



## Phraseology: Where Lexicon and Syntax Conjoin

David Singleton<sup>\*a</sup>, Justyna Leśniewska<sup>b</sup>

---

### Article Info

*DOI:*

*Article History:*

Received: 10 Dec. 2020

Accepted: 21 Dec. 2020

Published: 1 Feb. 2021

*Keywords:*

Lexicon and grammar,  
Lexis and syntax,  
Lexicogrammar  
Idiom principle  
Phraseology

*Article Type:*

ORIGINAL

### Abstract

This article discusses the definitions and scope of phraseology as a domain which is primarily lexical but which very much connects with grammatical issues. We discuss the gradual widening of the scope of phraseology, resulting from the emergence of data-driven, bottom-up analyses of large language corpora. We also note that statistical models of learning and usage-based models of language use converge to provide theoretical support for the claim that language use is to a considerable extent formulaic. An overview of empirical research which supports this claim is provided. We also examine the various schools of linguistic thought which blur the distinction between lexis and grammar, such as the Hallidayan concept of lexicogrammar, valency grammar, pattern grammar, and construction grammar. Our conclusion is that while these and other developments in the field of syntax acknowledge the role of lexis and even attach great importance to it, academic publications which cover typically lexical issues, especially with respect to second language acquisition and use, rarely acknowledge the crucial interconnection between lexis and grammar.

---

<sup>\*a</sup> Corresponding author: Prof., Trinity College, Dublin, Ireland. dsnglton@tcd.ie.

<sup>b</sup> Assoc. Prof., Jagiellonian University, Krakow, Poland.

## Introduction

Some forty years ago Paul Meara (1980) complained that second language vocabulary acquisition was neglected by research. Some two decades later, he had been cheered by progress on this front (Meara, 2002), recognizing that the topic had attained a position of some importance, although he worried that most vocabulary research in applied linguistics was based on a rather narrow linguistic agenda. The areas he mentioned as having had little impact on second language lexical research, psycholinguistics and computational linguistics, have more recently been much more to the fore in this domain. The main point, however, and one not especially salient in Meara's account, is that the vital connections between the lexicon and grammar, in first and second language, the basic oneness of the lexicon and grammar, continue to be underplayed in the area of lexical research. In fact, most of the movement towards recognizing the essential unity of lexicon and grammar has come from the syntacticians. From a very wide range of theoretical positions, syntacticians have in fact increasingly come to regard the lexicon as integral to the workings of syntax. At the same time, research on lexis, though no longer in any way scarce, in fact rather abundant, seems to be far less concerned with grammar.

There is of course an area of study which is primarily lexical but where the lexicon and syntax certainly connect and overlap: the domain of phraseology, defined by Gledhill as follows:

Broadly speaking, phraseology involves the study of formulaic sequences of words, including idiomatic phrases and proverbial expressions, which stand in contrast to other more prosaic constructions in the language in that they have a highly conventionalised form and frame of reference. (Gledhill, 2011, p. 2)

In fact, in general diction, this term has a broader, more inclusive meaning than Gledhill's definition suggests, as the Oxford English Dictionary (OED) confirms:

The selection or arrangement of words and phrases in the expression of ideas; manner or style of expression; the particular language, terminology, or diction which characterizes a writer, work, subject, language, place, etc.

The OED definitions of phraseology capture the fact that it needs to be conceived of fluidly, characterizing *all* processes underlying the "arrangement of words", even when the outcome is not self-evidently formulaic.

We shall return later to a closer examination of the notion that the way in which we "arrange words" is essentially phraseological in nature. Let us first probe a little more deeply into the relationship between lexicon and grammar.

### The Lexicon And Grammar

It is clear that when people think of language they think of words (see Singleton, 2016, pp. 1-12). This is one of the implications of the association between the notions of "word" and "speech" to be found in many languages - for example, the dual meaning "word"/"speech" attaching to French *parole*, Italian *parola* and Spanish *palabra*. A similar phenomenon can be noted in the terminological structure of the academic specialism of grammar. Thus, grammarians have traditionally seen their specialism as having two branches - syntax and morphology - in both cases the very definition of terms being lexically based. The term *syntax*, from Greek *syntaxis* ("putting together in order"), denotes the range of regularities which can be observed in the combination of sentence components (and the study of such regularities), these components being largely identifiable as words and groups of words. The term *morphology* derives from Greek root *morphé* ("form", "shape"), and denotes the internal structure of words (and the study thereof).

