson, 1993, for a similar proposal).

### (1) *Collocate Rule (CR)*

Based on the categorial features of  $\alpha$  and  $\beta$ , (i-viii) hold:

- i. A particular noun  $\alpha$  is usually modified by a particular adjective  $\beta$ .
- ii. A particular adjective  $\alpha$  usually modifies a particular noun  $\beta$ .
- iii. A particular verb  $\alpha$  is usually modified by a particular adverb  $\beta$ .
- iv. A particular adverb  $\alpha$  usually modifies a particular verb  $\beta$ .
- v. A particular verb  $\alpha$  usually takes a particular subject  $\beta$ .
- vi. A particular noun  $\alpha$  usually occurs as a subject of particular verb  $\beta$ .
- vii. A particular verb  $\alpha$  usually takes a particular object  $\beta$ .
- viii. A particular noun  $\alpha$  usually occurs as an object of particular verb  $\beta$ .

Now, it becomes clear that a particular lexical item collocates with another lexical item based on each's categorial features.

However, given (1) above, it seems to me that P&P could not account for collocations. In other words, in P&P, it was assumed that lexical items are mapped from lexicon onto syntax, *uninflected* for anything in a ready-made chunks (i.e. phrase markers), and then on syntax they get inflected for all features. **This is an empirical weakness of P&P framework.** Thus, it is not even clear how, where and even when mapping of lexical items from the lexicon takes place and how it could be accounted for. This actually seems to be problematic since every lexical category can take any constituent to combine with. In other words, in P&P framework, every lexical item is inserted separately and it is only in syntax that some kind of assignment like  $\theta$ -assignment, case assignment, etc. is done. As far as collocations are concerned, there are no clear restrictions as to which lexical item that can collocate with which lexical item, and hence, there is no any signal of preference to any particular collectability, given the fact that the meaning of a collocation is compositional (i.e. no obligatoriness is maintained). However, all these problems seem to be eliminated under minimalism.

## 5. MINIMALISM

Starting this section, let me first state why this study is employing a formal approach to the study of collocations, hence, integrating syntax and semantics. In fact, it is so because of a current trend (see e.g. Chomsky, 2005;

Lardiere, 2009; Shormani, 2014b&c, in press; Fitch *et al.*, 2005) that language acquisition studies and researches be tackled under the PF and LF interfaces. Where these interfaces meet is, in fact, at the heart of any language acquisition study, specifically, if what matters most is feature specification shared by both interfaces. In other words, feature specification encoded on every lexical item is what determines how interface takes place. As far as collocability is concerned, what makes a particular lexis collocate with any other lexis is the feature specifications encoded on both. For why this study is employing minimalism is because minimalism, as a theory of UG, is at the heart of language acquisition simply because the core of generative grammar in general and minimalism in particular has been how to account for how language acquisition process takes place, be it either of L1 or L2, and what natural language in general is all about.

As far as generative syntax is concerned, Chomsky (1965) proposes that "study of language is no different from other complex phenomena" (p.4). Since then, language has been subjected to scientific and logical methodology and undergone several developments up to date. Thus, the latest of such developments is *minimalism*. Minimalism "takes the language to consist of [two modules, namely] the lexicon and a computational system" (Ouhalla, 1999, p. 404f), thus, "eliminating DS and SS levels of representation, which leads to minimizing the load placed on language faculty" (Shormani, 2014d, p. 1), maintaining only PF and LF as interfaces. The major concern of minimalism is that "linguistic theory should make use of as few primitive notions as possible" (Kremers, 2003, p. 41). Very crucial to minimalism is what has been called by Chomsky (2000, p. 97) the Strong Minimalist Thesis (SMT) which is outlined in (2).

- (2) Language is an optimal solution to legibility conditions

In Chomsky's view, the notion of "legibility conditions" is related to interface properties. In other words, the idea tied to legibility conditions is that the core  $C_{HL}$  (or otherwise, the narrow syntax) makes available the optimal way of relating an arbitrary set of lexical items to the interfaces, i.e. LF (i.e. the conceptual-intentional system) and PF (the articulatory-perceptual system). In fact, (2) could be understood in such a way that language is an optimal solution of the  $C_{HL}$  to the constraints imposed by the two interfaces. These interfaces are related in such a way that meaning is tied to sound in such a way as to satisfy whatever conditions imposed by the *intrinsic properties* of the lexical items and the interfaces. In other words, minimalism seeks an adequate



linguistic theory, seeks to describe what is so called the I-language, i.e. the UG (or the linguistic endowment every child is born with) of “an ideal speaker-listener in a completely homogenous speech community who knows its language perfectly” (Chomsky, 1965, p. 3) and not his E-language, i.e. what he says in a particular situation, i.e. performance, though the latter has been paid much attention to in recent work specifically regarding SLA (I return to this point below). Chomsky (2013, p. 36) maintains that the fact that “[e]mbedding the study of I-language in the biolinguistic framework is entirely natural; an individual’s language capacity is, after all, an internal property” still persists. However, UG, as an internal and integrated part of human FL, and in its “technical sense” should “not be confused with descriptive generalizations” like those advocated in Greenberg’s universals, as stated so far. In addition, as noted by Chomsky (2005, p. 6), language is a matter of three factors interacting in its design: i) genetic endowment, ii) experience, and iii) principles not specific to the faculty of language. However, as far as language acquisition is concerned, Lardiere (2009, p. 217) argues, there are also principles that are made use of in language acquisition and other domains, and principles of structural architecture and developmental constraints that enter in canalization and efficient computation.

Thus, though the generative framework has witnessed several developments, the culminate state of which is minimalism, what makes the latter different from previous approaches is the fact that it develops a minimalist theory for UG, eliminating complexities and defines linguistic pieces of language as optimal realizations resulting from interaction between LF and PF levels of representation where the *Derivational Economy Principle* determines their optimality, which, in principle, enables the  $C_{HL}$  to select from a set of derivations the optimal ones (Chomsky, 2001). Thus, it is the  $C_{HL}$  which is responsible for mapping Lexical Arrays (LAs) in the numeration from the lexicon introducing them onto the syntax to generate longer structures finally spelled-out as pieces of language.

