Collocations in Generative Theory

Abdullah S Al-Dobaian
King Saud University

JALHSS مجلة الفنون والأدب وعلوم الإنسانيات والاجتماع
Journal of Arts, Literature, Humanities and Social Sciences – JALHSS
www.jalhss.com ISSN:2414 – 3383
العدد (14) أيار - مايو 2017

المتلازمات اللفظية في النظرية التوليدية

أستاذ مشارك
د. عبدالله بن سعد الضبيعان
قسم اللغة الإنجليزية وآدابها
كلية الآداب – جامعة الملك سعود
الرياض – المملكة العربية السعودية

الخلاصة

لم تخضع المتلازمات اللفظية لنقاش كافٍ في النظرية الأدنوية وغيرها من النظريات التوليدية السابقة لأنها كانت دائماً ما تُحال إلى المعجم الشاذ الذي يتميز عن التركيب المحكوم بقواعد بالرغم من أن المتلازمات اللفظية قد تجمع بين سمات معجمية وتركيبية. في هذا البحث، أبين أن ثمة حاجة ماسة لتخفيف التمييز الصارم بين المعجم والتركيب لنتمكن من تقديم تحليل مناسب للمتلازمات اللفظية. يمكن أن تعالج النظرية الأدنوية، كما فعلت النظريات المغايرة لتشومسكي، الخصائص الأساسية للمتلازمات اللفظية. وبناءً على ذلك، أبرهن على وجود قواعد عامة تُطبَّق على المعجم وعلى التركيب معاً لتحقيق أهداف بنيوية ودلالية. تعتبر قاعدة الضم إحدى أهم تلك القواعد التي تُطبَّق في المعجم لتكوين بنية مُرَكَّبيَّة (على سبيل المثال المتلازمات اللفظية والتي يتعلمها متحدث اللغة كتراكيب جاهزة). أوضح أن قاعدة الضم هي نتيجة مباشرة لمُسلَّمة التوافق الخطي لكاين والتي تتطلب أن تكون مكونات البنية المُرَكَّبيَّة مُرَتَّبة خطياً. تطبَّق قاعدة الضم أيضاً في التركيب كما تبين الأدبيات اللغوية لنظرية تشومسكي السائدة كنتيجة لمُسلَّمة التوافق الخطي لكاين.

Collocations in Generative Theory

Dr. Abdullah S. Al-Dobaian
Department of English and Literature
College of Arts – King Saud University
Riyadh – Saudi Arabia

ABSTRACT

Collocations are not examined adequately within the minimalist program and other earlier mainstream theories because they are relegated to an irregular lexicon that is distinct from a rule-based syntax despite the fact that collocations may have lexical and syntactic properties. In this paper, I argue that we need to relax this this strict division between the lexicon and the syntax in order to account for collocations. The minimalist program can address, as other non-Chomskyan theories, the basic properties of collocations. Therefore, I argue that there are general rules applying at the lexicon and the syntax to achieve specific structural and semantic goals. Merger is one important rule that applies in the lexicon to generate a phrase structure (i.e. a collocation structure that speakers of the language learn as ready-made chunks). I argue that merger is a direct result of Kayne's Linear Correspondence Axiom (LCA) requiring the phrase structure constituents to be linearly order. Merger also applies to the syntax as explained in Chomsky's mainstream tradition and is conditioned by the LCA.

1. Introduction[1]

Collocations are not well studied within the Chomskyan mainstream generative theory since they are regarded as a performance and not a competence matter. As a matter of fact, collocations pose a serious threat to the strict distinction between the lexicon and syntax which the mainstream generative theory assumes. Because they are not constrained by syntactic rules, they are considered to be marginal. However collocations show different ranges of regularity that may be better analysed in terms of a continuum. I argue that there are general principles at work in both the lexicon and the syntax that jeopardises the strict distinction between them. In order to account for what seems to be contrastive nature of collocations, I propose some modifications within the minimalist program[2] to be able to account for collocations. The purpose of this paper is twofold: one to understand the syntactic as well as the semantic aspects of collocations. Secondly, to provide a common account upon which Chomskyan minimalist theory as well non-Chomskyan generative theories such as Jackendoff's parallel structure[3], Culicover's *Syntactic Nuts*[4] may agree on despite their varying technical details.

The second section shows the Arabic collocations data. The third section provides a basic background on collocations. I review how different theoretical frameworks address collocations. Primarily I discuss Jackendoff's parallel structure[5], Culicover's *Syntactic Nuts*, and Construction morphology. In the fourth section, the problems of the Chomskyan mainstream generative analysis of collocations are illustrated. In the fifth section, I explain how the minimalist analysis can account for the syntactic and the semantic aspects of collocations.

2. Data

I provide below some examples of Arabic collocations of two basic types: verb + object collocations and adjective + noun collocations:

(1) verb + object collocations

a. *ḏarafa dumū'* 'shed tears'

b. *ǧada'a 'anfahu* 'cut one's nose'

c. *šarama šafatahu* 'cut one's lip'

d. *barā alqalama* 'sharpen a pencil'

e. *ṭāra ṣawābahu* 'lose one's mind'

f. *naqaḍa al'ahd* 'break a promise'

g. *kaẓama alġayḍ* 'suppress anger'

h. *ḍaraba maṯal* 'give an example'

i. *'axaḏa alḥaḏar* 'exercise caution'

j. *sanna qānūnan* 'pass a law'

(2) adjective + noun collocations

a. *xayāl wāsi'* 'rich imagination'

b. *qahwat sawdā'* 'black coffee'

c. *Jumhūr 'arīḍ* 'wide public'

d. *ǧabal šāhiq* 'high mountain'

e. *qawl sadīd* 'right saying'

f. *ḥāfay alqadmīn* 'bare-footed'

g. *ġanam qāṣīyah* 'straying sheep'

h. *sāyl 'arim* 'raging flood'

i. *maṭar ġazīr* 'heavy rainfall'

j. *ḍabāb kaṯīf* 'heavy fog'

3. Basic Background

In this section, I discuss some definitions of collocations, their basic characteristics or features, their boundary, and finally I examine some theoretical frameworks that address collocations.

3.1 Basic definitions and features of collocations

The study of collocations was first introduced by the contextualise framework.[6] Firth defines the term as: 'You shall know a word by the company it keeps'[7]. Evert proposes that 'a collocation is a word combination whose semantic and/or syntactic properties cannot be fully predicted from those of its components, and which therefore has to be listed in a lexicon'[8]. Gelbuk and Kolesnikova suggest that collocations are expressions on which words co-occur with other words to convey a special unpredictable meaning[9]. McCkeown and Radev indicate that collocation are 'cover word pairs and phrases that are commonly used in language, but for which no general syntactic and semantic rules apply'[10].

Based on these definitions and others, collocations are associated with the following important properties:

1. Collocations are ready-made chunks. Not only do children learn individual words but also they memorize and produce groups of words as prefabricated units that are not a result of grammatical rules[11]. In first language research as exemplified by the work of Tomasello[12], learning collocations apply at an early stage of language acquisition. As Handl and Graf observe, the language speaker learns collocations early on in order to serve three important goals: first, collocations minimize the cognitive burden for speakers since this helps to reduce the processing work by means of joining words together as units

rather than to combine them each time[13]. Secondly, collocations serve pragmatic and communicative function enabling the speaker to master his or her language. Thirdly, collocations are important acquisitional tools helping the child learn the language.