Although the word is central to the way in which non-specialists and the specialists alike think about language, defining what a word is raises many problems (cf. Carter, 2012, pp. 19-46; Lyons, 1968, pp. 400-442). Interestingly, the characterization of the word that seems to be least problematic is that which defines words in grammatical terms. The grammatical approach uses the criteria of "positional mobility" and "internal stability".

Words are said to be “positionally mobile” in the sense that they are not fixed to specific places in a sentence. The “internally stable” criterion refers to the fact that within words the order of morphemes remains consistent. The grammatical approach in this connection is not only the least problematic but also the one that works best across languages. This fact already speaks volumes about the connection between the lexicon and grammar.

In the grammatical discussion of the “London School” (see e.g., Halliday, 1961, 1994) and in valency grammar (see e.g., Allerton, 1982, 1995) the recognition of the lexicon-grammar overlap was present from the outset, but even in the work of Chomsky, widely thought of as the most syntactic of syntacticians, the lexicon became increasingly central as his model developed. While in the earliest version of the Chomskyan model (Chomsky, 1957) the lexicon was not recognized as an autonomous component, words being considered to be merely the elements through which syntax manifested itself, the subsequent evolution of Chomsky’s model consistently ascribed more and more importance to the lexicon (Cook, 1988; Cook & Newson, 2007). By the early 1980s (Chomsky, 1981), the lexicon was seen as having a crucial influence on syntactic structure. The so-called “Projection Principle” of the “Government and Binding” (GB) version of Chomskyan syntax states that the properties of lexical entries “project onto” the syntax of the sentence. The lexicalizing tendency in Chomskyan linguistics reached its zenith in the “Minimalist Program” (Chomsky, 1995), where Chomsky accepted (alongside many other schools of grammarians), that the particular lexical elements selected in any given sentence are the determinants of both the content and the form of the sentence.

Halliday’s notion of “lexicogrammar” is in ways akin to Chomsky’s 1995 conception, but it is in ways far more radical. Halliday was greatly influenced by Firth’s work, notably by his view (Firth, 1957a, 1957b, 1968) that “no aspect of lexis or grammar can be properly defined without reference to its typical context of use (or ‘co-text’) that is to say in actual stretches of texts or discourse” (Gledhill, 2011, p. 6). Halliday’s lexicogrammatical perspective incorporates Firth’s central tenet that that language signs are mutually dependent on and defined by those other signs with which they are habitually used:

Words must not be treated as if they had isolate meaning and occurred and could be used in free distribution. (Firth, 1957a, p. 18)

The collocation of a word or a ‘piece’ is not to be regarded as mere juxtaposition, it is an order of mutual expectancy. The words are mutually expectant and mutually prehended. (Firth, 1968, p. 181)

We now turn tack to the question of phraseology – the origins of the term denoting it, the history of its study, and its growing role in our understanding of the way in which language functions.

### **What is Phraseology?**

The word phraseology has a lot of ‘etymological appeal’, as it is neatly composed of the Greek term *phrasis* (“way of speaking”) and the well-known suffix *-logia* (“study of”). However, the meaning of this term is much less neat. As was mentioned earlier, the meaning of the term in general diction diverges somewhat from its meaning in linguistics. Worse still, the linguistic definitions of phraseology are also varied and at times vague. Granger and Paquot attribute this to “the highly variable and wide-ranging scope of the field and the vast and confusing terminology associated with it” (2008, p. 27). To understand why this is the case, one must look at the origins and history of the term.

The honour of being phraseology’s founding father is sometimes bestowed on the Swiss linguist Charles Bally (1865-1947), who came up with the notions of fixed and loose phraseological units (*unités phraséologiques* and *séries phraséologiques*). However, it was in Soviet Russia from the 1940s onwards that phraseology was developed into an important branch of linguistic enquiry by, among others, Viktor Vinogradov, Natalya Amosova and Igor Mel’čuk (Cowie, 1998). It also gained considerable popularity in East Germany and other countries of the communist bloc, and was eventually introduced to English-speaking scholars in the 1960s

(Cowie, 1998). Cowie points out that in the English-speaking world, major research projects focusing on phraseology first appeared only in the early 1990s, with phraseology being seen as a field of both “pure” and “applied” research.