The  $C_{HL}$  has three operations: first, the *Select* operation selects a LA only once and hence, meeting the *Inclusiveness Condition* (=IC) which prevents other new elements and/or features to be introduced into the computational system once more. Examples of this is the nouns’ [-/+Def], Case and  $\phi$ -features that are assigned in the numeration either as intrinsic features by lexical entry or optional features by the numeration operation. Second operation is called *Merge*. This operation merges two items (functional and/or lexical) ( $\alpha$ ,  $\beta$ ) (or otherwise word 1 (=W1) and word 2 (=W2)) forming  $K(\alpha, \beta)$  asymmetrically projecting into a single item, either  $\alpha$  or  $\beta$

where the projecting item becomes the head and hence, the label of the resultant complex, i.e. *a phrase* (i.e. a collocation) (Chomsky, 2013), and *Merge* is recursive, i.e. it takes place as many times as necessitated by the derivation and the nature of the piece of language under consideration. The third operation of  $C_{HL}$  is *Agree* which establishes a relation between a lexical item  $\alpha$  and a feature  $F$  in some restricted search space (Chomsky, 2001). This relation is then manifested in what is so called *agreement*. For instance, Case-checking is an *Agree* relation established between a lexical object say  $K$  labeled  $LB(K)$  and a feature  $F$  in some restricted search space (i.e. the  $K$ - $F$ ’s  $c$ -command domain in either directions) (Chomsky, 2000, 2001) where each is  $\phi$ -complete in order for the valuation process to take place. Since the  $LB(K)$  is the only element of  $K$  that is immediately accessible to a language  $L$ , it has to be the element that activates *Agree*, by virtue of its unvalued features: these constitute a *Probe* that seeks a matching *Goal* within the domain of  $LB(K)$ . The relation *Match* is taken to be an identity (Chomsky, 2001, p. 4). Matching of probe-goal induces *Agree*, eliminating unvalued features that activate them. However, what concerns us here is the first two operations, namely *Select* and *Merge* because these two operations are exactly where collocability in its technical sense takes place.

What language educators and teachers need to pay much attention to is that language is not a haphazard phenomenon, but rather a precise and concise system, perfect in itself. It is, in other words, and from a minimalist perspective, an *optimal solution to legibility conditions* manifested in the interaction between the UG modules, specifically syntax and semantics, which in turn interact with the  $C_{HL}$  leading to the satisfaction of the legibility constraints. The term *legibility* is crucial to this interaction in the sense that it is not true that any word collocates with any other word, but rather any resulted phrase (i.e. collocation) should meet and satisfy the interfaces, simply, *sound* and *meaning* and here lies the canonical nature of *Select* and then *Merge*. Since feature specifications are properties encoded on words, and since these specifications are often represented by inflections, it follows that they represent *sound*, i.e. PF. And since *sound* is not enough to “pass” a collocated object (i.e. phrase) of, say minimally, two words, and since these two words, as a priori, are selected based on each’s feature specifications, it follows that the resulted phrase (or collocate) necessarily satisfies LF. This is, I believe, very crucial to language educators, simply because if they are aware of the mental processes a particular piece of language undergoes when created, their task of teaching (i.e. orienting or reorienting) their learners will be planned earlier (i.e. in advance), so that when language learning process commences (usually in



classroom), they will be ready to perform their expected role. *Ready*, in the sense that they have the knowledge-base that enables them to deliver their sacred mission, and consequently, undoubtedly, such a mission, i.e. language acquisition process, will be successful, leading after all, to ultimate attainment.

### 5.1. Collectability and Minimalism

As has been stated above, what was a problem for P&P seems not to be there in minimalism. In other words, minimalism discards the earlier distinction between *DS* and *SS* and some syntactic notions like *government* (an X-bar notion where a head governs its complement, e.g. a verb and its object) and *projection principle* (a principle maintaining three levels of a lexical item). For instance a verb is said to have three projections, namely, maximal, i.e. VP, intermediate, i.e.  $V^1$  and minimal, i.e.  $V^0$ , and hence, what matters is only the interaction between *sound* and *meaning*, viz. PF and LF interfaces. To state it again, in P&P, lexical items are inserted from the lexicon uninflected for features, and it is only in syntax some kind of checking takes place. However, in minimalism, specifically, the recent assumptions (see Chomsky, 1995, 2001, et seq), prior to lexical insertion, every lexical item to be selected is specified for every feature it represents, and hence, mapped onto the syntax inflected for all *feature specifications* it carries. Having this in mind, we are in a position to introduce what *feature specification* is all about.

### 5.2. Feature Specification

In P&P, selectional restrictions are said to be represented by categorial features encoded on a lexical item. However, these categorial features are not specified in the lexicon. In other words, the mapping of such features takes place in the syntax which causes serious problems, as noted above. However, in minimalism categorial features are expressed by means of *feature specification*, i.e. every lexical item is specified for all the features it carries in the lexicon. In fact, feature specification is binary in nature, i.e. features are specified in a binary fashion [+/-] signs in square brackets. For instance, the noun *book* is [+N, +inanimate, -abstract]. The noun *honesty* is [+N, + inanimate, + abstract]. Nouns are specified as [+N, +A, -V]. In that, nouns can take adjectives as modifiers but not verbs. Verbs are specified as [+V, +Adv, -A], in the sense that verbs are modified by adverbs but not nouns or adjectives. Verbs, in case of arguments are specified as [+V, +/- Trns], in the sense that verbs can be transitive or intransitive and so on. The same is also applied to other lexical categories. But then, a question arises, i.e. does it mean that this is left open by UG? This is, in fact, an empirical question. I assume that there are two possibilities: i) UG leaves such feature specifications, and ii) it does not leave it. If we assume

that UG leaves feature specification, but, then, how is it that native speakers do not make errors (as different from mistakes, I return to this point below)? An answer to this question is that UG does not leave feature specification open. In other words, if UG has nothing to do with feature specification, it follows that native speaker of a language, say English, for instance, will commit errors in collocation, which is not the case. Thus, it seems that (i) is ruled out. Thus, (ii) is the only possibility remaining. In fact, UG, in addition to providing us with subcategorization rules, provides us with some kind of “selectability,” i.e. feature specification which makes us, as native speakers, able to “select,” as a priori, the context that “fits” a particular lexis to co-occur with. For instance, native speakers of English select the adjective *strong*, for instance, to collocate with *tea* rather than the adjective *powerful* to collocate with it. Thus, feature specification, in relation to collocability, can be formalized as in (3) below.

