

New perspectives on vocabulary: Formulaic sequences in the structuring of advanced learners' written production

Magdalena Zinkgraf, Silvina Rodríguez, Analía Castro y Ma. Angélica Verdú¹

Introduction

The study of native language corpora has shown that language presents repeated strings of words across texts and contexts, which might indicate that much more of language is patterned around clusters of words than was originally thought. The concept of the lexicon as made up of words in isolation is slowly giving way to a different perspective which has as its units recurrent sequences of words, some of which performing as discourse organizers. In this article we analyse a corpus of 237 academic essays (114,514 words) written by Spanish-speaking advanced learners of English at Comahue University in order to characterize their use of certain discourse-organizing formulaic sequences. First, an overview of the major changes in perspective is offered with respect to depth of knowledge of vocabulary, the nature of the lexicon and the role of lexis in language. Then, the learner corpus of formulaic sequences (FSs) is described in terms of the context, the corpus description and the selection of FSs. Results are presented for each of the selected FSs and compared to occurrences of those same strings in the British National Corpus. Finally, in the light of this study's findings some suggestions for the teaching of vocabulary are provided that will contribute to learners' acquisition of large chunks of language.

1. Changing paradigms: what is new about vocabulary learning

1.1. What knowing a word implies

The study of vocabulary and its acquisition in L2 contexts has traditionally centred on the lexical item. Knowing a word used to imply basically having knowledge of the word's form, its meaning and its pronunciation. In time, linguists became aware of a set of other features of lexical items that enriched speakers' knowledge. In the 80s, Nation and Meara shed light on new aspects of the knowledge of a word that should be mastered by learners apart from its meaning. In time, knowing a word came to include being familiar not only with its pronunciation, spelling and perhaps translation, but also with features like collocation (which adjectives, nouns and verbs typically co-occur with it), and connotation, among others (Nation 2001; Meara 1983, 1992; Schmitt 2000).

If we take the verb to *provoke* as an example, the dictionary definition for its Spanish counterpart may confirm our hypotheses as to its behaviour in English: its meaning reflects the same as in Spanish, "if you provoke someone, you deliberately annoy them and try to make them behave aggressively"². Through this **definition** we may gather information as to the pattern in which this verb is used: you provoke someone, as in (1).

(1) The lawyer claimed that his client was provoked into acts of violence.

¹ This paper stems from the research project J023 "Secuencias Formulaicas y su Adquisición en Estudiantes Universitarios de ILE", directed by M. Zinkgräf and subsidised by the Secretaría de Ciencia y Técnica (UNCo).

² *Cambridge Advanced Learner's Dictionary (CALD)* (2008)

The dictionary goes on to explain there is a **second meaning** of this verb, “to cause something to happen”. An **example** such as (2) enlightens us further:

(2) His election success has provoked a shocked reaction.

This sentence shows that a) the verb takes a complement, i.e. it is transitive; b) the complement it takes is inanimate and c), as a few more examples of how *provoke* is used are analysed, we will realize that the complement typically has a negative connotation (typical collocates it prefers are *reaction, chaos, a scandal, anger*, etc.). This fact points to an intrinsic characteristic of the noun phrase that accompanies the verb. But, more importantly, it reveals a type of link that is determined by the very selection of *provoke*. The verb itself implies something negative as part of its meaning and use, in this case realized in the adjective that modifies *reaction*, and, therefore, it mostly selects complements with a negative connotation. This type of semantic constraint is what Sinclair (2004) refers to as **semantic prosody**.

Each of the meanings of *provoke* above determines the way in which the verb will behave grammatically. In the first case (see (1) above), we observe how an animate noun (his client) is provoked by another animate noun (which is not overtly expressed) INTO something negative, which implies violence. Even though this prepositional phrase is optional in a number of variations of the pattern, what is essential to this sense of the verb is the feature of animateness of the complement (his client).

The second meaning evident in (2) occurs in a totally different grammatical environment: both the subject and the complement are inanimate, and the complement has a negative connotation. These two different patternings of the language around the verb *provoke* reflect what Hoey (2005) refers to as **grammatical colligation**. Words or sequences of words are surrounded by typical grammatical environments which are intimately connected to the meanings of the word/s analysed. The scope of this (grammatical) colligation is not restricted to grammar patterns, but also applies to functions and positions in sentences and in discourse at large.