The school of linguistic thought developed in the Eastern bloc is usually referred to today as “traditional” phraseology. Proponents of this school seek to present the various types of word co-occurrence phenomena as being on a continuum ranging from the relatively free to the most restricted. While different terms were used to describe these categories (a problem further compounded by translation issues), as Cowie (1998) observes, whatever the terminology, the taxonomies used by all authors were based on distinguishing between “word-like” and “sentence-like” combinations. Furthermore, “word-like” combinations would roughly always fall into one of three categories: 1) opaque, invariable combinations 2) partially idiomatic combinations 3) phraseologically bound combinations, i.e., items which are collocationally restricted only to some extent. The truly free combinations, the ones that have “only syntactic and semantic restrictions” (Granger & Paquot, 2008, p. 29) were seen as falling outside the realm of phraseology.

Howarth (1996) illustrates this taxonomy with examples of lexical (verb + noun) and grammatical (preposition + noun) combinations at the various levels of idiomaticity:

free combinations	restricted collocations	figurative idioms	pure idioms
<i>blow a trumpet</i>	<i>blow a fuse</i>	<i>blow your own trumpet</i>	<i>blow the gaff</i>
<i>under the table</i>	<i>under attack</i>	<i>under the microscope</i>	<i>under the weather</i>

Traditional phraseology is closely connected to lexicography. Phraseological investigation into the English language helped drive home the fact that much of the language that we deploy and encounter in everyday speech and writing clearly consists of combinations of words that we have available as more or less prefabricated chunks, and made it possible to document those chunks in the form of dictionaries. An analysis of the authentic data processed in preparation for the *Oxford Dictionary of Current Idiomatic English* (Cowie, Mackin & McCaig, 1975/1983), for example, yielded literally thousands of such stable multi-word units (see Cowie, 1988). Similarly, the *Oxford Dictionary of Phrasal Verbs* (Cowie & Mackin, 1993) and the *Oxford Dictionary of English Idioms* (Ayto, 2010) between them contain some 15,000 multi-word expressions. Our reliance on “ready-made” material, rather than taking the trouble to concoct combinations of elements on every occasion, probably relates to the fact that similar situations recur in life and tend to be referred to in similar ways.

An important pioneering application of traditional phraseological theory to the study of second language was made by Howarth (1996, 1998), who looked at what types of word combinations are the most problematic for language learners, and established that it was the ‘restricted collocations’ that posed most problems.

It is important to note that within this traditional approach, word combinations were distinguished on the basis of the linguist’s judgment. An entirely new perspective was provided by the bottom-up, data-driven approaches to the analysis of language which were enabled by the advent of large, computerised corpora. As will be shown below, a new understanding of the term “phraseology” emerged as a result.

### **Sinclair and His Phraseological Aftermath**

We know from our everyday experience that particular words are especially frequently to be found in the company of certain other words, depending on the context. Thus, a given selected word is likely to have a habitual potential “entourage”. This phenomenon of words “keeping company” together is referred to as *collocation*. Words which form collocations are repeatedly “placed with” each other; that is to say, they often co-occur within a given distance of each other in speech and in written texts. Collocation was an important topic for the neo-Firthians, and this school of thought was continued by Sinclair (1991), who initiated research based on very large samples of authentic texts, which revealed that collocational patterns were omnipresent. Many word combinations which would not necessarily be noticed by a linguist working within the traditional phraseological approach, because they were relatively transparent and syntactically regular, emerged from such analyses.

Sinclair is thus credited with the change of emphasis in linguistics from paradigmatic choice to syntagmatic constraints in linear sequences (Stubbs, 2009). The general importance of the contribution of collocational knowledge to linguistic competence has been discussed by, among others, Benson (1985) and Kjellmer (1991).

Patterns which emerged from actual language data suggested that many compositional, regular word combinations may in fact be pre-constructed to some extent, in other words, may be formulaic. The scope of phraseology was thus considerably extended – to apply to any patterning of words in language. This wider understanding of phraseology is usually referred to as the “distributional” or “frequency-based” approach (Granger & Paquot, 2008).

The existence of collocational patterning has to do with the fact that human beings prefer to save effort whenever possible. Also, it is certainly connected with the huge demands made on us by the extreme rapidity of speech production, which are such that we *have to* exploit every opportunity to make savings on processing time. This notion was given prominence by Sinclair, who developed the idea of economy of effort via prefabrication into the so-called “idiom principle”. He used the term *idiom* with a much wider application than as a label for fixed expressions with meanings that cannot be deduced from the meanings of their component parts. The *idiom principle*, as defined by Sinclair (1991, pp. 110-115) states that, when we are putting together phrases in a language we know, although it may look as if we were operating on the basis of open choices at every stage, the way we operate most of the time is to draw on our knowledge of pre-constructed or semi-preconstructed phrases, varying lexical content within the relative environments to a fairly limited extent.