2. Collocations are idiosyncratic. They are irregular because words co-occur with other words arbitrarily[14]; furthermore, collocations are associated with irregular syntactic and semantic features as Evert observes. As a result, they have to be listed in the lexicon given its unpredictability that may not accounted for by grammatical principles[15].
3. Collocations are recurrent. They are group of words used frequently[16].

3.2 Theoretical frameworks

There are many linguistic approaches that study collocation. Below I refer to some selected frameworks.

3.2.1 Contexualism approach

This framework assumes that context plays an important role in the study of language in general and collocations in particular. Firth was the first to suggest that collocations involve a lexical meaning at a syntagmatic level, which relates the meaning of a word with another word that surrounds it in a context. Firth argues that collocations' meaning is a result of co-occurrence relation among words[17]. Sinclair suggests that language uses two basic principles to build structures: the open-choice and idiom principles[18]. The first principle applies regular grammatical rules to produce syntactic units like a phrase or a clause. Sinclair calls this principle a 'slot-and-filler' model assuming that a large range of regular lexical choices fill the syntactic slots in accordance with grammatical rules[19]. The idiom principle however is used to construct restrictive "semi-preconstructed phrases" as single words. To Sinclair, the idiom principle accounts for the recurrence of collocations and reflects "a natural tendency to economy of efforts"[20].

3.2.2 Meaning-Text theory (MTT)

As mentioned above, collocations involve words co-occurring together unpredictably in order to convey a specific meaning. For example, we say *to carry out a survey* but not **to give a survey* with the meaning *to do or complete a survey*. MTT argues that lexical function (LF) explains the semantic and the structural relation of the constituents of the collocation[21]. To begin with, the meaning of the collocation is not derived directly from its components. Instead, there is a special meaning assigned and LF associates the lexical meaning to the collocation unit. To illustrate, a collocation structure consists of a base/ keyword (of LF) that is selected independently from the lexicon and a collocate or value (of LF) which is dependent on the base[22]. For example, *lend a hand* consists of a base/ keyword *hand* that is independently taken from the lexicon and a collocate *lend* that is semantically contingent or dependent on the base. LF can be represented mathematically as:

(3) LexicalFunction (keyword= value)

The collocational semantics is accounted for by different meanings represented by

different LFs. There are two kinds: Paradigmatic LF describes lexical relations as a result of morphological derivations (e.g. watch – watching) and semantic relations of synonymy and antonymy. Another type is syntagmatic LF that are used for lexical items in close proximity as collocations. There are more than sixty LFs of this type[23]. LF then captures the meaning of the collocation. MTT proposes different meanings that collocations may be associated with. For example, a collocation can have the LF **Oper** from Latin *operare* which means *do* or *perform* as in the collocation *to lend support*. The lexical function in accordance to (3) is represented as: **Oper** (support) = to lend. Other examples of LF are introduced in table (4).

(4) Examples of LFs (adopted from Gelbuk and Kolesnikova 2013: 26)

LF	Name description	Meaning	Keyword	Value	Collocation
Caus	From Lat., *causare*, to cause	to cause to exist	*elections*	*hold*	*to hold elections*
Real	From Lat., *realis*, real	to fulfil the requirement contained in the argument	*invitation*	*accept*	*to accept an invitation*
Fact	From Lat., *factum*, fact	to accomplish itself	*dream*	*come true*	*the dream came true*
Magan	from Lat., *magnus*, big	intense, intensely, very (intensifier)	*temperature*	*high*	*high temperature*

We have explained briefly some of the meanings of LF, LF also explains the structural relations of the base and the collocate of the collocation. To illustrate, the subscript 1 of the LF **Oper** is used for the subject when it is an agent. For example, *to lend support* has the LF: **Oper$_1$** (support) = to lend. However if the subject is the recipient of the action, it is represented by subscript 2 as in *to receive support.* The LF becomes **Oper$_2$** (support) = to receive.[24]

3.2.3 Jackendoff's Parallel structure[25]

Jackendoff argues against the Chomskyan generative grammar basic assumption that

syntax is the only generative component of grammar deriving phonology and semantics. Instead, Jackendoff believes that there are three independent parallel generative components of language: syntax, phonology, and semantics (conceptual structure). These structures have their own principles and they are all linked together via correspondence rules. One of the evidence Jackendoff provides to support his model of grammar is the account of free fixed expressions, i.e. lexical phrases like idioms and collocations. Mainstream generative theory does not address such expressions as they are considered marginal and not belonging to the core aspects of language but to the periphery of language since they cannot be constrained by grammatical rules[26].

In parallel structure, Jackendoff relinquishes the fundamental assumptions of mainstream theory from *Aspects*[27] up to the Minimalist Program (MP)[28]. Specifically, he assumes that fixed expressions are VPs that are lexically licensed. Unlike the MP and other Chomskyan theories, Jackendoff believes that such expressions are easily handled in his parallel structure theory. Namely, collocations and fixed expressions in general pose serious problems to mainstream theory, more specifically the MP[29]. To illustrate, collocations are multiword constructions contradicting the traditional rigid view of the lexicon consisting of single words. Moreover, traditional lexical insertion inserting lexical items in single syntactic slots cannot explain how collocations which are more than one word are inserted. Furthermore, lexical insertion or merger of words assumes meaning to be composed by compositionally merging the meaning of a word to another. However this cannot possibly work for collocation since their meaning is not determined unless they are grouped together as a unit[30]. The model that he uses adopts certain assumptions: This model is constraint-based rather than rule-based as that of the MP. As a result, collocations or other fixed expressions cannot be explained by grammatical rules but by means of violable constraints. For example, usually (regular) phrases are syntactically merged together but such constraint is weakened to allow another constraint to lexically form the collocation unit. As a result, the lexicon, as Jackendoff assumes[31], is redundant in which lexical entries are arranged according to regular principles; hence, idioms or collocations are considered to be lexical VPs. Syntactically, collocations are regular syntactic structures by which words are combined together just like regular phrases. However semantically, the meaning of collocations differ from regular phrases with compositional meanings. Even though collocations involve some kind of redundancy or predictability, it is not possible to translate such redundancy into grammatical rules[32]. With these assumptions in mind, Jackendoff argues that parallel structure and lexical licensing explain free expressions. For instance, a collocation as *commit suicide* is a lexical VP that is formed as a regular syntactic structure. Lexical licensing then unifies (Jackendoff's own term for *merge*[33]) the independently derived phonological, syntactic, and semantic structures of the VP *commit suicide* in the lexicon where correspondence rules linking the three different structures apply. Otherwise if one of these structures fails to be unified, the derivation crashes[34].

3.2.4 Culicover's *Syntactic Nuts*[35]

Culicover criticizes the sharp distinction that mainstream generative theory assumes between core grammar and the periphery. Core grammar pertains to regular rules that follow from UG while the periphery represents the irregular aspects of the language that do not follow the general principles of the language. Culicover presents some pieces of evidence that such distinction is hard to maintain. He provides what he calls *syntactic nuts* or non-canonical phrases about which native speakers (of English) have strong intuitions. These constructions have some degrees of generality with irreducible idiosyncrasy. Below I only provide one example of many syntactic nuts that Culicover discusses: endocentricity.