Learners’ awareness and knowledge of all the aspects that knowing a word involves will help them produce idiomatic, and more native-like, versions of the intended meanings. Each of these aspects cannot be measured in an all-or-nothing manner (Nation 2001; Schmitt and Carter 2004: 4). This is exactly what the term **depth of vocabulary knowledge** transmits: knowing a word or sequence of words involves different types of knowledge which will be mastered to a certain degree on the cline of that particular aspect. However, none of them should be disregarded when teaching vocabulary, an issue to which we shall turn in later sections of this paper.

1.2. Vocabulary in textual context: recent findings

Since the 1990’s, research into foreign language vocabulary acquisition and, thus, the teaching of vocabulary, has shifted away from a focus on individual items towards a wider conception of what vocabulary implies and what units it comprises. In 1991, after carrying out research on his native speaker corpus, Sinclair introduced the two principles that changed linguists’ perspectives as

regards the lexicon, its contents and the process of language production: The Idiom Principle and the Open Choice Principle. In his view, the combination of these two principles accounts for most language use.

On the one hand, the *Idiom Principle* establishes that “a language user has available to him /her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair 1991: 110). In other words, a lot more of the language we produce as native speakers is constrained: language is composed of combinations or strings of words that depend on speakers’ selection of one lexical item at a certain point in the process of language production, which in turn carries with it the selection of a sequence of words, meanings and forms (due to its collocation, colligation and semantic prosody) that are not open for choice to the speaker. The *Open Choice Principle*, on the other hand, postulates that “at each point where a unit is completed (a word or a phrase or a clause), a large number of choices opens up and the only restraint is grammaticalness” (Sinclair 1991: 109). The balance between what is constrained and what is free is very delicate and, even if subject to restrictions, language production is creative, though probably not as much as linguists used to believe.

The insights obtained through corpus analysis studies in the 90s have contributed to this significant shift in paradigms in the realm of teaching a foreign language: whereas in the past, linguistic research oriented at the teaching and learning of vocabulary had centred around the word, applied linguists slowly became aware of the need to broaden the scope of their analysis to the behaviour and characterization of strings of words, especially those typically occurring together in similar contexts in authentic language use. Moreover, while language was supposed to be made up of grammatical categories that were filled in with lexical items, corpus linguistics and its outburst in the 90s provided sufficient evidence of recurrent patterns that had been overlooked until then. This has revealed that much of language is lexically determined and phraseological or pre-fabricated and, so, if words in text behave as clusters at times, this probably reflects how they are stored in our minds. This conception has an impact on how they should best be taught and learnt.

1.3. On the formulaic nature of language

Different studies into corpus analysis and the way words are learnt and stored have offered some insights into what language is really like and what the lexicon is exactly comprised of (Weinert 2010). What follows is a summarised account of the discoveries made in this field.

- **Words do not occur alone**

Some function words like determiners, auxiliary verbs and even prepositions never occur alone in text and have dependent elements that are required for them to be used in a grammatically correct way. But content words also take these types of company words. Take for instance the adjective *afraid*. If asked to complete a sentence like (3), nobody would hesitate to fill the gaps with the verb *to be* before it and the preposition *of* after it.

(3) I _____ afraid _____ the dark.

This shows that we unconsciously have the knowledge that this is one of the typical uses of this adjective.

- **Typical words appear in typical linguistic contexts/ genres and text types**

In order to establish typicality of use of *afraid*, it is necessary to obtain information from a native language corpus like the BNC (British National Corpus)³ Most of the concordance lines show that the phrase 'I'm afraid' is very frequent in sentences like (4). As can be deduced from its punctuation, it usually appears in spoken discourse.

(4) AT7 713 'He'll soon have to go, I'm afraid.'

In fact, there are far more instances of this spoken use of 'I'm afraid' (with the meaning of '*I think*') than of the one exemplified in (3). It is quite unlikely to find this phrase in an argumentative essay, which evinces the types of restrictions of use that apply with respect to text types.

- **Combinations of words are typically used to convey one meaning and these are preferred over a random and creative word-by-word selection**

To convey the idea that someone is very much afraid and that, due to this fear, they cannot react to something, we could use the words: He + is+ so + afraid + that-clause. But a more frequent and idiomatic option from the BNC is, in fact, (5), where an adjective with a stronger meaning (*scared*) is modified by another one (*stiff*) to express the same in a more emphatic and economic way. Among native speakers this version will be more often found than the construction above.

(5) A61 1648 The young lady behind the bar was scared stiff by our presence.

It becomes obvious that cross-linguistic issues affect the language production process to such an extent that unawareness of the existence of these idiomatic phrases – perhaps due to lack of exposure to them - may lead to unnatural combinations that may not be understood when interacting with a native speaker.