### (3) Collocating Feature Specification Rule (CFSR)

For two lexical items  $\alpha$  and  $\beta$ ,  $\alpha$  can collocate with  $\beta$  iff:

- (i)  $\alpha$  corresponds to the feature specification of  $\beta$ .
- (ii)  $\beta$  corresponds to the feature specification of  $\alpha$

In addition, every two lexical items selected are combined by means of *Merge* operation which is strictly local. In other words, and as far as collocability is concerned, collocations are determined in the lexicon. After two collocates are selected, they are then mapped onto syntax not in ready-made chunks (as was assumed in the P&P), but rather in a step-by-step fashion, and in a bottom-up manner, in such a way that there is no ready-made slot that any collocate can occupy. If this is true, it follows that this allows us to postulate that ***derivation (and hence collocability) is done in the lexicon, and that syntax and semantics are a matter of the two interfaces.*** The former is the ‘phonetic representation’ legible to the sensorimotor system and the latter the ‘semantic representation’ legible to the conceptual system (Chomsky, 2000). Based on the fact that every word has feature specification, it is not true that a word can collocate with any other word, in such a way that a word can be inserted in a ready-made slot as was assumed in P&P. In other words, a word selects the category of the word to collocate with, in addition to selecting the particular type of such a word based on feature specification. In addition, one principle of UG is the *Binarity Principle which states that every node is branched into two and only two branches.* In other words, this principle means that every node is projected into two branches. Every branch points to, goes to, and entails the existence of another node.



As far as SLA is concerned, central to minimalism is the assumption that language acquisition is seen as acquiring features peculiar to L2, which could be different from those of L1. As Lardiere (2009) argues, language acquisition is now viewed as a matter of feature configuration and the very task of L2 learner is to “discern these specific configurations of features from the properties and placement of particular lexical items present in the linguistic input.” This is due to the fact that given the availability of principles (in terms of UG universal specifications), what makes L1 different from (or similar to) L2 is the feature specifications. These are the most important concern of the now-proposed developments and assumptions of syntactic theory. As argued by Travis (2008, p. 23), features and feature specifications are considered “the heart of recent Chomskyan syntactic theory and within this theory at the heart of language variation. Therefore, any study of language acquisition done within this framework is now a study of the acquisition of features.”

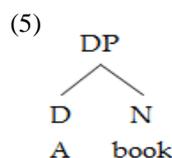
Now, if what makes L1 different from L2 and even L3 is features and feature specification, the question is where these features come from? A question that has been addressed by Lardiere (2009) who tries to answer it stating that since every human, if exposed in early childhood to “any human natural language” will acquire such a language in the same way a native child of this language does, it follows that “there is a universal set or inventory of linguistic Features” every human is biologically, genetically and innately endowed with “as part of the human genetic endowment, along with a species-uniform computational mechanism that combines and interprets the relevant features in a highly constrained way.” Now, since every language is different from any other, as has been stated above, “the child’s acquisition task is to select only that subset of features actually detectably deployed in the particular language(s) being acquired,” and that he/she disregards, discards or even forgets those which do not belong to his/her L1 because they are not in the linguistic input he/she is exposed to (Chomsky, 2001, p. 10; Rizzi, 2005, p. 74; Lardiere, 2009, p. 174).

Now, given the above argument that every human child is endowed with a C<sub>HL</sub> (i.e. a computational mechanism) that combines (collocates) LAs in a hierarchical fashion, and *interprets* them in terms of the relevant features encoded on these LAs, it follows that as far as syntax is concerned, what applies most in collocability is *Select* and *Merge* operations. For semantics, every child is equipped with an ability to *select* and/or *interpret* these LAs. Thus, the operation *Select* selects LAs (i.e. lexical items/words) from the lexicon *only once*, as maintained by IC, and the operation *Merge*, then, comes to play,

hence, merging these two LAs (and if the collocation consists of more than two words, *Merge* operates accordingly). Now, consider the syntactic CFSR applied to (4a) and (4b). Consider also that CSFR applies to (4a) but not to (4b). The former is syntactically and semantically well-formed while the latter is not so, and hence, cannot be interpreted because it violates CFSR, thus, not spelled-out.

- (4a) A book
- (4b) \*An book

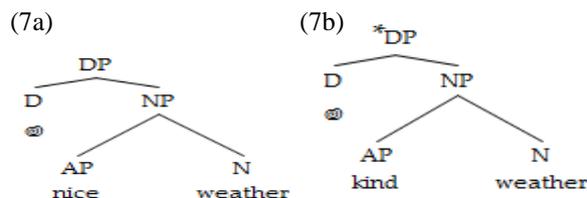
Now, the derivation of (4a) proceeds as follows. The head D, i.e. *a* is selected from the lexicon and merges with its complement *book*, and hence resulting in a syntactically and semantically grammatical collocation because there is a subcategorization property, i.e. the D *a* having a complement, viz. the N *book*, and there is a selectional property, i.e. *a* is collocated with words beginning with consonant sounds, both being satisfied. This is schematized in (5).



However, (5) is ungrammatical because the subcategorization property is satisfied while the selection property is not (qua the D *an* is collocated only with words beginning with vowel sounds). Thus, one step further is in order here by considering collocations like (6a) and (6b)

- (6a) nice weather
- (6b) \*kind weather

In (6a), there is collocation in which the adjective *nice* collocates with the noun *weather* and the result is a grammatical collocation. However, the collocation in (6b) is ungrammatical. This is schematized in (7a) and (7b), respectively.



In (7a), the noun *weather*, i.e. the head of the DP is selected from the lexicon via the *Select* operation which is merged with the A *nice* via the operation *Merge*, and hence, constituting the DP *nice weather*. Now, after the derivation is complete, the DP is sent to the PF and LF



for processing where it is assigned a phonetic and logical (semantic) representation, and hence, spelled out. However, the A *kind* and the N *weather* schematized in (7b) are selected and merged in the same way like (7a) but when the whole DP *kind weather* is sent to LF it crashes, i.e. it is banned, and hence, not sent to PF, and consequently, not spelled out. Now, the question is why is it impossible for the DP or otherwise the collocation *kind weather* to be spelled out? In fact, the reason is that *kind weather* is not a possible collocation in English. In other words, the N *weather* cannot collocate with the A *kind* because the A *kind* is specified for features like [+human], [-nonhuman]. Interestingly, note that the A *nice* is specified for features like [ $\pm$ human]. In other words, we can say *a nice man* and *nice weather*. However; we can say *kind man* but not *kind weather*. Note also that the noun *weather* is specified for [-Art] which means that articles like *a* or *an* cannot collocate with it, hence, indicated with  $\phi$ .