Endocentricity is one of the core aspects of X' theory. It states that every phrase has a head. For instance, a VP has a V head, NP a N head and so on. There are however exceptions to endocentricity. For example, there are non-DPs showing up in noun positions like the following examples[36]:

(5) a. [Under the bed] is a good place to hide.
b. [That the world is round] is obvious.
c. [Seeing] is believing.

Even though the phrases between brackets occur in DP/NP position, they are not DPs: *under the bed* is a PP, *that the world is round* is a CP (complementizer phrase), *seeing* is a VP. This is a violation of endocentricity because the subject position of the PP, CP, and VP is DP/NP requiring the head necessarily to be D/N. In order to solve this dilemma, we may assume that there is a D head taking these expressions as complements[37]:

(6) a. $[_{DP}$ D $[_{PP}$ Under the bed]]
b. $[_{DP}$ D $[_{CP}$That the world is round]]
c. $[_{NP}$ D $[_{VP}$ seeing]

The assumption of an empty D just to salvage the endocentricity principle invokes what Culicover calls *the Encoding Argument*[38]. For Culicover, knowing language is a direct result of the learnability theory by which a learner learns a language based on his exposure or experience with the linguistic data. If however the syntactic analysis focuses instead on the formal representation of the theory as a means of encoding the linguistic knowledge, the explanatory adequacy of the linguistic data is jeopardized in favor of a mere explanation of the formal structural representation of the theory. In other words, regardless of how the linguistic knowledge is encoded in terms of theoretical structural representation, this knowledge has to be accounted for based on the speaker's experience.

Culicover argues that endocentricity as other cases of *syntactic nuts* refers to a basic fact that a clear-cut distinction between the rule-based aspects of language and the irregular cannot be established. This applies to syntax and to the lexicon where free expressions have varying degrees of regularity. Culicover reconciles the two aspects of language, the general and the idiosyncratic, in a markedness theory. For example, endocentricity is the unmarked case (regular) for phrases while the marked (irregular) are exocentric phrases as those in example (5).

3.2.5 Construction morphology (CM)

CM is a word-based model of grammar in which the word is the minimal unit of analysis. CM is not a single theory but a group of wide range interconnected theories like the Cognitive Construction Grammar[39], Radical Construction Grammar (Croft 2001)[40] among others. CM attempts to provide a theory of how syntax, morphology, and the lexicon interact in the formation of words and phrases. The notion 'construction' plays an important role in CM. The construction is a pairing of form and meaning. The form consists of the phonological form and its morphosyntactic properties (e.g. category like N or V). As for the meaning, it includes semantic, pragmatic, and discourse information[41].

Based on the tripartite architecture of grammar (Jackendoff 2002a; 2007; Culicover and Jackendoff 2005), the construction, in CM, links three types of information: phonology, syntax, and semantics. Booij replaces the rules that Jackendoff uses to join the three components of grammar with 'schema'. That is, a schema pairs these three aspects of information together. For example, the word *player* and the compound *book shelf* are derived as a result of the following schematic representations[42]:

(7) a. $<[x]_{Vi}\ er]_{N\ j} \leftrightarrow [\text{Agent of SEM}_i]_j>$
b. $<[[x]_{Nk}\ [x]_{Ni}]_{Nj} \leftrightarrow [\text{SEM}_i \text{ with relation R to SEM}_k]_j>$

The angled brackets define the schematic representation. The double arrow marks the form and meaning relation. (7a) illustrates the schema for affixation of an affix to a base while (7b) shows the schema for a compound word. The variable X represents major categories (V, N, P, and A). The indexes $_{i, k, j}$ represent shared lexical properties of phonology, syntax, and semantics. Therefore, the base *play* in (7a) a verb with the index $_i$ is associated with the meaning represented by *SEM* with the same index. The whole word becomes a noun marked with the index $_j$ and is associated with the whole meaning specified with the same index $_j$. As for the compound in (7b), the head of the compound *shelf* is specified for the category N with the index $_i$ and is linked to the meaning *SEM* with the same index. The meaning of the head has a specific relation *R* to the meaning of the non-head *book* that is marked with the index $_k$. The derived meaning is associated with $_j$ that is co-indexed with the whole compound noun.

Syntactic nuts or non-canonical syntactic structures that Culicover discusses, collocations can be added to the list, are accounted for in CM by listing them in the lexicon with their meanings[43]. In doing so, CM does not distinguish between words and rules. In other words, lexical entries are fully specified words with their phonology and semantics; simultaneously there are rules that build on lexical words to generate more complex morphological structures. This contrasts with the basic idea of mainstream theory that structures generated by the grammar are not listed in the lexicon[44]. According to CM framework, there is no distinction between the grammar and the lexicon. There is a continuum of productivity ranging from non-productivity, productivity, and semiproductivity.

4. Chomskyan mainstream theory and collocations

I discuss the problems that collocations pose for Chomskyan mainstream theory. I

illustrate that the theory's basic assumptions are incapable of handling collocations adequately. Let us first examine the peculiar aspects of collocations and then determine if the mainstream theory can account for them.

4.1 Collocations

Collocations are associated with peculiar properties distinguishing them from regular phrases and free-word combinations. The first property is the limited compositionality. Unlike phrases or free word combinations (e.g. *work until morning*) whose meanings are compositionally derived from their components, the meaning of a collocation is not fully compositional. That is, the meaning of the collocation is not derived from the meaning of the base and the collocate but instead the meaning of the base and some additional meaning[45].

For example in *strong tea*, *strong* ads, as Manning and Schütze observe, the meaning of 'rich in some active agent'. *Strong* is different from the meaning of strength in *physical strength*[46].

The other property of collocations is non-substitutability in that one component cannot be substituted or replaced by another word. For example, we say *to make a mistake* and we may not replace *make* with *do*. Also we say *to carry out a survey* but not *to give a survey*. Therefore, collocations work as one lexical unit in which lexemes cannot be freely replaced.

Collocations behave differently from idioms in the semantic contribution of their internal constituents to the meaning of the entire structure[47]. Idioms are widely considered to be non-compositional since its components do not derive the meaning of the idiom, e.g. *kick the bucket* means *to die* and has nothing to do with either *kicking* or *the bucket*. However and despite the limited compositionality, the internal constituents contribute to the meaning of collocations[48]. Nunberg et al observe that idioms (phrasal idioms) and idiomatically combining expressions (collocations) contrast in terms of contribution of their internal components to the overall meaning[49]. Namely in idioms like *kick the bucket* and *saw logs* that means *sleep*, their constituents do not participate in the meaning of idiom. Instead, the whole VP idiom is assigned the meaning. However, as for idiomatic combining expressions such as *pull strings* meaning to exploit personal connections and *take advantage*, their components identify with the meaning of the whole idiom. In other words, the meaning of the idioms are distributed among their constituents.