- **Underlying recurrent sequences of words lie grammatical patterns**

A careful analysis of the surrounding context for recurrent strings of words like *under no circumstances* will show that grammatical patterns that are equally recurrent accompany these sequences. If we study the concordance lines from the BNC below ((6) to (8)), the significant use of modals stands out. This should come as no surprise as the very meaning of the phrase is related to prohibition. Results from searches in the BNC indicate an overwhelming majority of the modal *should*, which clearly emphasises the meaning of the whole sequence. Besides, together with these modals appears the grammatical inversion that is related to the initial position of this adverbial.

³ British National Corpus, "a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written.", available at <http://www.natcorp.ox.ac.uk/>.

- (6) A22 113 Under no circumstances **will** any of those competitors who have been chosen be told they cannot go.
- (7) A61 1817 That is, under no circumstances **must** I mention to anyone that I have turned down an invitation to return to Blighty.
- (8) C96 877 Under no circumstances **should** you add methylene blue to your tank.

One cannot ignore the fact that a certain degree of constraint also operates within the phrase itself: there is no room for grammatical variation for there exist no examples of a singular version of *circumstance*. This grammatical information in relation to this sequence is what has been referred to above as colligation (Hoey 2005).

- **Similar meanings are conveyed through specific sequences and grammatical patterns**

Whenever native speakers want to put across the idea that a process, situation or action is not allowed to take place, sentences like the following will be recurrently found:

- (9) J6P 1096 No rule of law or the profession, therefore, prevents him from entering into a contract by which he undertakes not to act for a specified class of person.
- (10) CN2 1163 This was to stop people from withholding information from the inspector's investigation in order to present it at review, with the intention of discrediting the inspector's report.
- (11) ABE 1770 This could discourage people from driving ten miles to buy their groceries.
- (12) A22 74 I suspect they would not have deterred people from travelling without tickets.'

In (9) to (12) from the BNC, it becomes obvious that the same pattern recurs: each of these verbs takes an agentive complement (*him, people*) and a prepositional phrase introduced by *FROM*: no other preposition is acceptable. All these cases convey the same meaning, perhaps each with their special shades and it is through this pattern that the intended message is transmitted. This feature is widely present in language use and, therefore, it deserves teachers' and learners' attention.

- **Native speakers have an unconscious perception for these combinations**

Hoey (2005) postulates that, in speakers' minds, words have what he terms 'lexical priming', which will link each of the lexical items selected in language production to their context and co-text. This priming occurs as a result of multiple encounters throughout their lives with those same words in varying contexts and situations. Words in our minds resemble slots that need to be filled with multiple sources and types of information. Every new encounter with a word or string of words provides these slots with data related to some of the aspects implied in knowing it (its collocations, colligation and semantic prosody). The slots are then loaded with new information and links among the items are created and reinforced. This seems to result in a network of words and strings of words that is revised and constantly updated. In L1 speakers' mental lexicons these words are stored in connection to one another and the connections established among them are so strong that they cannot possibly find other less idiomatic ways of putting across their meanings. If some slots are not frequently revisited, temporal retention of words' use and combinations in memory might become latent.

This priming occurs naturally for native speakers of a language but in the case of foreign language learning, the situation is quite different: in EFL contexts, the degree of exposure to the target language is much more reduced than, of course, in a mother-tongue setting. This will obviously have a bearing on the way words are primed for these learners. Even though language learners may be concerned about their knowledge being 'incomplete', Hoey holds the view that it is precisely the interaction between native and target language that enriches our linguistic experience. By consciously or unconsciously comparing knowledge about a word in the L1 with evidence of occurrences of the L2 equivalent, speakers of an FL might find this information useful to guide their linguistic choices in both languages.

1.4. Formulaic Sequences

Throughout the literature, researchers have coined a series of terms to refer to these strings of words so recurrently found in corpora: prefabricated patterns and routines (Nattinger and DeCarrico 1992), lexical bundles (Biber et.al 1999) and multi-word units - MWUs (Moon 1998), among many others. In an attempt to "establish the larger pattern into which they all fit", Wray (2002: 9) proposes 'formulaic sequence' (FS), which is the term that has been adopted in this paper, to refer to

"a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p. 9).

The central role that these sequences play in communication is evinced in the vast repertoire of memorized language sequences that native speakers of a language make use of (Ellis 2008: 6). However, it is a fact that in the case of non-native language learners, serious difficulties arise in the acquisition of these sequences (Schmitt, Dörnyei, Adolphs and Durow 2004; Jones and Haywood 2004; Ellis 2008; Schmitt 2010). Conversely, good command of these FSs, of their typical occurrence and of the constraints that delimit their use is believed to contribute to fluency and therefore is taken as evidence of a high level of proficiency in the foreign language (Boers et al 2006; Eyckmans 2007; Stengers 2009; Wray and Fitzpatrick 2008). Due to this, it is advisable for foreign language learners to become familiar with their possible uses and patterns.