Now, recall from section 3.1 that UG provides human with a set of specifications called *principles* being universal along with *parameters* that specify such principles, and if this is true, it follows that the environment which provides the child (in the case of L1 acquisition) with the linguistic input necessary to set such parameters is needed to provide L2 acquirers with an “equal” linguistic input capable of making them reset such parameters given that principles are universal. Accordingly, L1 acquisition is a parameter-setting and L2 acquisition is a parameter-resetting and retriggering process (for a comprehensive discussion on this, see Shormani, 2014b). Now, the question is whether or not UG is still available to L2 learners. In fact, it has been proved true that UG is still available to L2 learners (see Shormani, 2012a&b, 2014b, in press; White, 2003; Cook, 1988, 2003; Saville-Troike, 2006; Gass & Selinker, 2008, among many others). Given this, and the fact that L2 acquisition takes place when there is already a language (i.e. the learner’s L1) existing in the brain of L2 learners, it follows that L2 acquisition is nothing but “parameter-resetting” (see Shormani, 2014b for an approach to LA manifesting all these assumptions). However, if, as we have assumed above, UG is still available to L2’s learners, it follows that the ability of collocating lexical items is still available for L2 acquirers. However, a crucial issue remains! How can syntax and semantics be integrated, and hence, accounting for collocability? To answer such a question is the main concern of what follows.

### 5.3. Integrating Syntax and Semantics: the Proposal

It has been widely held that English is a richly collocated language (see e.g. Lyons, 1977; Leech, 1981; Kreidler, 2002; Shormani, 2012a; Shormani & Sohmani, 2012; McCarthy, 1990; Lewis, 1997; Mahmoud, 2005; Firth, 1957; Griffiths, 2006, among many others). In addition, as has been noted above, collocability is a characteristic of native speakers *per se* for the fact that every speech community establishes its own ways of expressing meaning. If this is true, acquiring collocations by L2 learners of English would be expected to be of a considerable difficulty, because their nature involves formulaic language, and the way they are formed involves arbitrariness and is considered one of the most difficult areas for L2 learners of English. Thus, collocation difficulty encountered by L2 learners is best manifested in errors committed by L2 learners, whatever the language being learned might be. In addition, collocation errors have been classified into different types. Such types include (but not limited to) interlingual vs. intralingual (see Ridha & Al-Riyahi, 2011; Shormani & Sohmani, 2012; Sonaiya, 1991; Mahmoud, 2005, 2011; Zughoul, 1991; Shormani, 2012a&c; James, 1998; Wang & Wen, 2002; Li, 2005, just to name a few). However, a very comprehensive classification has been done by Shormani (2014c) based on their sources into four categories, viz. *L1-transfer*, *L2-influence*, *Mutual* and *Unrecognized*. However, what concerns us most here is those ascribed to *L1-transfer* and *L2-influence* (see **Appendix** for some examples, from Shormani, 2014c). Thus, based on such two sources, I assume that making use of the proposed approach of integrating syntax and semantics presented in Figure 2 below makes us understand the nature of collocation acquisition and how they could be best acquired, hence, it is hoped, providing linguists, language educators and teachers with some sort of knowledge-base of how understand and teach collocations to their students.



Figure 2: Integrating Syntax and Semantics

How the proposal in Figure 2 above works can be simply put forth as follows. The whole of it represents the knowledge an acquirer/learner has about second language (say, English). This knowledge allows him/her to produce collocations but not all of such linguistic production meets UG. However, since he/she still has access to UG, things can be changed and/or reoriented. In other words, since what matters is parameters and not principles, and since, the still-access to UG comes by means of language acquisition/learning, such parameters of L2 can be reset (see Shormani, 2014b for a comprehensive discussion on how these parameters are reset and retriggered in the case of L2 foreign setting). Thus, this knowledge is represented by the lexicon (i.e. words), syntax (i.e. the syntactic rules he/she acquires) and semantics (i.e. the semantic rules of what goes with what, superficially feature specifications) which, I assume, can be reset and retriggered once the learner is exposed to that kind of acquisition/learning that is able to “orient” his/her presumed and/or perceived L2 knowledge in an inappropriate way. This is, in fact, interacted in the proposed approach presented in Figure 2.

In a minimalist technical sense,  $C_{HL}$  sees *Select* and *Merge* as the primitive operations all subsequent operations/relations result from. If this is true, it follows

that a given phrase (i.e. collocate) is made up of words ( $W_1, W_2, W_3...W^n$ ), starting building such a phrase in pairs of  $W_1+W_2$ , and so on. For instance, when *Merge* combines  $W_1+W_2$ , the result is  $aP$  (i.e. a collocate, where  $a$  is the label given to the newly constructed phrase). In figure 2, given the fact that lexical items are stored in the lexicon (qua a dictionary-like organ in the L2 learner’s brain), the operation *Select* selects first the head word symbolized by the capital letter  $W_1$  (i.e. the first collocate). This  $W_1$  can be a verb, noun, preposition, etc. Then, it selects the second word symbolized by the capital letter  $W_2$  (i.e. the second collocate). This  $W_2$  can be either an argument (i.e. a subject or object) or a modifier (i.e. an adjective, adverb, etc.). Now, the operation *Merge* merges  $W_1$  and  $W_2$  constituting the new lexical object  $W_1+W_2$  as a collocation. Given that the operation *Merge* takes place in the syntax, the new created lexical object is then sent to the semantics. In semantics, if this new created lexical object “passes” CFSR given in (3) (i.e. semantics), it is a correct collocation and will proceed as output, being syntactically and semantically grammatical (as indicated by the black arrow). In such a case, the new created lexical object is sent then to the phonological and semantic components for processing, and hence, spelled-out as a syntactically and semantically grammatical collocation.

On the other hand, if the new created collocation by *Merge* does not “pass” CFSR, specifically, it will proceed as an incorrect collocation (as indicated by the blue arrow, in the opposite direction). This incorrect collocation will have to be corrected, but how? In fact, this is a very substantial point where things change dramatically and radically provided that the L2 acquisition setting provides the learner with all that he/she needs of assistance to correct this misperceived piece of language (specifically regarding the collocation in question). Why this huge importance the proposal gives to L2 acquisition setting is due to the fact that the assistance the learner receives could change his knowledge about a particular concept in the target language or completely the otherwise. In other words, the proposal places huge emphasis on the rule played by L2 acquisition setting (usually the teacher) in reorienting the learner and reguiding him/her to correct his misconceived information about such-and-such (I return to this point below) simply because if such misconception about the collocation in question is not corrected (specifically in the acquisition process), it may get fossilized (see Shormani, 2013a; Han, 2000, 2004).

Thus, if the learner gets the semantic reorientation that allows him/her to correct the misperceived knowledge, he/she will correct the wrongly produced collocation. In



other words, correcting the produced collocation depends heavily on the linguistic input (as a remedial procedure) provided to the learner in L2 acquisition setting (usually a foreign one). Now, since this depends in the first place on the teacher (as is usual maintained in a forging learning setting), and since the learner is aware of the syntactic rules (i.e. the subcategorization properties), the teacher should provide the learner with the CFSR features of both collocates and raises his/her learners' awareness of what goes with what (i.e. both collocates) by means of different and several ways (I shall return to this point below).