Nunberg et al argue that the semantic compositionality of idiomatic expressions explains their syntactic productivity. In other words, Nunberg et al observe that idiomatic expression, and not phrasal idioms, are subject to general syntactic principles like: modification, passivization, topicalization and so on[50]. Now let us consider how Arabic collocation *darafa dumu* 'shed tears' work in these examples:

(8)

(1) . *ḏarafat alxansā'u dumū'an ġazīratan 'ala ṣaxar wa ma ḏarafataha 'ala 'awlādaha.*

shed-3sf xansa-nom tears-acc plentiful-acc over Saxar and neg shed-3sf-it over children-her

Xansa shed plentiful tears over Saxar and shed no tears over her children.

b. *aldumū'u, ḏarafataha alxansā'u 'ala ṣaxar.*

tears-nom shed-3sf Xansa over Saxar

Xansa shed tears over Saxar.

c. *aldumū'u ḏurifat 'ala ṣaxar.*

tears-nom were shed over ṣaxar

Tears were shed over Saxar.

d. *ḏarafat alxansā'u aldumū'a 'allati kānat ġazīratan.*

shed-3sf xansa-nom the tears-acc that were plentiful

Xansa shed tears which were plentiful.

These examples illustrate that part of the collocation is accessible by different syntactic operations. In example (8a), part of the collocation *dumu* is modified by the adjective *gazirατan* and even *dumu* is referred to by the pronoun *–ha* that is attached to the verb. In (b), *aldumu* is topicalized and in (c) it is passivized. Finally, a relative pronoun refers to *aldumu* in (d). Grimm argues that collocations in Arabic are closer to free word combinations than to idioms[51]. She supports her claim by using pronominalization, modification, and extraction of parts of the collocation. For all these reasons, we conclude that the parts of the collocation *ḏarafa dumu* derive the meaning of the whole collocation unit. The semantic analysability explains the syntactic transformations of the internal parts of the collocation.

Nonetheless not all collocations can be accessible to syntactic principles:

(9) a. **'aqaḍa almağrimu maḍği'a Sāmi alhādiʔ' wama 'aqaḍahu albardu.*

deprived-3sm the criminal-nom (of) sleep Sami peaceful and neg deprived-3sm-it the cold-nom

The criminal deprived Sami of peaceful sleep and the cold did not deprive him of sleep.

b. * *maḍği'ahu, 'aqaḍa almağrimu.*

his sleep deprived-3sm the criminal-nom

His sleep, the criminal deprived.

c.* *maḍa'ahu 'uqiḍa.*

his sleep was deprived of

His sleep was deprived.

d. **'aqaḍa almağrimu maḍği'a Sāmi ʔallaḏi kāna yahana'u bih.*

deprived-3sm the criminal-nom his sleep Sami that was enjoying in it

The criminal deprived Sami of his sleep that he was enjoying.

As we can observe in (9), the internal part of the collocation *aqada madgia* cannot be targeted by syntactic rules. *madgia* cannot be modified in (9a), and it cannot be referred to by a pronoun *–hu* attached to the verb *aqada*. Similarly, *madgia* cannot be topicalized in (9b) and passivized in (9c). Finally a relative pronoun may not be used to refer to *madgia* in (9d). Therefore we conclude that collocations do not have a consistent behaviour regarding the semantic analysability of their constituents. Some of them as in (8) contribute to the meaning of the collocation even though the meaning is not totally composed from its parts while others as (9) do not.

Similar to idioms, the internal parts of the collocation in (9) are not semantically analysable and thus they become inaccessible to syntactic rules as observed in (9).

4.2 Chomskyan Mainstream theory and collocations

Chomsky distinguishes between regular rule-governed core language and idiosyncratic periphery of language[52]. Chomsky believes that the focus should be on the core system since it instantiates the general properties of Universal Grammar (UG) principles whereas the periphery marks the exceptions of language in the lexicon and hence they should be marginalized. Therefore lexical phenomena like idioms, collocations, Culicover's syntactic nuts[53] and Jackendoff's *construction after construction* [54] are not subject to UG rules and therefore be ignored as mere exceptions. It is no wonder then that collocations for example received so little attention in the mainstream tradition despite its importance in language learning in general[55]and in particular corpus linguistics, natural language generation, parsing, and computation lexicography[56]. Since collocations are not viewed as non-productive and non-universal varying from one language to another, they are not addressed in the transformational-generative grammar[57]. There is a strict division between an irregular lexicon and a transparent grammar[58]. As it stands, the generative approach faces serious problems in handling collocation.

Let us examine how Chomsky's generative theory (in particular Government and Binding (GB) and the minimalist program (MP)) accounts for collocations. To begin with, Chomsky argues that language has two basic components: a lexicon and a computational (syntactic) system[59]. The lexicon is the repository of all lexical items and the irregular properties of language. The computational system selects the lexical items from the lexicon to generate syntactic structures by means of UG rules. One important problem that generative approach faces is how to account for the lexical insertion of multiword construction like idioms and collocations. Even though that the insertion of lexical item applies in syntax, there are however slight differences in how the GB and MP view lexical insertion of words. In GB, the lexical items are inserted at Deep-structure which is an underlying syntactic level. The lexical items are then placed in X° slots to make syntactic X-bar trees or phrases. As for the MP, lexical insertion takes place by merging lexical items in syntax forming a syntactic structure. Lexical insertion or merger of lexical items in syntax can explain how single words are inserted in syntax but not multiword structures[60]. To illustrate, idioms or collocations for example are XP constructions that are larger than a word; hence they are not placed in X° slots in order for lexical insertion or merger to operate on them. Therefore multiword constructions pose problems for lexical insertion. To solve this problem, Chomsky proposes that an idiom *kick the bucket* is replaceable by a verb in syntax by means of an 'idiom rule' assigning special semantic features to the verb which determines the meaning of the idiom at Logical Form[61]. Chomsky's analysis applies to other types of multiword expressions like collocations for example. Challenging Chomsky's proposal, Jackendoff refers to many problems[62]. One difficulty Jackendoff discusses is discontinuous idioms that Emonds[63] observed in which a direct object intervenes between the idiom's constituents. For example, in *take NP to task, take NP for granted*, assuming Chomsky's replacement rule, the idiom underlying form should look like *[$_V$ take to task] NP* where the direct object moves to the right of idiom. Jackendoff believes that such movement is forced and cannot be justified within the universal Grammar. Besides, the movement rule of the direct object implies that syntax access the internal part of idiom suggesting that we are dealing here with a phrase and not a single head verb replacing the idiom. In fact, Jackendoff provides idioms which are complete sentences: *that's the way the cookie crumbles, keep your shirt on, the jig is up*[64]. He argues that these types of idioms may not be at any circumstance inserted or replaced by V. Jackendoff concludes that in order to avoid the aforementioned problem it is best to analyse an idiom as a stored VP.