2. Our study

The purpose of this paper is to present the findings related to advanced university learners' use of some discourse-organizing formulaic sequences in a corpus of essays and compare it to native speaker occurrences in the BNC. This section describes the study in terms of the context, the corpus built, our object of study, i.e. formulaic sequences, and the criteria that guided the selection procedure.

2.1. Our context

The learners participating in this study were 237 students who took the annual subject English IV, for the Teacher-Training and Translation Courses at Facultad de Lenguas (Universidad Nacional

del Comahue), between the years 2008 - 2014. These EFL advanced learners gave their written consent for the inclusion of only one written text as they completed a student profile with information about their experience as learners of English as a foreign language, their characteristics as students and the conditions under which data were gathered. The essay collected for the database was the first practical assignment of the year. Participants' age varied between 21 and 34. This corpus represents a transversal sample of learners taking this course in this particular EFL university setting.

2.2. The learner corpus

A data base of written texts of 114,514 words in 5,176 sentences was compiled and studied by means of *Wordsmith Tools 6.0* in search of instances of formulaic sequences used to organise learners' discourse. This software provides a series of statistical data and indices that characterize the corpus under analysis.

Following the requirements established by the University of Louvain (Belgium) for the systematic compilation of texts to be included as part of the International Corpus of Learner English (*ICLE*), each essay of a length of 300 to 500 words (mean 358.54, Std: 202.28) was written under exam conditions in no more than 3 hours and without recourse to dictionaries. The possible topics for learners to write about were provided by the teachers as a list of thesis statements to be supported or challenged.

With the aim of analysing this corpus in terms of the use of discourse-organising formulaic sequences, we used the *Wordsmith Tools*' command 'find', which allows the identification of complete words, contracted forms and also affixes. Once the search expression, i.e. the key word of the selected FSs, was defined and found in the alphabetically arranged wordlist, all the instances of each keyword in the formulaic sequence were located by means of the functions *compute/concordance*. The final selection of examples was done manually and the length of the resulting concordance lines was set to 100 characters or more to ensure enough context of use for the FSs.

2.3. Selection of Formulaic Sequences

The formulaic sequences studied were selected *a priori*, based on evidence of multi-word units found universally and systematically in texts written by native speakers (Wray 2008: 96-98). The different FSs in this study (appearing in Table 1 below) function in discourse as indicators of the logical organization of ideas either at sentence or paragraph level. Some sequences coincide with what Biber et al. (1999; 2006), using a frequency-driven criterion, label as "lexical bundles". The sequences we have selected not only recur frequently in written texts but also serve a special organizing purpose when linking elements in one sentence or relating ideas inside or between paragraphs. As will become evident, some of these sequences are not complete structural units, but they have a discoursal intra- and inter-sentential role which authors like Granger (1998),

Howarth (1998), Nation (2001) and Römer and Arbor 2009 have analysed within the scope of English for Academic Purposes.

As a consequence (of)	All things considered	All the same,
As a result (of)	based on	In contrast (to)
in an attempt	TAKE/TAKING/TAKEN into	On the contrary
At the same time	account	As a conclusion
as a means of/to	BEAR(ING) in mind	In conclusion
give rise to	On the other hand	To conclude/We can conclude
First of all	On the one hand	

Table 1. List of formulaic sequences searched for in the learner corpus

3. Results

In this section we will broadly characterize the way in which learners have used these formulaic sequences in our learner corpus. In order to establish a point of reference against which we can draw conclusions in this respect, we have also made use of the BNC. A comparison with the BNC of some of the figures obtained in this study and of the typical collocations is presented in the table below.

