

## The Incorporation of the Corpus-Based Approach in the Teaching of Second Language Writing: A Case Study of Second Year L.M.D Students

Received: 22/11/2019 ; Accepted: 04/04/2020

### Abstract

Writing is an indispensable skill in the realm of foreign language learning. Achieving a high level in such a skill is a complex undertaking. This complexity stems from the high standardised and conventionalised nature of writing, which entails awareness and mastery of the cognitive, linguistic, and psychological factors associated with such a process. This paper looks at the practicality and possibility of applying the Corpus-Based Approach to the teaching of L2 writing for second year L.M.D students. A corpus-oriented paradigm of teaching writing uses a variety of corpora as a method of instruction which would raise the students' awareness of lexico-grammatical competence in terms of vocabulary, grammar, and the lexico-syntactic means of establishing textuality in terms of cohesion and collocation.

**Keywords:** Writing; corpus-based approach; corpora; lexico-grammatical competence; collocation.

**Abdeldjalil BOUGHEZAL \***

Faculty of Letters and Languages  
Department of Letters and English  
University of Djillali Liabes Sidi  
Bel-Abbes,  
Sidi Bel-Abbes, Algeria-

### Résumé

L'écriture est une compétence indispensable dans le domaine de l'apprentissage des langues étrangères. Atteindre un niveau élevé dans une telle compétence est une entreprise complexe. Cette complexité découle de la nature hautement standardisée et conventionnelle de l'écriture, qui implique une prise de conscience et une maîtrise des facteurs cognitifs, linguistiques et psychologiques associés à un tel processus. Cet article examine donc l'utilité pratique et la possibilité d'appliquer l'approche fondée sur le corpus à l'enseignement de la rédaction en L2 pour les étudiants de deuxième année L.M.D. Un paradigme pédagogique d'écriture basé sur un corpus utilise une variété de corpus comme méthode d'instruction afin de sensibiliser les étudiants au vocabulaire, à la grammaire, aux déplacements progressifs, à la collocation, et aux moyens lexico-syntaxiques d'établir la cohésion.

**Mots clés:** Écriture ; approche par corpus ; corpus ; cohésion ; cohérence ; e, collocation.

### ملخص

الكتابة هي مهارة لا غنى عنها في مجال تعلم اللغات الأجنبية. تحقيق مستوى عالٍ في مثل هذه المهارة هو مهمة معقدة. ينبع هذا التعقيد من الطبيعة المعيارية والتقليدية العالية للكتابة، والتي تستلزم الوعي وإتقان العوامل الإدراكية واللغوية والنفسية المرتبطة بهذه العملية. تبحث هذه الورقة، بالتالي، في التطبيق العملي وإمكانية تطبيق النهج القائم على أساس الإحضرار على تدريس الكتابة لطلاب السنة الثانية. يستخدم نموذج تعليمي موجه نحو الإحضرار للكتابة مجموعة متنوعة من الكلمات كوسيلة من وسائل التعليم التي من شأنها أن ترفع وعي الطلاب بالمفردات والقواعد النحوية والتحويلات السلسلة والالتصاق وأجهزة الإقناع والوسائل المعجمية لإقامة التماسك.

**الكلمات المفتاحية:** الكتابة؛ المقاربة المستندة إلى البرمجيات؛ قاعدة البيانات؛ التماسك؛ التجميع.

\* Corresponding author, e-mail: [abdeldjalil.boughezal@gmail.com](mailto:abdeldjalil.boughezal@gmail.com)

## Introduction

The ability to express oneself in a foreign language through an adequate and accurate writing remains at the heart of the academic achievement. Mastering a foreign language writing competency is a long and complex undertaking. One's whole person is affected when struggling to reach beyond the confines of one's first language into proficiently writing in a foreign language. The Algerian higher educational context emphasises the extreme necessity of mastering the skills of writing. The latter is regarded as a backbone of the academic as well as the professional success.

Notwithstanding the substantial need to get a grip on such an intricate skill in language teaching and learning, students, in the department of English at the University of Hadj Lakhdar, Batna, still demonstrate great deficiencies in expressing themselves through writing. These difficulties may stem from their poor lexical repertoire, unawareness of the collocation mechanisms needed to exploit the available vocabulary, and the lack of grammatical rules and conventions to reach the desired syntactic maturity. The aforementioned deficiencies, when combined together, seem to make students consider writing as a thorn stuck in their throats. A lot of researches and studies have suggested various solutions to the previously mentioned puzzling writing problems. One such example is the advocacy of technology epitomised in the corpus-based approach as a remedial tool to enrich the students' knowledge of collocation and lexico-grammatical features.