Thus, in Sinclair’s *Collins COBUILD English Grammar* (3rd edition, 2011) there is discussion (pp. 160-161) of the relationship between single words and related multi-word combinations, such as:

*arrange – make an arrangement*  
*confess – make a confession*  
*suggest – make a suggestion*

Clearly, in processing terms, cases where the multi-word synonym is used result from integrated, unitary choices similar to cases where the single word is deployed. Sinclair concludes his 1991 book as follows:

While grammars and dictionaries continue to report the structure of language as if it could be neatly divided, many of those people who are professionally engaged in handling language have known in their bones that the division into grammar and vocabulary obscures a very central area of meaningful organization. (Sinclair, 1991, pp. 137)

Sinclair argues (*ibid.*) that “when we have thoroughly pursued the patterns of co-occurrence of linguistic choices there will be little or no need for a separate residual grammar or lexicon”. The implication is that the choices in question lie principally in the area of collocational/phraseological patterning.

Indeed, psycholinguistic evidence suggests that fixed expressions and formulas have an important economizing role in speech production (see, e.g., Peters, 1983); that is to say, they enable us to produce speech which is very markedly more fluent than it would be if we had to build up piece by piece every expression we employ (see Nekrasova, 2009; Pawley & Syder, 1983). Conklin and Schmitt (2012, p. 45) point out that:

It makes sense that our brains would make use of a relatively abundant resource (long-term memory) to compensate for a relative lack in another (working memory) by storing frequently occurring formulaic sequences. These could then be easily retrieved and used without the need to compose them online through word selection and grammatical sequencing.

It should be noted at this point that the idiom principle, and related models, connect an external reality (the patterning of words in language) to assumptions about how the human mind processes language. Myles and Cordier caution against the potential confusion of speaker-internal and speaker-external aspects of formulaicity, observing that many researchers “take as their basis what usually happens in the language surrounding the speaker, extrapolating that this preferential status has consequences for the storage of these sequences in the speakers of that language” (Myles & Cordier, 2016, p. 4). While it is true that it should not be taken for granted that the formulaicity observable in language has a specific psycholinguistic reality, there is support from both theoretical and empirical studies for this being the case. A comprehensive review of those studies is beyond the scope of this paper, but the most important points will be covered below.

### **Support for Formulaicity / The Psycholinguistic Reality Behind Phraseological Phenomena**

There is a substantial body of empirical findings which support formulaic processing. First of all, idioms are known to be processed faster by native speakers than comparable non-idiomatic strings (see Conklin & Schmitt, 2012, for a review). Even more interestingly, faster processing has also been found in the case of word combinations which are formulaic (i.e., which form frequently used strings), but are not otherwise idiomatic. One study (Bod, 2000, as cited in Bod, Hay & Jannedy, 2003) found that participants responded faster to compositional, three-word subject-verb-object sequences when the sequences were frequent (e.g., *I like it* generated a faster response than, for example, *I keep it*); similar results were obtained by Arnon and Snider (2010) with four-word phrases. Similar implications arise from a study by Tremblay et al. (2011), in which native speakers read passages containing lexical bundles (i.e., frequently occurring strings) and comparable passages with ordinary word combinations, with the latter being read more slowly. Tremblay and Baayen (2010) found that the corpus frequencies of phrases correlate with the native-speaker participants’ recall times for those phrases, while Sosa and MacFarlane (2002) obtained similar results when testing native-speakers’ reaction times to phrases with different frequencies.

Another source of support for the existence of formulaic processing can be found in studies involving patients with brain damage (e.g., Van Lancker-Sidtis & Postman 2006; Van Lancker-Sidtis 2012), which suggest a different representation of familiar phrases in the brain from that of non-formulaic language, and point to the right hemisphere as the place where formulaic language is processed, while indicating that non-formulaic language is handled by the left hemisphere.