FS	Learner corpus: 114,514 words		BNC: 100 million words
	f(x)	N	BNC
As a consequence,	introducing the effects of a certain course of action	37	868
As a result,		40	7915
As a result of this/that		5	242
As a result of + NP		14	5095
in an attempt to	introducing purpose	5	1355
At the same time,	introducing a fact that should be taken into account	19	6923
As a means	introducing the means through which a result is achieved	6	1212: 37/50 as a means of
Give rise to	introducing the effects of an event or situation	4	1529
First of all,	introducing the first of a number of things that you want to say	25	1380 (spoken) referential
All things considered	evaluating the positive aspect despite drawbacks	13	34 in all Few in comparison Spoken/ informal (email)
Based on	offering evidence or support for something	18	11441
Taking into account	thinking about a situation or considering it when deciding what to do	26	521
TAKE INTO ACCOUNT		10	0
		39	417
Taken into account		8	1032
Bear in mind		16	74/747
Bearing in mind	9	651	
On the other hand	presenting the second of two contrasting points, facts or ways of looking at something	42	5308
On the one hand		26	1417
All the same,	presenting a different case from what has been mentioned	8	1031
In contrast,	mentioning a very different situation from the one you have just	2	1183

	mentioned		
On the contrary	presenting an opposing point of view/ idea/	23	776
To conclude	Ending an	52	648
We can conclude	presentation/argumentation	2	30
As a conclusion,	summarising facts or ideas	8	0
In conclusion,	ending presentation/argumentation + spoken	32	314
On account of	introducing the reason or explanation for something	2	487

Table 2. Number of occurrences of selected FSs in the learner corpus and in the BNC.

The table summarises the findings of our corpus search, where in the first column appear the selected formulaic sequences. In the second the meaning or function that sequence serves in discourse have been included together with the number of occurrences in our database (N), and in the third appear the number of occurrences in the BNC.

3.1. Some interesting insights into learner use of FSs

In this section we will describe the findings in terms of frequency of occurrence in the learner corpus. As can be observed, the most frequent FSs in our corpus (in dark grey in Table 2) are *as a result*, *on the other hand*, *take/-ing/-en into account* and *to conclude*. Most of these have been used appropriately in learners' essays to indicate precisely what each of these sequences is primed for. Out of these four, *as a result (of)* is the clearest example of how well learner use reflects the frequency of this discourse-organiser in the BNC, for this is the sequence with the highest number of instances found both in the learner corpus and the British database. The different variations offered by learners for this discourse-organizing function present similarities with native use in the BNC and respect their frequency rate: *As a result*, (40 cfr. 7915); *As a result of this/that* (5 cfr. 242) and *as a result of + NP* (14 cfr. 5095).

- (13) se women are very tall and extremely skinny. **As a result**, not many people have access to thei
- (14) that the students who have paid a basic one. **As a result of this**, differences among students
- (15) fore, the chaotic world where we live is not **as a result of** the poor teaching practice but the po

A similar situation is noticed in the case of *on the other hand*, as learner use seems to mirror the proportion in the BNC (with the obvious differences relative to corpus size). Moreover, if we compare *on the one hand* with its counterpart, our corpus also reflects the difference there is between one and the other in native language (1417 vs. 5023), a difference which Biber has discovered in his own studies (Biber *et al.* 1999). While *on the one hand*, can typically be omitted, *on the other hand* is essential to present "two opposite ways of thinking about a situation" (CALD 2008: 651).

The FS *take into account* has varied grammatical realizations, some of them allowing a direct complement (*take something into account*). A few sentences containing this FS in our corpus do not necessarily organize discourse, as shown in (16) and (17) below.

- (16) their products and make their clothes **taking into account** only the bodies of skinny models.
- (17) they are just to be accepted or to be **taken into account** without knowing that this may affect

However, most of the concordance lines in our database, like examples (18), (19) and (20) evince the fact that learners have used this FS and its variations to structure ideas inter- and intra-sententially:

- (18) affairs as regards education, we should **take into account** three aspects that reflect three differ
(19) uring unto the end. A further aspect to **take into account** as regards the students' welfare has t
(20) me goal that is to provide knowledge. **Taking into account** what has been mentioned, it is pos

In sentence (18), the FS is accompanied by a modal verb and the whole phrase organises the presentation of the supporting arguments that ensue. In (19) this learner has chosen a more complex and native-like construction with the same function, inserting an adjective to modify *aspect*. In the learner corpus there are a number of similar examples which contain other nouns like *issue* and *factor* and adjectives like *special* and *important*. Notice that these two lines organise the upcoming discourse in terms of what will be developed next, while (20) exemplifies a very frequent use among learners for this FS: summarising previous argumentation. Such is the case of (21) and (22), except for the fact that the formulaic sequences contain the nouns *this* and *everything* within them, which are not frequent in written text among native speakers.