### 1. The Nature of Corpus Linguistics

Prior to highlighting the notion of corpus linguistics, it is necessary to very briefly elaborate the meaning of the word "Corpus". The term corpus is a Latin word meaning "body". Today, it is used to refer to representative collection of texts in a particular language or dialect to be subject for linguistic analysis. Chiefly, a corpus is "a finite collection of machine-readable texts sampled to be representative of a language or variety", McEnery and Wilson (1996, p. 218). Corpora stand for the process of using of a vast collection of representative samples of texts encompassing several varieties of a language used in discrepant trends of linguistic interactions. Theoretically, a corpus is "Capable of Representing Potentially Unlimited Selections of text", Dash (2005, p. 35).

As for corpus linguistics, it is believed to be a multidimensional area with a huge spectrum that includes all varieties of language use in all linguistics contexts, communication, and comprehension. The implementation of corpus linguistics into language study has yielded a new understanding of what is language, its use, nature, design, and how it should be studied. Chiefly, Corpus Linguistics is a philosophy which has the aim of investigating language and all its properties by describing large collections of text samples. Such a philosophy has been adopted in a variety of research areas for ages: from descriptive study of a language, to language education, to lexicography, etc.

Throughout the years, the notion corpus linguistics has long been conceived as a perplexing and confusing concept. Lexicographers, linguists, and language professionals have made tremendous endeavours to decode some of the intricacies associated with such a term. In this respect, Stubbs (2006) emphasises the frequent reticence of corpus analysts in establishing a scientific elaboration of their operational methods. This ambiguity which governs the discussion of the methodological framework adopted is very surprising with rapport to the scientific frame that corpus linguistics claims to hold.

When addressing the issue of what corpus linguistics 'is', not only have discrepant descriptions been provided, but alternatives have been explicitly suggested and rejected. These involve statements which regard corpus linguistics as either a tool, a method, a methodology, a methodological approach, a discipline, a theory, a theoretical approach, a paradigm (theoretical or methodological), or a combination of these.

## **2.The Corpus-Based Analysis and Methodology**

Corpus-based analysis is seen as a sophisticated method of answering questions that have long been asked by linguists. Corpora can serve the purpose of testing hypotheses and adding a quantitative dimension to many linguistic studies. As such, it is legitimate to argue that corpus software offers the researcher with language in a form that is not normally encountered and that this might bring to light patterning that often skips the realm of linguistic detection. Corpus-based research, thus, has led to a reassessment of what language is like, its nature, use, and design.

### **2.1. Searches, Software, and Methodologies**

Corpus based studies are manifested through the use of dedicated software. The latter inevitably mirrors assumptions about methodology in corpus investigation. At its basic modus operandi, corpus software:

- Searches the corpus for a given target item,
- Counts the number of instances of the target item in the corpus and calculates relative frequencies,
- Displays instances of the target item so that the corpus user can carry out further investigation. (Meyer, 2002)

### **2.2. Search Items**

All corpora may be searched for use instances of a single word (e.g., week). The majority of search software, also, have the potential of conducting a single search to find sets of words (e.g., week, century, decade) and strings of words (e.g., the coming week).

Using the software, one can search for a word when it is tagged for a particular word class only, such as 'light' when it is tagged as an adjective, not as a noun or a verb; given sequences of tags, such as 'preposition, determiner, noun'; or individual words followed or preceded by given word classes, such as fundamentally followed by an adjective. Similarly, a corpus that is parsed will allow searches for particular clause types or structures. For example, one may search in the International Corpus of English for sentences that contain an if clause before or after the main clause (Nelson et al., 2002, p. 57).

Further, corpora are well equipped with the potential of annotating other types of information, such as semantic categories, categories of cohesion, collocation, parallelism (Garside et al., 1997). Software calculates the frequency of occurrence of such categories and, usually, compares this frequency in other corpora.

### **2.3. Word Lists and Frequency Information**

A word list is a list, generally ordered either alphabetically or in frequency order, of all the words in a particular corpus with feedback about the frequency with which that word occurs in the corpus. The most basic word lists interpret 'word' as merely a collection of letters; so, for example, the occurrences of walk is searched without no difference between the noun and the verb, and the occurrences of walks, walking, and walk are given separately. More specialized lists make a distinction between the noun and verb occurrences of walk (Leech et al., 2001). Much more sophisticated are word lists that provide distinction in meaning and use (e.g., between walk meaning 'move in slow motion' and other meanings of walk).

### **2.4. Comparative Frequencies**

Unless it is comparative, information about frequency can never be authentic and representative. For example, In the study of 'if clauses', it is believed that these clauses are more frequent before the main clause than after in either written registers or formal spoken registers. However, it is not the case in informal spoken registers (Nelson et al., 2000). Therefore, the expediency of frequency analysis is to compare

one corpus with another and, by necessity, to compare two languages, varieties of a language, or text types.