Finally, research on the prosody of formulaic language also has something to contribute to the discussion on the formulaic character of language processing. Stretches of language which are formulaic may be characterised by specific prosodic features, speed, stress patters, alignment of pauses, etc. (Kuiper, 2004; for discussion, see Lin, 2012). In a study by Hallin and Van Lancker Sidtis (2017), speakers of Swedish read out sentences with proverbs and comparable sentences without proverbs, and were found to use different speeds and tonal patterns for the former than for the latter, which was taken by the authors as an indication of the holistic storage of formulaic sequences. The authors also propose a tentative hypothesis concerning how formulaic language might be processed in neurolinguistic terms, based on the observation that patients with damage to a particular area of the brain - the basal ganglia – lose their command of formulaic language. Since the basal ganglia play a key role in subcortical motor systems, the authors suggest that formulaic language may be produced using holistically stored motoric gestures, which would explain the gains in speed of processing.

The above studies provide strong support for the existence of the “idiom principle” – the existence of a formulaic mode of processing language. Very importantly, this formulaic mode is unlikely to be restricted to the items studied within the traditional approach to phraseology – idioms, set phrases, and restricted collocations. A formulaic mode of processing also seems to exist for word combinations which are transparent, idiomatically unmarked, but frequent. At the same time, it is worth noticing that frequency as such is not the sole predictor of formulaicity, as some of the items most likely to be stored holistically – for example, idioms – are rare in terms of their overall frequency of occurrence in language.

Frequency plays an even more central role in some theoretical models of language acquisition and processing, which also provide a potential explanation for the existence of phraseological phenomena. The postulated formulaic nature of language processing finds theoretical support in probabilistic models of language acquisition and processing, which explain how language is acquired through sensitivity to patterns in the input and their frequencies (see Chater and Manning, 2006, for a review). These models are compatible with connectionist approaches to language processing (Rumelhart & McClelland, 1986), which “see knowledge in terms of soft connection strength rather than rules or patterns” (Singleton, 1999, p. 123).

A very interesting question concerns the extent to which language – in speaker-external terms – is formulaic. At the moment, there is hardly a consensus on the matter, and the answer depends on the particular method used to classify the formulaic stretches. Conklin and Schmitt (2012) provide an average based on a number of studies with different methodologies, and arrive at the figure of between one third and one half of discourse. An interesting study by Forsyth and Grabowski (2015) explores the possibility of automatically measuring the extent to which a text is formulaic by using (or adapting) various existing indices.

In speaker-internal terms, if the language user, as Sinclair suggested, alternates between relying on the idiom principle and the open-choice principle, one could posit a similar question: to what extent do speakers rely on either principle? Is one more dominant than the other? According to Wray, “formulaic processing is the default,” and “construction out of, and reduction into, smaller units by rule occurs only as necessary” (Wray, 2002, p. 119). The definitive solution to this question remains unknown, but it is possible that the directions of research outlined above will eventually provide some answers.

### Phraseology and Grammar

The classical approach to phraseology, because of its clearly delineated boundaries, was clearly dealing with lexical issues. However, the distributional approach, with its much broader approach, enters the territory usually demarcated as the domain of syntax. The links between phraseology and grammar have been explored in a number of recent articles (e.g., Dobrovolskij, 2017; Zerlina and Kostina, 2015). The trend of such discussion tends towards undermining the traditional position that grammar is autonomous and prime.

Some researchers have posed the question whether “grammar” is in fact simply an outgrowth of collocation/formulaicity/phraseology. For instance, Hoey (2005) suggests that the structure of language actually *derives from* lexical patterns.

...as a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. (Hoey, 2005, p. 8)

This radical perspective is elucidated thus:

Each of us, presumably to different extents and with different outcomes, constructs a grammar – leaky, inconsistent, incomplete – out of the primings we have for the sounds, words, phrases that we encounter. This grammar, or perhaps one should say these grammars, may in turn be used to regulate ... our linguistic choices. (Hoey, 2007, p. 31)

This line of thinking is in tune with recent understanding of the way in which second language grammar is established, on which Lantolf and Thorne have the following to say.

From the perspective of emergent grammar ... learning an additional language is about enhancing one's repertoire of fragments and patterns that enables participation in a wider array of communicative activities. It is not about building up a complete and perfect grammar in order to produce well-formed sentences. (Lantolf & Thorne, 2006, p. 17)

Such a view ties in with the probabilistic models of language acquisition, mentioned above, and with the view that the primary language learning mechanism is statistical in nature. Within this framework, the process of learning is seen as "a gradual process of accumulating linguistic knowledge based on the distributional properties of the input" (Andringa & Rebuschat, 2015, p. 188). Patterns of grammatical structure emerge from the frequency information which human beings accumulate without being conscious of this process (Rebuschat & Williams, 2012).