Another problem is the semantic interpretation of multiword expressions. Given that lexical insertion within the mainstream theory applies in syntax either by placing the lexical item in X° slots in D-structure or by merger of lexical items, the semantic interpretation follows compositionally in LF. In other words, after lexical items are inserted, the meaning of the syntactic phrase is compositionally determined by the constituents making up the structure. Therefore, lexical insertion and semantic interpretation are designed to address compositional regular phrases. But what about non-compositional phrases like idioms and collocations? How do they receive their non-compositional meaning by Logical Form? In fact, there is no satisfying answer in the mainstream theory[65]. These irregular multiword constructions do not conform to the Frege's compositionality principle stating that the meaning of a syntactic structure is achieved by adding the meaning of their constituents and is a result of syntactic rules[66]. The meaning of the idiom is not composed from the meaning of its parts and therefore this is problematic for lexical insertion or merger that builds the structure semantically from the meaning of its components[67]. As we discussed above, Chomsky (1981: 146 note 94) suggests an ad hoc 'idiom rule' for *kick the bucket* assigning special semantic features to the verb which determines the meaning of the idiom at Logical Form[68]. However such rule cannot explain or predict whether the constituents of the idiom contribute to the meaning of the collocation as a whole or not. As explained in 4.1, Nunberg et al distinguish two types of idiom: idiomatic combing expressions and idiomatic phrases in which the constituents of the former type identify with the meaning of the idiom unlike the case in the other type[69]. The evidence that Nunberg et al base their idiom distinction on come from the use of syntactic rules such as modification, topicalization, pronominalization, passivization of the internal constituent in the case of idiomatic combing idioms as in *take advantage* proving its semantic analysability unlike the case in idiom like *kick the bucket*. Likewise, the Arabic examples in (8) and (9) reinforce such distinction. Interestingly, the application of syntactic rules to idioms cast more doubt to the mainstream theory of rigidness of idioms and that the lexicon cannot be constrained by rules leading to its basic proposal of strict distinction between the lexicon and the grammar. A more practical assumption is to relax such distinction to allow a continuum of productivity where we expect to find within the lexicon idioms ranging from high productivity to less productivity.

In the reminder of this section, I briefly discuss one mechanism that was proposed in the mainstream theory to account for the combinatorial properties of lexical items. Subcategorization rules and selectional restrictions were introduced to specify the syntactic and semantic properties of a lexical item.

Subcategorization rules indicate the idiosyncratic complement and categories that are associated with a lexical item while the selectional restrictions specify the semantic restrictions on the arguments a lexical item takes as (10) below shows[70]:

(10) *Murder:* CATEGORIAL FEATURES: [+V, -N]

SUBCATEGORIZATION FRAME: [V, ____NP]

SELECTIONALRESTRICTIONS:<HUMAN——Human>

Unfortunately this system of rules cannot account for collocations because the collocational relation is determined by lexical constraints of lexical items co-occurring together which are not explained by subcategorization rules or selectional restrictions[71].

5. A modified MP analysis of collocations

Collocations have special syntactic-semantic combinatorial properties that are determined by lexical constraints and not by general syntactic rules. As explained above, lexical insertion or merger being used syntactically does not account for how collocations are built. I therefore propose some modifications in the MP to be able to handle collocations better. First I discuss the syntactic and semantic properties of collocations. Then I argue that collocations are full lexical phrases or XPs that are derived by means of a lexical merger operation.

5.1 Syntax and semantics of collocations

Collocations, unlike regular phrases, are structures in which words lexically select others to represent specific meaning. Therefore words co-occur together syntactically in order to convey special meanings. For example, *make* and *decision* co-occur as a viable collocation but not *make* and *walk*[72]. But what links the constituents of the collocation to its unique meaning? LF (Lexical function) is used as a tool that links the syntactic structure to the meaning of the collocation[73]. In other words, LF establishes the syntactic as well the semantic relation between the constituents of the collocation. For example, LF maps the base/ keyword *decision* in *make a decision* unto the value/ collocate *make* which is dependent semantically on the base[74]. Structurally the base *a decision* selects the collocate *make* but not *do* for instance. To provide some details on the meaning of the collocation, let us examine Seretan's following definition[75] that he adapts from Mel'čuk[76]:

(11) Let AB be a bipartite language expression, where A and B are lexical items of the language L, and let 'S' be the meaning of AB, 'A' the meaning of A, and 'B' the meaning of B. The expression AB is a collocation iff the following three conditions hold:

(i)'S'⊃'A' (the meaning of S contains the meaning of A);
(ii) A is selected by the speaker in a regular and non-restricted way;
(iii) B is not selected in a regular and non-restricted way, but depending on A and the meaning 'S' to express.

In this definition or semantic and structural rules of collocations, the base of the collocation say *coffee* in *black coffee* is selected by the speaker independently and regularly by the general rules of grammar. It is also used unrestrictively if it can be used alone without any other lexical item.

On the other hand, the collocate *black* is dependent on the base; thus it is irregularly and restrictively selected by the base. This restriction on the selection of collocate, determined by the base, is structural and semantic. As a result, *black* means 'without milk' that we get in association with *coffee*[77]. It is interesting to compare collocations to regular phrases in which both components of phrases are independently selected and are used regularly and unrestrictively by the rules of grammar.

Finally, it is worth distinguishing semantic analysability and semantic non-compositionality of the constituents of the collocation. The meaning of the collocation is non-compositional in the sense that it is not totally derived from the meaning of the components of the collocation in a regular and non-restrictive way. For example, *black coffee* is not fully composed from the meaning of both parts but rather *black* is semantically dependent on the base to have a unique meaning of 'coffee without milk'. Therefore even though both parts share the meaning of the collocation, the meaning is not fully compositional. The Arabic collocation *ḏarafa dumū*' 'shed tears' is similar to *black coffee* involving semantically analysable constituents as evidenced by the access of the rules of syntax in examples (8) and it is semantically non-compositional. However we may have collocation with non-analysable constituents as *'aqaḍa maḍği'a* 'deprive of sleep' as evidenced by non-accessibility of syntactic rules as shown in examples (9) and it is non-compositional. For instance, the collocation in (9) is non-compositional since the collocate is dependent on the base *maḍği'a* to establish the special meaning of the collocation.

5.2 Collocations as lexical XPs

In this final section, I argue that treating collocations as phrases derived lexically by means of a lexical merger avoids the problems imposed by the lexical insertion/ merge analysis assumed by the mainstream linguists. First, I illustrate that this analysis is not totally new but it finds its roots in the literature even though it has not been developed to reach its full potential. Then I show that the proposed analysis handles the syntactic as well as the semantic restrictions of collocations.

Hale and Keyser argue that verbs especially denominal verbs such as *shelve*, *saddle* are projected lexically as full maximal phrases or VPs[78]. Furthermore, denominal verbs are formed by syntactic rules like incorporation and Empty Category Principle (ECP). Namely, the noun moves to incorporate into a verb and leaves a trace back that is antecedent-governed in accordance to ECP. For instance, *shelve* is derived lexically from *put something on the shelf*. To illustrate, the lexical entry *put something on the shelf* involves a VP structure: with a specifier, head, and complement. The noun *shelf* moves to incorporate into the verb thus deriving the denominal verb. According to Hale and Keyser, the lexicon is syntactic or lexical-syntax (l-syntax) since it is constrained by syntactic rules and that the XP structure represents the argument information as well as the structural relations of the lexical head; simultaneously, the lexicon is lexical because the incorporation associated with denominal verbs for example does not affect all nouns. In other words, not every noun develops denominal verb but only some.

Despite Hale and Keyser's assumption of XP structure of lexical entries and the submission of the lexicon to syntactic rules, they are quick to suggest that the XP structure is not realised until D-structure where lexical insertion applies and the phrase is inserted in its syntactic slots in accordance with the conventional generative wisdom. Regardless of the level at which the structure is inserted, Hale and Keyser assume that all verbs involve a lexical phrase and that these verbs are like phrasal idioms in that their syntactic phrasal structures should be learned. They base their claim empirically on the observation that in many languages (e.g., Igbo) lexical items are phrasal.