Thus, assume that the assistance the language learner receives from the teacher (usually in the classroom) is good enough to make him/her realize what goes wrong with what that makes such produced collocation(s) ungrammatical, and hence, being able to correct the collocation in question, this corrected collocation will proceed to semantics again. The linguistic assistance provided to the learner should be sufficient to make him/her reset/retrigger the parameters that have been acquired wrongly, so that the wrongly produced piece of language (i.e. the collocation) is corrected. Thus, given this, if it, once more, "passes" CFSR (i.e. semantics), then it will proceed as a correct output, i.e. a correct collocation. In other words, like in the first correctly produced collocation, the second new created lexical object is sent then to the phonological and semantic components for processing, and hence, spelled-out as a syntactically and semantically grammatical collocation.

However, a very substantial question that should be addressed here is that suppose the linguistic reorientation provided to the learner were not that good enough to enable him/her to reset the wrongly acquired parameter(s), and the learner continues to produce semantically ungrammatical collocation(s), what is the solution provided by the proposed approach to such a problem? In fact, the proposed approach does not ignore this aspect. In other words, if the resultant collocation after receiving L2 linguistic assistance from the teacher (or whosoever) is not enough and the collocation does not "pass" the semantic component, it has to go back to L2 acquisition setting (as indicated by the curved blue arrow). However, this time, the matter differs. In fact, this happens, I assume, only with weak students, but let us assume that it happens, and thus, I propose that there are two possibilities: i) either the teacher could do the task of reorienting by providing the (weak) learner(s) with more examples that illustrate his/her misconception, and hence, giving him/her some assignments for practice to be able to reset the wrongly learned parameter(s), or ii) the learner himself/herself has to do the task of resetting or retriggering by reading and/or learning about such a

difficult area. In fact, the same thing happens even with Ph.D scholars after college study. As has been stated earlier (see section 1), Ph.D scholars, specifically those whose mother tongue is not English, the author of this article is one of them, continue to produce ungrammatical collocations (either syntactically or semantically), and they got corrected by their Ph.D mentors, but still much of this correction comes through self-reading and self-researching!

#### 5.4. Competence and Performance

Central to collocability is what Chomsky (1965, p. 3) first draws our attention to, i.e. the distinction between competence and performance as:

Linguistic theory is primarily concerned with an ideal speaker-listener, in a completely homogenous speech community who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of language in actual performance.

Thus, as can be elicited from the above excerpt, performance can be simply defined as the production and understanding of utterances in particular languages in concrete situations (Cook, 1996; Schachter, 1996). This gives us also a clue that Chomsky seems to reject or abstract from Saussure's concept of *langue* as merely a systematic inventory of items and returns rather to a more traditional conception of underlying competence as a rule-governed system of generative processes (Chomsky, 1965, p. 4). However, it seems that what Saussure calls *parole* is maintained by Chomsky in what he refers to as performance, i.e. the use of language in concrete situations. However, a question to be raised here: how is linguistic competence acquired or stored in the brain? To answer such a question, therefore, competence must be distinguished from performance and the directionality of the relations of presupposition of dependency within this trichotomy (Alptekin, 2002). In fact, the acquisition of linguistic competence allows for the possibility that the acquisition of competence is partly or even wholly dependent on performance. However, to distinguish between sentences as abstract theoretical entities and sentences as the products of utterance is to introduce what is arguably a false distinction between grammar and language, on the one hand, and between what Chomsky has distinguished as the I-language, i.e. internal mental language, and the E-language, viz. external language, on the other hand. Further, Chomsky draws what is at first sight a puzzling distinction between grammars and languages or alternatively between I-languages and E-languages, arguing that while the former are real and can



be the object of scientific enquiry, the latter are indeterminate, amorphously suspect or mysterious.

Consequently, Chomsky distinguishes between *error* and *mistake* where the former is attributed to *competence* and the latter to performance. Errors are “rule-governed,” “systematic” and repeatedly committed. This has been proved true by the evidence that *competence* errors are not self-corrected and the learner is unaware of (Shormani, 2012a). *Mistakes*, on the other hand, are not “systematic.” Nor are they “rule-governed” and the learner is aware of their committing. Moreover, mistakes are attributable to the speaker’s fatigue, distraction, or inattention. All in all, this lends us a strong piece of evidence that performance imperfectly reflects competence where the latter is intact and perfect, while the former is not. What I want to suggest here is that ‘error’ should be treated differently from ‘mistake’ and that it is our duty to figure out where a wrongly produced collocation is an error that has to be paid much attention to, as it is caused by imperfect competence. In this case, learner’s knowledge of L2 has to be reoriented. However, we, as educators, should also figure out where it is just a mistake as a performance-based deviation. In the latter case, we do not have to worry about reorientation.

Still, however, there remains a very crucial question to be addressed here, i.e. why is it that L2 learners (even advanced ones) commit errors in collocation? This is, in fact, an empirical question the answer to which requires us to go through the following discussion. In fact, it has been held that even 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 (Shormani, 2014c). Recall that UG provides native speakers of a language with feature specification, in addition to the native authentic environment that makes them able to adequately generate (collocate) any given set of items in a well-formed piece of language (maximally a sentence). As far as SLA is concerned, L2 acquirers have to be exposed to “equal” (i.e. the same quality and quantity) linguistic input that native speakers have had. This is due to the fact that if they do not have that ‘equal’ linguistic input, it follows that the SLA process they are in or have been through is not adequate. What I want to suggest is that since L2 acquisition process takes place in a foreign context (i.e. where no authentic exposure is available), I assume that SLA is *often* characterized by “imperfectness.” Thus, if we assume that this is true, (i.e. L2 acquirers do not have that opportunity to be exposed to that “equal” linguistic input), it follows that they are certainly expected to commit errors in their L2 linguistic system in general and in collocation in particular. Thus, it is strongly

recommended that L2 acquirers be exposed to that sufficient, adequate, authentic and accurate linguistic input.