The whole usage-based approach to language acquisition has similar things to say:

Language knowledge involves statistical knowledge, so humans learn more easily and process more fluently high-frequency forms and 'regular' patterns which are exemplified by many types and which have few competitors. Usage-based perspectives of acquisition thus hold that language learning is the implicit associative learning of representations that reflect the probabilities of form-function mapping. (Ellis, 2017, p. 130)

On this view too language "rules" derive from structural regularities emerging from unconscious analysis of distributional characteristics of language input. Usage-based models (e.g., Bybee & Hopper, 2001; Goldberg, 2006; Langacker, 1987, 2000; MacWhinney, 2006; Tomasello, 2003) stress the fact that learning is inductive, data-driven, and enabled by experience. As language users come into contact with samples of language, certain form-meaning mappings become gradually automatic for them, or "entrenched" (Langacker, 1987). Entrenchment is thus obviously related to the frequency of forms in the input. The question which presents itself in this context is that of the extent to which what we have been taught to think of as "grammatical rules" - i.e., morphosyntactic patterns and usage - are related to the phenomenon of habitual multi-word co-occurrence. We also have to consider, of course, the implications of this relationship.

The concepts mentioned above – the findings concerning the relationship between frequency and formulaicity, as well as the idea that what we see as the "rules" of language actually derive from the unconscious analysis of distributional characteristics of language input – tie in with some approaches to language description which blur the distinction between grammar and lexis. Two developments are particularly worth mentioning in this context.

Pattern grammar (Hunston & Francis, 2000) emerged from research carried out as part of the Collins COBUILD project (see Sinclair 1991). It is described as "an approach to language which maintains the generalising characteristics of grammatical descriptions while prioritising the behaviour of individual lexical items" (Hunston, 2002, p. 167). An example of a pattern is the behaviour of a verb in terms of being followed by finite or non-finite clauses, to-infinitive clauses, '-ing' clauses, etc. Words come with their "attendant phraseology", which is the grammar pattern associated with a given word. This approach puts lexis at the centre of attention, because the phraseology of a given word – the patterns it appears in – is seen as one of the many elements of lexical knowledge: "knowing a word means, among other things, knowing the patterns a word has" (Hunston, 2002, p. 176). This approach to language description is very useful in terms of second language instruction, as it has obvious pedagogical applications.

Construction grammar (Goldberg, 2006; Tomasello, 2003) is a group of approaches to grammar which are based on the idea that arbitrary pairings of form and meaning occur at every level of linguistic description, including syntax. Constructions are exactly that – they are units which pair a linguistic form with a meaning. All constructions are part of a lexical-syntactic continuum, and they may be of different levels of abstractness. Constructions are acquired as the result of repeated experiences, which means that frequency is again seen as

playing an important role (see e.g., Gries & Ellis, 2015). Constructions which are encountered more often become represented in the lexicon with stronger connections.

A core feature of construction grammar is that it “militates against .... The dictionary-and-grammar model” (Hilpert, 2014, p. 3). The connection with phraseology is evident in the following quote from a construction grammar textbook (the author uses the term “idiomatic” in the more general meaning of “formulaic”):

... ordinary language is fully permeated by a large number of idiomatic expressions whose forms and meanings are not entirely predictable on the basis of wither the word meanings recorded in a dictionary or the rules of syntax provided by a grammar. (Hilpert, 2014, p. 4)

For example, expressions such as “That’s quite useful a lesson” or “too big a shock” are treated as idiomatic – because they are syntactically unusual, that is, they deviate from the canonical patterns of grammar – yet they allow any lexical item that meets certain requirements to occur in them. Understood this way, idiomatic expressions are not fixed; rather, they are strings with some systematicity, “schemas with slots that can be filled with some elements but not others” (Hilpert, 2014, p. 6).