- (21) e a decent and comfortable life. **Taking this into account**, it can be said that, in spite of having
(22) s produced in our country. **Taking everything into account**, I think that the general belief in our

When we analyse learner preferences for *To conclude*, what stands out is the fact that there are more instances in our corpus of this sequence with the purpose of “ending a presentation or argumentation” than in the British database. While *To conclude*, appears in initial position in only 23 out of the 49 instances in the BNC, learners seem to have used this FS more pervasively. A few of the instances of this sequence in the native database may indicate *that* it is more frequently used in spoken than in written language, which has clearly been overlooked by learners.

One of the main differences in use between the learner corpus and the BNC is perceived in what learners include after the phrase in question. When they write examples like (23) and (24), native speakers prefer “It is possible to conclude that”, “it seems right to conclude” and “it is fair to conclude”. In the case of this formulaic expression, learners seem not be aware of these other sequences which are more recurrent in native language.

- (23) look elegant and comfortable at the same time. **To conclude, it must be said that** the fashion
(24) considering teaching a non-valued profession. **To conclude, we can say that** teaching is no lo

A careful analysis of learner use of this FS and its co-text shows there is a significantly larger number of modals accompanying it in the learner corpus than there is evidence of in the BNC. Learner use also implies more explicitly referential language to what follows (Biber 2006:139), as in *we can say that* in (24) above, than is observed in those examples of the BNC reflecting the same summarizing function.

The **least frequent formulaic sequences** in our database are *all the same, in contrast, in an attempt to, as a means (of/to), give rise to, as a conclusion* and *on account of*. These formulaic sequences have a common feature with respect to the discourse organising functions: with the exception of *All the same* and *in contrast*, they function within sentence level, linking propositions

in sentences, rather than operating at paragraph level, as in some of the cases discussed above. Even though *All the same* is quite frequent in the BNC, there are very few instances in the learner corpus and, out of these, some, like (25) are not even examples of this organising use.

(25) ons and the products they make are very expensive **all the same**. One of the industries that sel

In the case of *In contrast*, in half of the uses found in the BNC it functions as an adverbial followed by a comma, which some learners appear to be aware of, as shown in the correct form in (26). What is missing in learners' repertoire, it might seem, is the formulaic variation that is followed by a prepositional phrase introduced by either *to* or *with* as in (27), equally recurrent in the BNC (1053),

(26) f a positive and enjoyable atmosphere at work. **In contrast**, many people claim that success

(27) FA6 1048 Physics was perceived as exciting, progressive and fundamental, **in contrast to** other disciplines which were perceived as routine, static and lacking substance.

Discourse-organising formulaic sequences that function intra-sententially like *in an attempt to*, *as a means*, *give rise to* and *on account of* are scarce in the learner corpus. Although, as shown in Table 2, there are some instances of each of them (2 to 8 depending on the FS), the figures do not reflect its use in the native database, where all these lexical realizations of connectors are very frequent.

What happens with *As a conclusion*, is noteworthy: although these are non-existent in the BNC, there are eight occurrences in the learner corpus. *In conclusion* appears in the BNC, though not altogether frequently, and could be an appropriate alternative for *as a conclusion*.

3.2. General discussion

The careful analysis of concordance lines from the learner corpus has shed light on characteristic phrasing in learners' written language. One of these features is related to the discursial opposition they present when introducing points of view in the company of some of the FSs studied. There is a widespread use of 'we' followed by the modal 'should' to introduce new factors, issues and aspects that need to be considered in the discussion. In contrast, when students need to present a negative version of what is to be taken into consideration, they turn to the pronoun *they* or clearly make reference to *people* who do not share their views. These linguistic conditions typically accompany the FSs *take into account* and *bear in mind*.

Across a number of the sequences explored, more elaborate and complex examples have been found like *It is vital /essential / important for (sbdy) to bear in mind/ take into account*. These are rather unusual in the learner corpus. Yet the fact that there are some instances may point to a growing sense of awareness in some of the students about native-speaker use. These options, which are recurrent in the British database, may seem to be a way round learners' overuse of modals, which is not as frequent in the BNC in the company of the FSs analysed.

Overall, learners may appear to prefer connectors as discourse-organisers rather than linking formulaic sequences like the ones discussed as the least frequently used in 3.1.. There is evidence to claim that they feel more confident in the use of sequences like *As a Consequence*, *As*

a result, On the other hand, since they require less linguistic and discourse manipulation. By inserting these sequences in the appropriate positions, they make sure their meanings are put across with a minimum of risk.

One word of caution is needed here as regards the conclusions that can be drawn from this study. Due to the varied nature of text selection in the BNC, native use of discourse-organising formulaic sequences may not faithfully represent their canonical use in essays by native speakers. The findings derived from the comparison between the BNC and our learner corpus seem to shine a somewhat dim light on our students' formulaic competence and provide us with a few guidelines as to how to develop it further. We must not forget, however, that these two databases are not entirely comparable. Future lines of inquiry should be undertaken to guarantee that the EFL learner corpus is matched against an equivalent corpus of essays written by English-speaking advanced university learners, like COCA.