Even though that Chomsky has assumed early on that the lexicon is a repository of irregularity of the language, he thinks that the lexicon can be constrained by rules generating regular complex words lexically merging regular inflections to roots as in *played*[79]. Therefore the lexicon does not include simple lexical items but lexical items merging with inflections deriving a complex structure. This complex word formation is not restricted to inflections but causatives, compound nouns, and noun incorporation (Chomsky 2015). We have already discussed denominal verbs as an example of noun incorporation (NI) (Hale and Keyser 1993); furthermore Al-Dobaian examines data of denominal verbs in Arabic and Hebrew and argue that they are lexically formed[80]. As for causatives, Al-Dobaian discusses deadjectival causatives of Arabic and Hebrew in which some adjective lexically merge with the verb head[81]. Rosen argues that NI is lexically derived by means of merger between the noun and the verb[82].

Based on this limited review of literature, we can observe that within the generative theory the lexicon dos not only list simple lexical items but also has redundant rules merging lexical items with other morphemes deriving phrasal or complex morphological structure. If this is the case, then the official periphery-core distinction assumed by mainstream tradition becomes at least shaky. The lexicon has its own rules and it is not just a repository of exceptions. Furthermore, the lexicon possesses some regularity aspects like building phrases that are subject to X-bar theory: a lexical phrase structure has its head and may have a complement or specifier. Also, syntactic rules modify the internal structure of some collocations as in (8).

Lexical merger is used as mentioned above to derive a lexical morphological complex words as explained with denominal verbs and causatives. Moreover, I argue that it can be also used to from a multiword constructions like collocations. The end result would be a lexical phrase. In other words, the MP would distinguish between types of XPs: An XP that is derived in the lexicon by means of lexical merger [83]. The other type is the regular XP formed syntactically by merger. Below I examine how lexical merger applies to collocations, the principles controlling such process, and the implications on MP and Chomsky's generative tradition in general.

As we have discussed above, the base of a collocation is selected independently and unrestrictively from the lexicon and it selects another word (a collocate) to form a special meaning. The collocate is dependent on the base for its selection; hence, it is restrictively chosen by the base to establish a specific meaning for the collocation[84].

Therefore, the base and collocate form a lexical unit or multiword structure that cannot be derived as a regular phrase in syntax in which its component words are selected independently from the lexicon and combined unrestrictively with compositional meaning derived from the phrase constituents[85]. In other words, collocations are formed as lexical XPs implying that language speakers acquire them as a combination of words. Indeed this is what the literature suggests. For example, Mel'čuk observes that collocations or *phrasemes* are not formed from regular phrase structures consisting of simple words but they have to be stored as a group of words in the lexicon and memorized[86]. Seretan observes that collocations come as ready-made chunks that the speaker of a language learns besides single lexical items[87]. Sinclair assumes that these chunks are 'semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments'[88]. The language speaker learns collocations early on in order to minimize the cognitive burden since this helps to reduce the processing work by means of joining words together as units than to combine them each time[89].

Now we are in a position to provide an analysis within the MP of how and why collocations are formed as lexical multiword structures or XPs. The lexical selection between the base and the collocate or their co-occurrence restriction may be viewed as a lexical merger creating a collocational unit. But why such merger applies in the first place?
Before answering this question, it is fruitful to compare lexical and syntactic mergers. Both types of merger create a phrase structure. However not all phrases are equal. On one hand, the merger applying in the lexicon is very restrictive. That is not all words can be lexically merged but only specific words co-occur with others in order to develop a unique meaning with which the collocation is associated. On the other hand, syntactic merger, controlled by syntactic rules, freely applies to words creating a phrase whose meaning is compositionally derived from its components. Collins argues that syntactic merger is motivated by an *Integration* condition that he defines below[90]:

(12) Every category (except the root) must be contained in another category.

According to the definition in (12), a constituent structure is contained by another constituent. For example, a verb *play* and its complement *football* are two constituents contained by *VP* and therefore they are *integrated* together to form a bigger constituent structure[91]. Merger builds larger sentence structure by means of a series of a successive binary procedure combing a pair of constituents[92]. Collins relates the condition of Integration to Kayne's Linear Correspondence Axiom (LCA) in which asymmetric c-command of terminal nodes is mapped into their linear order[93]. According to Collins, the integration of constituents into a clause allows their terminal nodes to have the right order with other nodes. If integration fails, then no linear order is possible among terminal nodes. For instance, the constituents *play* and football are integrated deriving a VP; thus the verb asymmetrically c-commands and consequently precedes its complement. Even though, Collins in an earlier version (1995) proposes that integration may possibly reduce to LCA, he no longer assumes this in his (1997) work. Collins does not derive Integration from LCA because he believes that Integration 'applies at every step of the derivation'[94]unlike LCA that applies after Spell-Out as Chomsky[95] argues.

Having examined the syntactic merger, I discuss lexical merger. I argue that the *Integration* condition motivates the lexical merger of collocations. The ultimate goal of merger is to create a bigger phrasal structure whether it is at the syntactic level or at the lexicon deriving multiword structures as in collocations. The integration follows from LCA since the constituents of the collocation must be linearly ordered. To see how LCA works, let us examine the linear order of the Arabic collocations: *darafa dumu* and *aqada madgia*:

(13) VP
V NP
darafa / *aqada* *dumu* / *madgia*

In the tree diagram, the verbs asymmetrically c-command their complements because only the verbs asymmetrically c-command the complements but not the other way around. The definitions of c-command and asymmetric c-command are given in (14):

(14) a. A c-commands B if neither A or B includes the other and every node (phrase) dominating A also dominates B[96].

(15) b. A asymmetrically c-commands B if and only if A c-commands B and B does not c-commands A[97].

As the definitions show, the verbs in (13) c-command their complements because according to (14a) the verbs and the complements do not dominate each other and the phrase VP dominates both of them. Therefore, the verbs asymmetrically c-command the complements since only the verbs c-command the complements but the complements do not dominate the verbs since the NP only dominates the complements and not the verbs as (14b) illustrates. As a consequence of the asymmetrical c-command, the verbs are linearly ordered before their complements. That is why the verbs precede their complements. So the LCA applies not just after Spell-Out as argued by Chomsky but also in the lexicon as it controls the linear order of the collocations' constituents[98].

The implications of this discussion is that merger is a general rule that works in the lexicon as well as the syntax. Merger is motivated by the Integration condition that is derived in turn by the LCA. Because merger creates phrasal structures, the constituents of the phrase should be arranged in terms of word order in accordance to the LCA. More importantly, the lexicon is not just a repository of irregularities of language but also a level equipped with its own rules. It is necessary to relinquish the strict division between a rigid lexicon and a rule-based regular syntax. Instead, the lexicon is viewed as a continuum with different ranges of regularity. Such view allows us to account for the learnability of collocations rather assuming them as mere exceptions of the language stored in the lexicon.

6. Conclusion

The MP does not address collocations mainly because they are considered to be part of the periphery and thus treated as exceptions. The MP as well as Chomskyan theories assume a strict division between an irregular lexicon and a rule-based grammar/ syntax. In order to account for collocations, this strict division must be relaxed so that the MP can share with other non- Chomskyan theories as Jackendoff's parallel structure among others the basic conceptions of collocations. To do so, I attempted to modify the MP by proposing that there are general rules applying in both the lexicon and the syntax. One of these rules is Merger that applies in the lexicon to create a phrase structure that speakers of language, besides learning single lexical items, learn these ready-made chunks. Merger is motivated by the LCA ensuring that the derived phrasal structure of the collocation has the proper linear order. Therefore, it is necessary to view the lexicon as a continuum where various ranges of regularity are present. So we find collocations with their semantically analysable constituents behaving like regular phrases and collocations whose constituents are not semantically analysable behaving just like opaque idioms.