### Discussion and Conclusion

As we can see from the above review, phraseology has done a great deal to provide an angle on the study of lexis which is also inclusive of grammar. Moreover, the interconnectedness of grammar and lexicon has long had a place in linguistic theory – as evidenced by the concept of lexicogrammar, and the approaches propounded by valency grammar, pattern grammar, and construction grammar. Alas, this inseparability of grammar and lexicon is, as we have mentioned repeatedly, to a large extent ignored by researchers working on vocabulary issues in applied linguistics. Many treatments of the lexicon in relation to second language teaching and learning focus on “bean counting” – the frequency of individual words – and steer totally clear of syntactic and phraseological matters. For example, in the (otherwise in many ways excellent) book on vocabulary in the curriculum recently edited by Dodigovic and Agustin-Llach (2020) the items *grammar*, *syntax* and *phraseology* do not even appear in the index! Many sources provide information on what it means to know a word (see e.g., Nation 1990; Richards, 1976) – but such lists, even if they do include collocations, do so only in the sense of immediate, lexical collocations; for example, *torrential rain*, where it is part of the knowledge of the word *torrential* that it collocates with *rain*.

In fact, second language pedagogy itself is an area where the division between lexis and grammar is very firmly delineated, or rather, etched in stone. A brief look at language textbooks reveals material divided into sections on grammar, vocabulary, and idiomatic expressions. It is important to note that the latter are treated as the “weird stuff” of language, the exceptions to the otherwise neat workings of a system of grammatical rules operating on individual lexical elements. Even the praiseworthy inclusion of collocations in some language textbooks does not make any dents in the grammar-lexis barrier, as they are mostly isolated typical cases of adjective-noun, verb-noun and adverb-adjective combinations. In the same vein, pedagogical grammars present productive patterns, while learners’ dictionaries contain the building blocks, with the odd idioms dictionary thrown into the mix. This is a situation which is ripe – indeed over-ripe – for radical change!

## References

- Allerton, D. J. (1982). *Valency and the English verb*. Academic Press.
- Allerton, D. J. (1995). Valency Grammar. In E.F.K. Koerner and R.E. Asher (Eds.), *Concise history of the language sciences* (pp. 280-289). Kidlington.
- Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning. *Studies in Second Language Acquisition*, 37(02), 185–196.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67–82.
- Ayto, J. (Ed.). (2010). *Oxford dictionary of English idioms*. Oxford: Oxford University Press.
- Benson, M. (1985). Collocations and idioms. In R. Ilson (Ed.), *Dictionaries, lexicography and language learning* (pp. 61-68). Oxford: Pergamon/The British Council.
- Bod, R. (2000). *The storage vs. computation of three-word sentences* [Paper presentation]. Architectures and Mechanisms in Language Processing Conference 2000 (AMLaP'00), Leiden, The Netherlands. <http://staff.science.uva.nl/~rens/amlap00.psi>.
- Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. MA: MIT Press.
- Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: J Benj.
- Carter, R. (2012). *Vocabulary: Applied linguistic perspectives* (2nd ed.). Routledge.
- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, 10(7), 335–344.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The Minimalist Program*. MIT Press.
- Conklin, K. and Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32(1), 45–61.
- Cook, V. (1988). *Chomsky's Universal Grammar: An introduction* (1<sup>st</sup> ed.). Blackwell.
- Cook, V. and Newson, M. (2007). *Chomsky's Universal Grammar: An introduction* (3rd ed.). Blackwell.
- Cowie, A. (1988). Stable and creative aspects of vocabulary use. In R. Carter and M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 126-139). Longman.
- Cowie, A. and Mackin, R. (Eds.). (1993). *Oxford dictionary of phrasal verbs* (revised ed.). Oxford: Oxford University Press.
- Cowie, A. P. (1998). Introduction. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 1-20). Oxford: Clarendon Press.
- Cowie, A., Mackin, R. & McCaig, I. (Eds.) (1983). *Oxford dictionary of current idiomatic English* (Vols.1-2). Oxford: Oxford University Press. (Original work published 1975)
- Dobrovoskij, D. O. (2017). Grammar and phraseology: Cross points. *The Bulletin of the Russian Academy of Sciences: Studies in Literature and Language*, 76(1), 5-14.
- Dodigovic, M. and Agustin-Llach, M. (2020). *Vocabulary in curriculum planning: Needs, strategies and tools*. Cham: Palgrave Macmillan.