Even though possible areas have been highlighted in which learners' competence might be improved, we should bear in mind that this is a cross-sectional study that captures four cohorts of language learners' written production at one static point in time. The essays compiled were the first in the academic year, which indicates that these learners might have reached a higher level of competence throughout their subsequent courses. Further research should explore these students' developing formulaic competence across time.

4. Pedagogical implications for the teaching of vocabulary

A synthesis of the main issues has been presented in relation to the shift in paradigms with respect to the lexicon. The different aspects involved in knowing a word have been discussed. In addition, learner use of a number of discourse-organising formulaic expressions has been analysed in an attempt to describe similarities and differences with the use native speakers make of these same FSs. The deviations in learners' versions seem to indicate some degree of unawareness of the constraints these sequences present, which might show the need for teacher intervention. We now turn to ways in which this task can be approached.

In his book *The Lexical Approach*, Lewis (1993) proposes that language should be recognized as grammaticalized lexis instead of lexicalized grammar, revolutionising the perspective on foreign language learning and teaching. The findings mentioned in section 1.1 present substantial evidence to support Lewis's (1997, 2000) emphasis on teachers focusing not only on individual words but also on multi-word units. He then puts forward a number of principles that should guide teachers' work with respect to the teaching of vocabulary in context and, in particular, of these sequences.

To begin with, teachers should be able to identify those strings that are most frequent in native language use and that are appropriate to their students' level, abandoning the habitual though unnatural tendency to teach words in isolation. Words not only need to appear in a context that prepares the ground for their use but as teachers, we should also pay attention to the *linguistic* context immediately surrounding them.

Learners' noticing of the existence and recurrence of these strings of words is essential to their acquisition and this should be teachers' main concern. Boers and Lindstromberg (2009) assert that "for learners of a second language to benefit from this fluency-facilitating feature of chunks, chunks that are met, noticed and learned must then be adequately entrenched in the learners' long-term memory" (p. 10). Lack of noticing might lead to inability to use them as units. It is our task to provide students with as much and as varied input as possible, containing target expressions for them to notice.

Bearing in mind the fact that not all the input presented to learners will become intake, the more repetitions of the selected phrases they are exposed to, the more chances there will be for learners to incorporate them so that these sequences become intake. For this to happen, we should contemplate some of the following suggestions:

- Shift from a teacher-centred lesson to a more learner-centred approach in which students are in charge of **observing** the input (analyzing the examples provided), **making hypotheses** as regards the regularities found in it, and finally **experimenting** with the language themselves (producing their own examples to convey personal meanings). A key issue when asking students to observe the input lies in the selection of examples.

- Present significant and numerous **examples** illustrating the natural uses of the sequences selected. In order to do this, Lewis promotes the use of corpora and databases, or at least, a quick, manual search across different sources of authentic material. Two or three instances of a new chunk are not enough for students to be able to hypothesise on its meaning and use and to make their own generalisations. It is advisable that examples should be presented in written form, even if they illustrate uses of oral language, to help learners manipulate the evidence at hand.

- Design tasks that will guide learners to **notice** the recurrence of the sequence and its salient features, **sort** the data (examples) to find categories of meaning or regularities in the language and **describe** their findings to their peers, pooling insights and discoveries. Lewis emphasises the need to direct learners' attention to the selected sequences, perhaps following certain input-enhancement techniques like bold type, bigger font size, underlining, etc.

- Ensure the inclusion of tasks that will **cyclically revisit** the sequences presented. Since language acquisition is non-linear, i.e. it does not necessarily follow a fixed order, and since learners have their own rate and speed of acquisition, we should guarantee multiple opportunities for everyone to meet the new strings of words at different times of the teaching year.

If we take all these considerations into account, we will be helping students to retain information in general and FSs in particular in their long term memory. Following Stevick (1996), if learners discover things on their own, and learning becomes significant because of relevance, patterns and connections established between old and new information and emotions, knowledge will be more efficiently stored. Moreover, as Morgan & Rinvolutri (2004) claim both depth in the treatment of the vocabulary item or sequence and learners' interaction with it "are necessary if the encounter [with a word] is to be meaningful" (p.5). It is precisely this degree of meaningfulness that

will contribute to retention in memory and, therefore, to easier access for production in learners' future linguistic endeavours.