---

[1] This research is supported by a grant from the Research Centre for the Humanities, Deanship of Scientific Research at King Saud University. I would like to thank King Saud University for this grant.

[2] Chomsky, N. *The Minimalist Program*.

[3] Jackendoff, R. *The Architecture of the Language Faculty* and *Foundations of Language*.

[4] Culicover, P, *Syntactic Nuts: Hard Cases in Syntax*.

[5] Jackendoff, R. *The Architecture of the Language Faculty* and *Foundations of Language*.

[6] For details, see Seretan, V. *Syntax-Based Collocation Extraction*.

[7] Firth J. R. *Papers in Linguistics*, pp. 179.

[8] Evert, S. *The statistics of word Cooccurrences*. pp. 17.

[9] Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*.

[10] McCkown K. and Radev, R. Collocations. pp. 507.

[11] Seretan, V. *Syntax-Based Collocation Extraction*.

[12] Tomasello, M. *Constructing a Language*.

[13] Handl, S. and Graf E.-M. Collocation, anchoring, and the mental lexicon - an ontogenetic perspective. pp. 119-147.

[14] See Seretan, V. *Syntax-Based Collocation Extraction* and McCkown K. and Radev, R. Collocations.

[15] Evert, S. *The statistics of word Cooccurrences*.

[16] See Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions* and Seretan, V. *Syntax-Based Collocation Extraction*.

[17] Gitaski, C. *The development of ESL collocational knowledge*.

[18] Sinclair, J. Corpus, Concordance, Collocation.

[19] *Ibid.* pp. 109.

[20] *Ibid.* pp. 110.

[21] Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*.

[22] Mel'čuk, I. Collocations and lexical functions.

[23] Seretan, V. *Syntax-Based Collocation Extraction*.

[24] Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*.

[25] See Jackendoff, R. *The Architecture of the Language Faculty* and *Foundations of Language*.

[26] Chomsky, N. *Knowledge of Language*. pp. 147.

[27] Chomsky, N. *Aspects of the Theory of Syntax*.

[28] Chomsky, N. *The Minimalist Program*.

[29] I discuss the problems of the analysis of the MP of free expressions in § 4. For now, I will just briefly mention these problems.

[30] Jackendoff, R. Construction after construction and its theoretical Challenges. pp. 8-28.

[31] Jackendoff, R. *The Architecture of the Language Faculty*.

[32] Jackendoff, R. Construction after construction and its theoretical Challenges. pp. 8-28.

[33] Jackendoff prefers to use this term over *merge* since *unify* is a more inclusive general term covering wide-ranging cognitive relations. For more details see Jackendoff, Alternative Minimalist visions of Language. pp. 189-226.
[34] Jackendoff, R. Construction after construction and its theoretical Challenges. pp. 8-28.
[35] Culicover, P, *Syntactic Nuts: Hard Cases in Syntax*.
[36] *Ibid*. pp. 96.
[37] *Ibid*. pp. 96
[38] *Ibid*. pp. 67
[39] See for example, Goldberg, A. *Constructions at Work*.
[40] See Croft, *Radical Construction Grammar*.
[41] Booij. G. *Construction Morphology*.
[42] Examples in 7 (a, b) are taken from Booij. G. *Construction Morphology*. pp. 3 and 7.
[43] Culicover, P, *Syntactic Nuts: Hard Cases in Syntax*.
[44] Booij. G. *Construction Morphology*.
[45] Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*. pp. 24.
[46] Manning, C. and Scütze, H. *Foundations of Statistical Natural Language Processing*. pp. 151.
[47] Given the limited space, I will not provide detailed account of idioms in this paper. I just mention the basic aspects of idioms that prove to be fruitful in the discussion of collocations.
[48] McCkown K. and Radev, R. Collocations. pp. 507.
[49] Nunberg, G. et al Idioms. pp. 491-538.

[50] For more details, see *Ibid*.
[51] Grimm, P. S. Collocation in Modern Standard Arabic revisited. pp. 22-41.
[52] Chomsky, N. *The Minimalist Program* and Chomsky, N. *Aspects of the Theory of Syntax*.
[53] Culicover, P, *Syntactic Nuts: Hard Cases in Syntax*.
[54] Jackendoff observes that NPN construction like *face to face* and *dollar for dollar* requires two nouns to be identical separated by five prepositions: upon, to, for, after, and by. He argues that even though such construction is lexical, it is constrained and productive. This presents a problem for the Chomskyan assumption that lexical phenomenon is arbitrary and irregular. For details, see Jackendoff, Construction after construction and its theoretical Challenges. pp. 8-28.
[55] Handl, S. and Graf E.-M. Collocation, anchoring, and the mental lexicon - an ontogenetic perspective. pp. 119-147.
[56] Manning, C. and Scütze, H. *Foundations of Statistical Natural Language Processing*. pp. 151.
[57] Gries, S. T. Phraseology and linguistic theory: a brief survey.
[58] Jackendoff, R. Compounding in the Parallel Structure and Conceptual Semantics. pp. 105-128.
[59] Chomsky, N. *The Minimalist Program*.
[60] Jackendoff, R. *The Architecture of the Language Faculty* and *Foundations of Language*.
[61] Chomsky, Lectures *on Government and Binding*. pp. 146 (note 94).
[62] Jackendoff, R. *The Architecture of the Language Faculty* and *Foundations of Language*.
[63] Emonds, J. Emonds, Joseph. 1970. Root and Structure-preserving Transformations.
[64] Jackendoff, R. *The Architecture of the Language Faculty* and *Foundations of Language*. pp. 160.

[65] *Ibid*.
[66] Vegge, T. *Idioms: Categorization, lexical representation and the question of compositionality* and Jackendoff, R. Alternative Minimalist visions of Language.

[67] Jackendoff, R. *The Architecture of the Language Faculty* and Alternative Minimalist visions of Language.
[68] Chomsky, N. *Lectures on Government and Binding*. pp. 146 (note 94).
[69] Nunberg, G. et al Idioms. pp. 491-538.
[70] Radford, A. *Transformational Syntax. A First course*. pp. 372.
[71] Bartsch, S. *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Cooccurrence*.
[72] Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*.
[73] Mel'čuk, I. Collocations and lexical functions.
[74] Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*.
[75] Seretan, V. *Syntax-Based Collocation Extraction*. pp. 19.
[76] Mel'čuk, I. Collocations: définition, rôle et utilité. pp. 23–32.
[77] Seretan, V. *Syntax-Based Collocation Extraction*.
[78] Hale, K. and Keyser, S. J. On the argument structure and the lexical expression of grammatical relations. pp. 53-110.