## 6. DISCUSSION

Given the above argument on how competence is different from performance as related to ‘error’ and ‘mistake,’ we are now in a position to proceed to how the minimalist proposal developed in this article can better account for collocation acquisition. As has been stated so far, the major concern of minimalism, as a linguistic theory, has been how to understand the nature of language acquisition, be it of L1 or L2, and consequently help in proposing appropriate approaches, models, procedures, etc. that could help (re)orient and (re)guide our students to accomplish their acquisition task. The article at hand is just an example of how minimalism could be made use of in proposing some sort of approach in relation to such an aspect (see Chomsky, 1965, et seq, up to 2013; Lardiere, 2009; White, 2003; White, et al, 2003; Cook, 1996, 2003; Gregg, 1993; Gruter, 2006; Hawkins & Hattori, 2006; Hegarty, 2005, among so many others, who have devoted themselves to trying to account for language acquisition and help in such process through syntactic theory in general and minimalism in particular). Thus, given the fact that knowledge of collocability is part of the native speaker’s communicative competence as a by-product, lexically and semantically, every “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 (Shormani, 2014c). In addition, it is a fact that collocations are amalgam of both syntax (manifested in terms of *Select* and *Merge* operations), and semantics (i.e. CFSR).

Thus, as far as syntax is concerned, syntax classes should be concerned with how to provide students with the correct way of how collocations are formed, and how *Select* and *Merge* operations each work. In addition, in syntax classes, learners should be exposed to how lexicon works and how it stores our L2 vocabulary in general and semi-fix expressions like collocations in particular. Syntax teachers are also advised to show their students how to enrich their lexicon by learning not only words but also the expected contexts in which they can be used. In addition, in relation to syntax (qua one of the sources of errors found in collocation, see Shormani, 2014a, b&c; Mahmoud, 2005; 2011; Llach, 2005), I agree with Lardiere (2009, p. 219) who rightly argues that it is not of a value to compare “the ways in which grammatical



features of the native language(s) and target languages are organized, or in suggesting that learners use L1 feature configurations as a departure point for what to look for in the L2.” Supporting Montrul & Slabakova (2002), Lardiere points out that those features of L1 signaling the properties of where L1 differs from L2 and where they meet in terms of contrastive analysis of such different and similar properties of both languages, simply because examining such points would lead to making the task of the teacher and the learner alike easier. However, I am not sure whether Lado’s (1957) first proposal could be made use of here or later developments contrastive analysis has witnessed like Wardhaugh (1970) or even more new one proposed by Fisiak (1981) for making such a comparison. For this, I propose that there should be a comparative syntax course of the students’ L1 and that of the L2 they are learning. Though understanding and mastering phonology and morphology of L2 is important, I suggest that syntax and semantics are more important and should be paid much attention to. This is so due to the fact that syntax and semantics are the two components that L2 learners remain in contact with in their use of L2 being learned. In that, principles are universal of which neither the teacher nor the learner should be worried about. What really matters is the other set of such UG properties, namely, parameters, and how to make our students have the ability of resetting and retriggering them if ill-perceived and/or acquired. This comparative syntax course focuses on the different (and similar, if any) parameters regarding collocability in both languages, i.e. the learners’ L1 and the L2 they are learning. This is so due to the fact that once a learner gets aware of how collocability works in his/her language, similar to or different from that of L2, he/she will be able to distinguish those of L1 from those of L2. For instance, while the *taqs-un latif-un* (kind weather) is possible in Arabic, \*kind weather is not possible in English.

As far as semantics is concerned, semantics classes should be utilized in teaching selectional restrictions in the form of feature specifications that are peculiar to words in sets such as verbs, nouns, adjectives, etc. Some kind of contrastive linguistics could also be utilized between the learners’ L1 and L2 of the type recommended by Lardiere (2009). Only in this way can students’ awareness be drawn and/or raised to the different (and perhaps similar, if any), aspects where L1 and L2 diverge and where they (may) converge. As put forth by Lardiere (2009, p. 209), there may be a chance “for missteps if the lexical semantics of a selected lexical counterpart in the L1 does not exactly match that of the L2.” For instance, a native speaker of Arabic may relate the English plural morpheme *-(e)s* to that of Arabic *-uu/ii(na)* which cannot hold true for the difference between both languages in terms of feature specifications

specific to each plural morpheme of each language. The Arabic plural morpheme *-uu/ii(na)* is known as sound plural (there is also another type called broken plural see e.g. Shormani, 2013b; McCarthy & Prince, 1990). In that, the Arabic plural morpheme is a feature of morphosyntactic analysis while English’s is purely morphological. In other words, the Arabic sound plural morpheme *-uu(na)* is attached to the noun when it is in the Nominative Case while *-ii(na)* is attached to the noun when it is in the Accusative or Genitive Case, whereas the English *-(e)s* is used in the three Cases. So, there is some kind of difference in the LF and PF interfaces and the identity each morpheme is spelled out. In addition, the bracketed *(na)* morpheme is used in all cases and deletes only if the noun is in the **Construct State** (a structure specific to Semitics like Arabic and Hebrew, known as *ʔidafa*) as in *muhanddis-uu l-maʔruuʔ-i* (the engineers of the project, see e.g. Fassi Fehri, 1993, 1999; Shormani, in press; Kremers, 2003), where *(na)* is deleted. However, the English bracketed *(e)* is used when preceded by sibilant (i.e. sounds like *s, z, ʃ*, etc.), i.e. purely phonological. Given this, teachers are advised to pinpoint such feature specifications peculiar to each language, and how each works differently from the other based on the feature specifications of a given noun in each language. This could be utilized even in morphology classes.

In *vocabulary* and *study skill* classes/courses, teachers (and language educators alike) should make it clear for students that they should avoid bilingual dictionaries like Arabic-English and/or vice versa for the fact that they are error-provoking rather than error-preventing and that the existence of one-to-one correspondence is not always error-free. This is so due to the fact that depending on such dictionaries leads to committing errors. So, I recommend that English-English dictionaries should be used, instead, specifically those which provide contextual examples where the looked-up word is possible to be used and where it is not.

As far as the syntax and semantics interfaces are concerned, the teacher should guide students to the fact that semantics starts where syntax stops. There is a tendency now that any language acquisition should be done within the interfaces. This will allow for a biolinguistic investigation into the human FL because it determines the portion of human language universality manifested in the interaction between syntax and semantics (Harley & Ritter, 2002; Lardiere, 2009, among others). In fact, the role of integration of syntax with semantics is not denied, particularly in recent studies. To quote Fitch *et al.* (2005, p. 203, also in Lardiere, 2009, p. 219f):



Syntax clearly plays a significant role in our ability to construct and express new meanings, but at least some of the restrictions and complexities of this process are plausibly inherited from conceptual structure, rather than being part of syntax *per se*. Just as the conceptual structure of objects and events surely influences and constrains the properties of nouns and verbs, it seems plausible to postulate that *linguistic devices expressing quantity, tense, aspect or comparison, or other temporal or logical relations, inherit at least some of their structure from the conceptual structure of time, space and logic*. The precise locus of such constraints is an active area of current research in linguistics (emphasis in the original).