References

- (2008). *Cambridge Advanced Learner's Dictionary- Third Edition (CALD)*. Cambridge: C.U.P.
- Biber, D. (2006). *A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Boers, F. y S. Lindstromberg. (2009). *Optimizing a Lexical Approach to Instructed Second Language Acquisition*, New York: Palgrave Macmillan.
- Boers, F, J. Eyckmans, J. Kappel, H. Stengers & M. Demecheleer (2006). "Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test". *Language Teaching Research* 10: 245-261.
- Ellis, N.C. (2008). Phraseology: The periphery and the heart of language. In Meunier, F. & S. Granger (eds.), *Phraseology in Foreign Language Learning and Teaching*. (1-13) Amsterdam: Philadelphia: John Benjamins Press.
- Eyckmans J. (2007) Taking SLA research to interpreting: does knowledge of phrases foster fluency? In: F. Boers, J. Darquennes and R. Temmerman (Eds.) *Multilingualism and Applied Comparative Linguistics, volume 1: pedagogical perspectives* (89-105). Cambridge: Cambridge Scholar Publishing.
- Granger, S. (1998), "Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases", in Cowie, A. P. (Ed.) *Phraseology: theory, analysis and applications*. (145-160) Oxford: OUP
- Hoey, M. (2005). *Lexical Priming: A new theory of words and language*. New York: Routledge
- Howarth, P. (1996). *Phraseology in English academic writing*. Tübingen: Max Niemeyer.
- Howarth, P. (1998). The Phraseology of Learners' Academic Writing, In Cowie, A. P. (1998), (Ed.). *Phraseology: Theory, analysis and application*. Oxford: OUP
- Jones, M. and S. Haywood. (2004). "Facilitating the acquisition of formulaic sequences" in Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use*. (269-299) Amsterdam: John Benjamins Press.
- Lewis, M. (1993). *The Lexical Approach*. Hove UK: Teacher Training Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach*. LTP.
- Lewis, M. (2000). *Teaching Collocations*. London: Thompson Heinle
- Li, J. and N. Schmitt (2010). The development of collocation use in academic texts by advanced L2 learners: a multiple case study approach. In Wood, D. (ed.) *Perspectives on Formulaic Language*. (23-46), New York: Continuum International Publishing Group.

- Meara, P. (1983). *Vocabulary in a second language (Vol. 1)*. London: Centre for Information on Language Teaching and Research (CILT).
- Meara, P. (1992). *Vocabulary in a second language (Vol. 3). Reading in a Foreign Language 9, 1*. (Complete issue)
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon.
- Morgan, J. and M. Rinvolucri. (2004). *Vocabulary*. Oxford: OUP.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*, Cambridge: CUP.
- Nattinger J. and J. DeCarrico. (1992) *Lexical phrases and language teaching*. Oxford: OUP.
- Römer, U. and A. Arbor (2009). English in Academia: Does Nativeness Matter? *Anglistik: International Journal of English Studies 20:2*, pp. 89-100.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: CUP.
- Schmitt, N., Z. Dörnyei, S. Adolphs and V. Durow (2004). "Knowledge and acquisition of formulaic sequences" in Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use*. (55-86) Amsterdam: John Benjamins Press.
- Schmitt, N. (2010). *Researching Vocabulary: a Vocabulary Research Manual*. London: Palgrave Macmillan.
- Schmitt, N. and R. Carter (2004) "Formulaic sequences in action: An introduction", in Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use*. (1-22) Amsterdam: John Benjamins Press.
- Sinclair, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford: OUP.
- Sinclair, J. M. (2004). *Trust the text: language, corpus and discourse*. London: Routledge.
- Stengers, H. (2009). "The Idiom Principle put to the test: an exercise in comparative applied linguistics. Unpublished PhD Dissertation, Free University Brussels.
- Stevick, E. W. (1996). *Memory, meaning & method: A view of language teaching (2nd ed.)*. Boston: Heinle & Heinle.
- Weinert, R. (2010). "Formulaicity and Usage-based Language: Linguistic, Psycholinguistic and Acquisitional Manifestations", in Wood, D. (Ed) *Perspectives in Formulaic Language: Acquisition and Communication*, (1-22) New York: Continuum.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: CUP.
- Wray, A. (2008). *Formulaic Language: Pushing the boundaries*. Oxford: OUP.
- Wray, A. and T. Fitzpatrick. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In Meunier, F. & S. Granger (eds.), *Phraseology in Foreign Language Learning and Teaching*. (123-148) Amsterdam: Philadelphia: John Benjamins Press.