- Ellis, N.C. (2017). Chunking in language usage, learning and change: *I don't know*. In M. Hundt, S. Mollin and S.E. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives* (pp. 113-147). Cambridge: Cambridge University Press.
- Firth, J. R. (1957a). *Papers in linguistics 1934-1951*. Oxford: Oxford University Press.
- Firth, J. R. (1957b). A synopsis of linguistic theory, 1930-1955. In *Studies in Linguistic Analysis, Special Volume of the Philological Society* (pp. 1-31). Oxford: Blackwell.
- Firth, J.R. (1968). A synopsis of linguistic theory, 1930-1955. In F.R. Palmer (Ed.), *Selected papers of J. R. Firth 1952-1959* (168-205). London: Longman.
- Forsyth, R.S. and Grabowski, Ł. (2015). Is there a formula for formulaic language? *Poznań Studies in Contemporary Linguistics*, 51(4), 511–549.
- Gledhill, C. (2011). The 'lexicogrammar' approach to analysing phraseology and collocation in EP texts. *La Revue de GERAS: ASP*, 59, 5-23.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Granger, S. and Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27-49). Amsterdam: John Benjamins.
- Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(S1), 228-255.
- Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 17, 241-292.
- Halliday, M. A. K. (1994). *Functional grammar* (2<sup>nd</sup> ed.). London: Edward Arnold.
- Hallin, A. E., & Van Lancker Sidtis, D. (2017). A closer look at formulaic language: Prosodic characteristics of Swedish proverbs. *Applied Linguistics*, 38(1), 68-89.
- Hilpert, M. (2014). *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Hoey, M. (2007). Grammatical creativity: a corpus perspective. In M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (Eds.), *Text, discourse and corpora: Theory and analysis* (pp. 31-56). London: Continuum.
- Howarth, P. (1996). *Phraseology in English Academic Writing: Some implications for language learning and dictionary making*. Tübingen: Niemeyer.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics* 19, 24-44.
- Hunston, S. (2002). Pattern grammar, language teaching, and linguistic variation: Applications of a corpus-driven grammar. In R. Reppen and S. M. Fitzmaurice (Eds.), *Using Corpora to Explore Linguistic Variation* (pp 167-183). Amsterdam: John Benjamins.
- Hunston, S. and Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Kjellmer, G. (1991). A mint of phrases. In K. Aijmer and B. Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman, 111-127.
- Kuiper, K. (2004). Formulaic performance in conventionalized varieties of speech. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 37–54). Amsterdam: John Benjamins.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. Stanford: Stanford University Press.
- Langacker, R. W. (2000). *A dynamic usage-based model*. Stanford: CSLI.

- Lantolf, J. and Thorne, S. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- Lin, P. M. S. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics*, 33(3), 342–347.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2006). Emergentism: use often and with care. *Applied Linguistics*, 27(4), 729-740.
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching* 13 (3-4), 221-246.
- Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research* 18 (4), 393-407.
- Myles, F., & Cordier, C. (2016). Formulaic sequence (FS) cannot be an umbrella term in SLA. *Studies in Second Language Acquisition*, 39(1), 1-26.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning*, 59(3), 647-686.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). Longman.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Rebuschat, P., & Williams, J. N. (2012). *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77-89.
- Rumelhart, D., & McClelland, J. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (Ed.). (2011). *Collins COBUILD English Grammar* (3<sup>rd</sup> ed.). London: Collins.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Singleton, D. (2016). *Language and the lexicon: An introduction* (republished ed.). London: Routledge.
- Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language*, 83(2), 227-236.
- Stubbs, M. (2009). The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics*, 30(1), 115-137.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: Continuum International.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613.
- Van Lancker Sidtis, D. (2012). Formulaic language and language disorders. *Annual Review of Applied Linguistics*, 32, 62-80.
- Van Lancker-Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and

right-hemisphere damaged subjects. *Aphasiology*, 20, 411–426.

Wray, A. (2002). Formulaic language in computer-supported communication: Theory meets reality. *Language Awareness*, 11(2), 114-131.

Zerkina, N. & Kostina, N. (2015). English Phraseology in teaching: Interrelation of theory and practice. [\*Procedia - Social and Behavioral Sciences\*](#), 199(3), 143-148.

**David Singleton** is Emeritus Fellow at Trinity College Dublin and Emeritus Professor at the University of Pannonia. He has served as Secretary General of AILA and as President of EUROSLA. In 2015 he received the EUROSLA Distinguished Scholar Award and in 2017 was awarded Honorary Life Membership of AILA.

**Justyna Leśniewska** is Associate Professor at the English Department of the Jagiellonian University in Kraków. Her research interests are related to the phraseological aspects of second language acquisition, the collocational competence of EFL learners, and the role of cross-linguistic influence in language acquisition and use.