## 7. CONCLUSION

Thus, as far as collocability is concerned, I have assumed that while syntax merges (collocates) lexical items in *binary fashion*; semantics specifies which lexical item goes with which lexical item. Interestingly, it seems that semantics behaves in *binary fashion*, too. In other words, a lexical item can or cannot collocate with another lexical item (and there is no possibility for a third state, in the sense *can* and *cannot* at the same time), be it in the form of an argument or a modifier. This lies in the fact that feature specification takes the form of binary choices. Take, for instance, the adjective *kind* and *nice* in *\*kind weather* and *nice weather*, the latter is grammatical while the former is not. This is so because the adjective *kind* is specified for [+human] and [-nonhuman] while the adjective *nice* is specified for [±human]. We say: *strong tea* but not *\*powerful tea*, *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. Further, cohesion is maintained in a text by means of either repeating a lexis, or relating it to another previous one or collocating this lexis with some others. Along similar lines, Halliday (1994) argues that there is some kind of semantic cohesion and/or continuity which can be established in a particular text by repeating a particular word, or by the use of such a word in relation to a previous one either in synonymy or collocability. Thus, if collocability in general and English collocations in particular are a matter of what goes with what in terms of syntax, i.e. *Select* and *Merge* operations, selecting and merging lexical objects, and of semantics, as well, manifested in feature specifications, and since both are endowed by UG, and hypothesizing that UG is still available to L2 acquirers, it could be concluded that we linguists, educators and language teachers can (re)orient our students to the correct way of resetting and retriggering the UG parameters of the target language that have been misconceived previously, hence, contributing in the success of L2 acquisition process. However, it should be taken into consideration that

whatever the proposal developed in this article leads to, it remains a fact that its application is mandatory, and whether or not it comes to our expectations has to be rethought and revisited, and I leave this for future studies.

## ACKNOWLEDGMENTS

I would like to thank the JTTE's Managing Editor Dr. Abdul Razzak for the great effort devoted to developing the journal. I would also like to thank two anonymous reviewers for their invaluable comments having helped improve the layout as well as the content of this article.

## REFERENCES

- Alptekin, C. (2002). Towards Intercultural Communicative Competence in ELT. *ELT*, 56(1), 57-64.
- Hornstein, N. (2009) *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981). *Lectures On Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1987). Transformational Grammar: Past, Present and Future. In *Studies in English Language and Literature*. (33- 80). Kyoto: Kyoto University.
- Chomsky, N. (1993). A Minimalist Program for Linguistic Theory. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 1-52. Cambridge, Mass.: MIT Press.
- Chomsky, N. (2001). Derivation by phase. In Kenstowicz, M., editor, *Ken Hale: a life in language*. Cambridge, MA: MIT Press, 1-52.
- Chomsky, N. (2000). Minimalist Inquiries: The Framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, (eds.) R. Martin, D. Michaels, & J. Uriagereka (89-155). Cambridge, MAA MIT Press.
- Chomsky, N. (2002). *On Nature and Language*. Cambridge: Cambridge University Press.
- Chomsky, N. (2004). Beyond Explanatory Adequacy. In A. Belletti (ed.). *Structures and Beyond: The Cartography of Syntactic Structures*, 3. (104-131). Oxford: Oxford University Press.
- Chomsky, N. (2005). Three Factors in Language Design. *Linguistic Inquiry*. 36, 1-22.
- Chomsky, N. (2008). On phases. In *Foundational Issues in Linguistic Theory*, C. Otero, R. Freidin & M. Zubizarreta (eds.). (133-166). Cambridge: MAA MIT Press.
- Chomsky, N. (2013). Problems of Projection. *Lingua*, 130, 33-49.
- Cook, V. (1988). *Chomsky's Universal Grammar: An Introduction*. Oxford: Basil Blackwell.
- Cook, V. (1996). Strategies and UG Theories. Unpublished Paper Presented at Seville circa Conference.
- Cook, V. (2003). (ed.). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Cowie, A. (1998). Creativity and Formulaic Language. *Linguistica e Filologia*, 8,159-170.
- Cruse, D. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.



- Fan, M. (2009). An Exploratory Study of Collocational Use by ESL Students: A Task-based Approach. *System*, 37(1), 110-123.
- Fassi Fehri, A. (1993). *Issues in the Structure of Arabic Clauses and Words*. Dordrecht: Kluwer Academic Publishers.
- Fassi Fehri, A. (1999). Arabic Modifying Adjectives and DP Structures. *Studia Linguistica*, 53.2.
- Fisiak, J. (1981). (ed.). *Contrastive Linguistics and the Language Teacher*. New York: Pergamon
- Fitch, W., Hauser, M & Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition*, 97, 179-210.
- Fries, C. (1945). *Teaching and Learning English as a Foreign Language*. University of Michigan Press.
- 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.
- Gregg, K. (1993): Taking Explanation Seriously; or, Let a Couple of Flowers Bloom. *Applied Linguistics* 14, 276-94.
- Griffiths, P. (2006). *An Introduction to English Semantics and Pragmatics*. Edinburgh: Edinburgh University Press.
- Gruter, T. (2006). Object (Clitic) Omission in L2 French: Mis-setting or Missing Surface Inflection? In O'Brien, M.G., Shea, C. and Archibald, J., editors, *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*. Somerville, MA: Cascadilla Proceedings Project, 63-71.
- Halliday, M. (1966). Lexis as a Linguistic Level. In C. E. Bazell et al (eds), *In Memory of J.R. Firth*. London: Longman. 150-61.
- Halliday, M. & Hasan, R. (1976). *Cohesion in English*. Longman, London.
- Halliday, M. (1994). Contexts of English. In K. Carlon, K. Davise & B. Rudzka-Ostyn (eds.) *Perspectives of English: studies in honour of Professor Emma Vorlat*. Peeters.
- 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.
- Hawkins, R. & Hattori, H. (2006). Interpretation of Multiple Wh-Questions by Japanese Speakers: A Missing Uninterpretable Feature Account. *Second Language Research* 22, 269-301.
- Hawkins, R. & Hattori, H. (2006): Interpretation of multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research* 22, 269-301
- Hegarty, M. (2005). *A feature-based syntax of functional categories: the structure, acquisition and specific impairment of functional systems*. New York: Mouton de Gruyter.
- 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.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins
- 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.
- Kremers, J. (2003). *The Arabic Noun Phrase*. Netherlands: LOT.
- Lado, R. (1957). *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, University of Michigan Press.
- Lardiere, D. (2009). Some Thoughts on the Contrastive Analysis of Features in Second Language Acquisition. *Second Language Research*, 25, 2, 173-227.
- 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.
- Leech, G. (1981). *Semantics: the study of meaning*. (2<sup>nd</sup> ed.) Harmondsworth: Penguin Books.
- 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.
- Llach, M. P. (2005). The Relationship of Lexical Error and their Types to the Quality of ESL Compositions: an Empirical Study. *Porta Linguarum*, 3, 45-57.
- Lyons, J. (1966). Firth's Theory of Meaning. In C. Bazell, J. Catford, M. Halliday & R. Robins (eds.), *In memory of J. R. Firth*, 288-302. London: Longman.
- Lyons, J. (1977). *Semantics*. London: Cambridge University.
- 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*, Reykjavík, 204-207.
- Mahmoud, A. (2005). Collocation Errors Made by Arab Learners of English. *Asian EFL Journal*. Available @ <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? *Investigationes Linguisticae*, 11, 1-12.
- McCarthy, J. & Prince, A. (1990). *Foot and Word in Prosodic Morphology: The Arabic Broken Plural*. Cambridge: Cambridge University Press.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.