[79] Chomsky (2015) discusses two possibilities of deriving complex words: one involves lexical merger and another syntactic merger. The former merges the root and the inflection in the lexicon and then checking applies in syntax. The other option is to draw a lexical item from the lexicon and then merge it with the affix syntactically.
[80] Al-Dobaian, A. On the Semitic Denominal Verbs: the case of Arabic and Hebrew, pp. 65-83.
[81] Al-Dobaian, A. *Semitic causatives and inchoatives: Their implication to the syntax-morphology interface and aspectuality*.
[82] Rosen, S. Two Types of Noun Incorporation: A Lexical Analysis, pp. 294-317.
[83] Lexical merger deriving a lexical XP can also be applied to account for Culicover's syntactic nuts (1999), and Jackendoff's *construction after construction* (2008). I assume that merger is a general condition that applies in the lexicon as well as syntax. However its nature varies depending on the level it belongs to. Lexical merger, for example, randomly combines two morphemes/ words together for specific structural and semantic reasons that I explain below. As for syntactic merger, we expect words to freely combine based on the rules of syntax, e.g. X-bar structure.
[84] Mel'čuk, I. Collocations: définition, rôle et utilité. pp. 23–32.
[85] Mel'čuk, I. Collocations and lexical functions.
[86] *Ibid.*
[87] Seretan, V. *Syntax-Based Collocation Extraction*.
[88] Sinclair, J. Corpus, Concordance, Collocation. pp.110.
[89] Handl, S. and Graf E.-M. Collocation, anchoring, and the mental lexicon - an ontogenetic perspective. pp. 119-147.
[90] Collins, C. *Local Economy*. pp.66.
[91] Collins assumes that merger can be motivated to satisfy two different types of properties: Integration and feature checking in the case of head movement. I shall only focus on merger for Integration property that does not involve feature checking. For more details, check (Collins 1997).
[92] Radford, A. *Minimalist Syntax. Exploring the Structure of English*.
[93] Kayne, R. *The Antisymmetry of Syntax*.
[94] Collins, C. *Local Economy*. pp.68.
[95] Chomsky, N. Bare Phrase Structure.
[96] Li, Y. *X°: A theory of the morphology-syntax interface*. pp. 160.
[97] Kayne, R. *The Antisymmetry of Syntax*. pp. 4.
[98] Chomsky, N. Bare Phrase Structure.

References

Al-Dobaian, A. *Semitic causatives and inchoatives: Their implication to the syntax-morphology interface and aspectuality*. Madison: University of Wisconsin Ph.D. dissertation. 2002.

Al-Dobaian, A. On the Semitic Denominal Verbs: the case of Arabic and Hebrew. In: *Journal of King Saud University*. Riyadh: King Saud University, 18(2): 65-83. 2006.

Bartsch, S. *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Cooccurrence*, Gunter Narr Verlag: Tübingen. 2004.

Booij. G. *Construction Morphology*. Oxford: Oxford University Press. 2010.

Booij. G. Construction Morphology. In: Hippisley, A. et al. (eds.): *The Cambridge Handbook of Morphology*. Cambridge: Cambridge University Press. 2015.

Chomsky, N. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. 1965.

Chomsky, N. *Lectures on Government and Binding*. Dordrecht: Foris. 1981.

Chomsky, N. *Knowledge of Language*. New York: Prager. 1986.

Chomsky, N. Bare Phrase Structure. In: Webelhuth, G. (ed.): *MIT Occasional Papers in Linguistics 5*. Also in *Government and Binding Theory and the Minimalist Program*, Oxford: Blackwell, (1995).

Chomsky, N. *The Minimalist Program*. *20th Anniversary Edition*. Cambridge, MA: MIT Press. 2015.

Collins, C. *Local Economy*. Cambridge, MA: MIT Press. 1997.

Culicover, P. *Syntactic Nuts: Hard Cases in Syntax*. Oxford: Oxford University Press. 1999.

Culicover, P. and Jackendoff, R. *Simpler Syntax*. Oxford: Oxford University Press. 2005.

Evert, S. *The statistics of word Cooccurrences: Word pairs and collocations*. PhD thesis, University of Stuttgart. 2004b.

Gelbuck, A. and Kolesnikova, O. *Analysis of Verbal Collocations with Lexical Functions*. Berlin: Springer. 2013.

Gitaski, C. *The development of ESL collocational knowledge*. PhD thesis, University of Queensland. 1996.

Goldberg, A. *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press. 2006.

Gries, S. T. Phraseology and linguistic theory: a brief survey. In: GRANGER, S. (eds.): *Phraseology: An Interdisciplinary Perspective*, Amsterdam: John Benjamin. 2008.

Grimm, P. S. Collocation in Modern Standard Arabic revisited. In: *Journal of Arabic* Linguistics, Heidelberg, Germany: Heidelberg University, 51: 22-41. 2009.

Hale, K. and Keyser, S. J. On the argument structure and the lexical expression of grammatical relations. In: *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, MA: MIT Press, 53-110. 1993.

Handl, S. and Graf E.-M. Collocation, anchoring, and the mental lexicon - an ontogenetic perspective. In SCHMID, H.-J. (eds): *Cognitive Foundations of Linguistic Usage Patterns*, Berlin: De Gruyter Mouton, 119-147. 2010.

Jackendoff, R. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press. 1997a.

Jackendoff, R. *Foundations of Language*. Oxford: Oxford University Press. 2002.

Jackendoff, R. Alternative Minimalist visions of Language. In: *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 41(2): 189-226. 2005.

Jackendoff, R. Construction after construction and its theoretical Challenges. In: *Language*. 84(1): 8-28. 2008.

Jackendoff, R. Compounding in the Parallel Structure and Conceptual Semantics. In: LIEBER, R. et al (eds): *The Oxford Handbook of Compounding*, Oxford: Oxford University Press, 105-128. 2009.

Kayne, R. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press. 1994.

Li, Y. *X°: A theory of the morphology-syntax interface*. Cambridge, MA: MIT Press. 2005.

Manning, C. and Scütze, H. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA. 1999.

McCkown K. and Radev, R. Collocations. In DALE, R. et al. (eds): *A Handbook of Natural Language Processing*, Marcel Dekker, New York, NY, 507–523. 2000.

Mel'čuk, I. Collocations and lexical functions. In Cowie A P (ed.): *Phraseology. Theory, Analysis, and Applications*, Claredon Press, Oxford, 23–53. 1988.

Mel'čuk, I. Collocations: définition, rôle et utilité. In: GROSSMAN, F. et al.: *Les collocations: analyse et traitement*. Editions De Werelt, Amsterdam. Grossmann F, 23–32. 2003.

Mel'čuk, IPhraselogy: Its place in the Language, in the Dictionary, and in Natural Language Processing. In GAVRIILIDOU, A. et al (eds.): *Selected papers of the 10th ICGL*, Komotini/Greece: Democritus University of Thrace, 62-77. 2012.

Nunberg, G., SAG, I., and WASOW, T. Idioms. In: *Language*, Vol. 70(3): 491-538. 1994.

Radford, A. *Transformational Syntax. A First course.* Cambridge: Cambridge University Press. 1988.

Radford, A. *Minimalist Syntax. Exploring the Structure of English*. Cambridge: Cambridge University Press. 2009.

Rosen, S. Two Types of Noun Incorporation: A Lexical Analysis. In: *Language*, 65: 294-317. 1989.

Seretan, V. *Syntax-Based Collocation Extraction*. Berlin: Springer. 2011.

Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.

Tomasello, M. *Constructing a Language*. Cambridge, MA: Harvard University Press. 2003.

Vegge, T. *Idioms: Categorization, lexical representation and the question of compositionality*. PhD thesis, University of Agder. 2012.