A much precise calculation takes place when comparing relatively small corpora of specialized texts with larger, more general corpora, using the Keywords program (part of the Wordsmith Tools suite, Scott, 1996). Keywords ranks the words in the specialized corpus in order of the magnitude of their difference from the general corpus. This illustrates the difference between specialized texts and other text types. For example, “a corpus of newspaper feature articles, when compared with a more general corpus of newspaper texts, is found to have Keywords such as tax, European, war, education, schools, and church” (Scott, 2001, p. 116). These give a clear idea about the orientation as well as the dominating themes of the articles in question. Other Keywords have a grammatical dimension such as pronouns, and prepositions (Scott, 2001, p. 126). Such words often occur in specific phraseological sequences that are more frequent in the specialized corpus than in the general one, Gledhill (2000).

### **2.5. Concordance Lines**

The aforementioned software has a number of statistical operations on items found in the corpus, ranging from simply counting the number of occurrence to calculating the degree of significance of occurrence. Contrariwise, software that provide concordance lines is used to identify the target item (usually a word or phrase) each time it occurs in the corpus and presents each instance, or as many as are required, to the corpus user. “Usually this is done with the target item in the center of the screen and a few words to the left and right of that item. This ‘key word in context’ presentation, as it is known, has a number of uses. Even the small amount of context is usually enough to show what the word or phrase means, what phrases it often occurs in, and/or the discourse function that it has” Hyland, (2007, p. 125).

### **2.6. Register**

Much of the conducted comparative investigations have used corpora to scrutinize the English language as it manifests in various contexts. The latter has been defined under the auspices of the linguistic theory (e.g., in Matthiessen, 2005, where register is defined according to systemic theory), or according to a less theoretical, ‘commonsense’ view of where clear distinctions might lie. Biber et al. (1999), for instance, use broad register categories of conversation, fiction, news reporting, and academic prose. Others have made more refined distinctions: the CANCODE corpus of spoken English, for example, distinguishes between transactional, professional, socializing, and intimate contexts (Carter, 2004; Hyland 2000). It distinguishes between academic genres such as research articles, book reviews, abstracts, and textbooks and between different academic disciplines. The major focus of works of this genre is placed on the explanation of quantitative results qualitatively.

## **3. Corpora and Second Language Teaching**

The primordial aim of second language acquisition is to build models of particular representations of learners at a particular stage of apprenticeship. The most salient evidence behind these mental representations is the language produced by learners whether spontaneously or through data elicitation procedures (Myles, 2005). The efficiency of the second language acquisition process is highly dependent on the reliability as well as the validity of these elicitations, procedures, and data collection.

Not before the mid-1980s that language corpora have shown tremendous potential in computer-assisted language learning, research, and teaching. Convergence between corpora and pedagogy has triggered radical changes in the way second language (L2) materials development, curriculum design, and teaching methodology are approached and designed. Despite the fact that classroom applications of corpora do not fully occupy the arena of SLA practice, they have been very attractive to language teachers largely due to such advantages as salience of linguistic phenomena and extensive exposure to authentic language use in various registers and genres. This availability has led educators to adopt corpus-based learning to boost inductive, discovery-oriented learning opportunities whereby students themselves conduct a

corpus-based analysis and, therefore, engage in active and autonomous learning (Chambers, 2010; Boulton, 2009; Braun, 2007). Nevertheless, empirical evaluations of hands-on uses of corpora by L2 learners have remained relatively marginal (Rodgers, Chambers & Le Baron-Earle, 2011).

Corpus-based analyses are conducted through the software programs such as Wordsmith Tools (Scott, 2004), which encompasses various text-handling tools to approach quantitative and qualitative textual data analysis. Wordlists provide information on the frequency and distribution of the vocabulary – single words but also word sequences used in one or more corpora. Wordlists for two corpora can be compared automatically to highlight the vocabulary that is particularly salient in a given corpus, i.e. its keywords or key word sequences.

In light of the above mentioned paradigms, corpus-based studies have had interest in variety of distinctive linguistic features of academic discourse as compared with other genres. Biber et al. (1999) have shown, for example, “that nouns, nominalisations, derivational suffixes and linking adverbials are particularly frequent in academic prose while private verbs, *that*-deletions and contractions occur very rarely” (p. 57). Studies of vocabulary in academic prose have emphasised the primordial role of a sub-technical vocabulary that is typical in a vast array of academic texts and disciplines which generally serve organisational or rhetorical functions prominent in academic writing, e.g. introducing a topic, hypothesizing, exemplifying, explaining, evaluating, concluding (cf. Thurstun and Candlin, 1998; Luzón Marco, 2001). Other investigations referred to the existence of an English for Academic Purposes-specific phraseology epitomised in word combinations that are semantically as well as syntactically compositional, e.g. *in the presence of*, *the aim of this study*, *the extent to which*, *it has been suggested*, *it is likely that* (Biber et al, 1999).