- Mel'čuk, I. (1998). Collocations and Lexical Functions". In A. P. Cowie (ed.) *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press, 23–53.
- Mitchell, T. (1971). Linguistic 'goings-on': Collocations and Other Lexical Matters Arising on the Syntagmatic Record. *Archivum Linguisticum*, 35-69.
- Montrul, S. & Slabakova, R. (2002). Acquiring Morphosyntactic and Semantic Properties of Preterite and Imperfect Tenses in L2 Spanish. In Pérez-Lerouz, A.T. and Licerias, J.M., editors, *The acquisition of Spanish morphosyntax: the L1–L2 connection*. Dordrecht: Kluwer, 113–49.
- Myers, E. & Blumstein, S. (2005). Selectional Restriction and Semantic Priming Effects in Normals and Broca's Aphasics. *Journal of Neurolinguistics*, 18, 277–296.
- Ouhalla, J. (1999). *Introducing Transformational Grammar: from Principles and Parameters to Minimalism*, (2nd ed.). London: Arnold.
- Pawley, A. & Syder, F. (1983). Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In *Language and Communication*, J.C Richards and R.W. Schmidt (eds), 191-225. London: Longman.
- Ridha, A. & 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.
- Rizzi, L. (2005). On the Grammatical Basis of Language Development: A Case Study. In Cinque, G. and Kayne, R.S., editors, *The Oxford handbook of comparative syntax*. New York: Oxford University Press, 70–109.
- Robins, R. (1967). *A short history of linguistics*. London: Longman.
- Rögnvaldsson, E. (1993). Collocations in the Minimalist Framework. *Lambda* 18:107-118.
- Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. New York: Cambridge University press.
- Schachter, J. (1996). Maturation and the Issue of Universal Grammar in Second Language Acquisition, in Ritchie, W. and Bhatia, T. (eds.). 1996, 159-193.
- Seretan, V. (2013). An Integrated Environment for Extracting and Translating Collocations. *Informatics* 1, 11-31.
- 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. (2014a). Lexical Choice Difficulties: A Psycholinguistic Study towards a Solution. *IISTE*, 3, 66-76.
- Shormani, M. (2014b). Mentalo-Behaviorist Approach to Language Acquisition. *JTTE, Scientific Publishing Center, University of Bahrain*, 2(2), 76-92.
- Shormani, M. (2014c). Collocability Difficulty: A UG-based Model for Stable Acquisition. *JLLL*, 4, 54-64.
- Shormani, M. (2014d). Structural Ambiguity in Arabic DPs: A Minimalist N-to Spec Approach. *Qalam Journal*, 1(1), 1-35.
- Shormani, M. (in press). *Introducing Minimalism. A Parametric Variation*.
- Shormani, M. (in press). Semitic Construct State: A Phase Approach.
- Shormani, M. & Sohmani, Y. (2012). Semantic Errors Committed by Yemeni University Learners: Classifications and Sources. *International Journal of English Linguistics*, 2(6), 120-139.
- Sinclair, J. (1966). Beginning the study of lexis. In C. Bazell, J. Catford, M. Halliday & R. Robins (eds.), *In memory of J.R. Firth* (pp. 410-430). London: Longman.
- Sonaiya, R. (1991). Vocabulary Acquisition as a Process of Continuous Lexical Discrimination. *IRAL*, 14(4), 269–284.
- Stroik, T. & Putnam, M. (2013). *The Structural Design of Language*. Cambridge: Cambridge University Press.
- 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.
- Valette, R. (1991). Proficiency and the Prevention of Fossilization – an Editorial. *Modern Language Journal* 75, 3, 326–328.
- Wang, W. & 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.
- Wardhaugh, R. (1970). The Contrastive Analysis Hypothesis. *TESOL Quarterly*, 4, 2, 123-130.
- White, L. (2000). Second Language Acquisition: From Initial to Final State. In J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory*, Blackwell, 130-155.
- White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- White, L. Goad, H., & Steele, J. (2003). Missing inflection in L2 acquisition: defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics* 48, 243–63.
- Wray, A. (2000). Formulaic Sequences in Second Language Teaching: Principles and Practice. *Applied Linguistics*, 21(4), 463–489.
- Zughoul, M. (1991). Lexical Choice: Towards Writing Problematic Word Lists. *IRAL*, 29(1), 45-60.
- Zughoul, M., & Abdul-Fattah, H. (2003). Collocation Strategies of Arab Learners of English: A Study in Lexical Semantics. (ERIC Document No. ED 479746).



**Appendix**  
**Collocation Errors & Sources**  
**(from Shormani, 2014c)**

**Lexical Collocations**

**L1-transfer**

- (a) \*...and our village has kind weather. (nice)
- (b) \*...and there are many pretty areas in Bada'n. (nice places)
- (c) \*because he has a big mind when he speaks. (reasonable thinking)
- (d) \*When my grandfather died, my father became the power man in the village. (powerful man)

**L2-influence**

- (a) \*I feel happy when the teacher gave questions to us. (asked)
- (b) \*.we established two homes: one in village and another in Ibb city. (built two houses)
- (c) \*...and I love syntax because it has a good plan. (convincing/easy syllabus)
- (d) \*and it makes a big different in my life. (a big difference)

**Syntactic collocations**

**L1-transfer**

- (a) \* I am not interested with syntax. (interested in)
- (b) \* He went # college. (went to)

**L2-influence**

- (a) \* Our country depends in Oil. (depends on)
- (b) \* ...this was my solution #the problem. (solution to)