These studies have the credit of speculating that the phraseology of academic discourse is highly conventionalised and that “novice writers differ from professional writers in their use of EAP-specific lexical bundles” (Cortes, 2002).

To date, over hundreds of corpora have been developed, and interest in using them has increased steadily, especially in the area of academic writing. Numerous corpus-based findings have emerged from contrastive interlanguage analyses (Granger, 1996) identifying lexical, grammatical, phraseological, stylistic, and pragmatic features of learner language. Research suggests that English language learners clearly exhibit problems of frequency, register, positioning, semantics, and phraseology (Gilquin et al., 2007). Although most of the findings are still largely at the level of implications, which may have delayed pedagogical use (Granger, 2009), corpus-based methodologies are slowly but surely making their way toward the classroom.

#### 4. The Corpus-Based Approach and Writing Pedagogy

This part of the article at hand profoundly highlights the way E.F.L students use corpus-based materials in L2 writing instruction so as to gain deeper insight into how learners may benefit from corpus-based writing activity. As learners are supposed to gain profit from such an experience, it is crucial to determine how they react to a classroom environment in which corpus use has the lion's share. As such, E.F.L teachers can dynamically adopt corpus-based techniques in such a way that creates more effective and meaningful learning conditions.

One of the central principles of the corpus approach to language descriptions is that vocabulary and grammar are interrelated rather than distinctive from each other (Halliday, 1992; Sinclair, 1991). Within this linking of vocabulary and grammar, also known as lexico-grammar, there is an emphasis on the co-occurrence or most frequent combinations of words, i.e., “collocation” (Biber & Conrad, 2001; Conrad, 2000). For instance, the noun “location” is often followed by the prepositions “of” and “for.” This combined focus on lexical input and grammatical function is of considerable value to someone acquiring English as a second or foreign language, as well as to teachers of ESL or EFL.

Recent attempts to connect the corpus approach with genre analysis have been particularly beneficial in such domains as English for Academic Purposes (EAP) or English for Specific Purposes (ESP). Genre-based corpus analysis looks for common

collocational frameworks within particular genres of communication, and in the process facilitates more effective communication of the kind expected inside specific genres, such as medical or legal English, where specialized word patterns are likely to occur. Given such emphases, the use of corpus data has become increasingly appealing in the context of L2 writing instruction, where the simultaneous focus on vocabulary, grammar, and discourse patterns provides second language writers with the kinds of target language input they especially need to achieve high levels of proficiency as L2 writers.

In addition to the lexico-grammatical expediency, the corpus-based approach has also been regarded as contributive to L2 writing by providing students with authentic experience of real language use and large quantities of real-life target language discourse.

“Exposure to these examples of genuine language use can (a) enrich learners’ understanding of specific uses of target words in a wide variety of contexts and (b) expand their L2 linguistic repertoire. Students’ encounters with these multiple samples of discourse combinations should then contribute to growth as L2 writers (and readers)”, Thurstun and Candlin, (1998, p. 256).

As corpora are proved to facilitate the process of L2 writing, scholars launched an appeal for an immediate implementation of such a paradigm in teaching materials and classroom activities. This does not mean an exclusive corpus-dominated writing pedagogy, but rather the inclusion of some corpus-based activity with respect to treating the language side of L2 writing. This shift toward a corpus component has been growing concern about the traditional way in which teachers and textbook writers have provided language input for students. Biber (2001) points out that, “these decisions have usually been based on the author’s gut-level impressions and anecdotal evidence of how speakers and writers use language” (p. 101). Similarly, teachers may include their own notions of which language items are most useful for students to learn. In this context, Biber argues “empirical analyses of representative corpora provide a much more solid foundation for descriptions of language use” (ibid, p. 101).

As noted earlier, another fruitful implementation of the corpus-based approach to the teaching of L2 writing resides in exposing students to various authentic texts in order to raise their familiarization with the functions of words in specific contexts or genres. “By helping students acquire contextualized grammatical knowledge, this corpus-based approach teaches them (indirectly or directly, depending on the explicitness of the instruction) how and where to put words into sentences, which a dictionary often fails to do Odlin, Tribble (2002, p. 25). In a study of a course employing corpora, Odlin and Tribble argued that a corpus-based activity increased “contextual and linguistic awareness raising during an EAP course” (p. 133). This is the very defining contribution of corpora: their potential of providing L2 writers with the concrete linguistic input necessary for high-quality compositions.

Including a corpus-based dimension in L2 writing instruction further boosts inductive language learning, which is a crucial facet of second language acquisition. Inductive learners acquire the ability of elucidating their own inductive discoveries about the target language as they are exposed to multiple instances of a specific linguistic item in use. The result is student-centered discovery learning, which fosters self-confidence and mastery of the learning process (Johns, 1991; Stevens, 1995).

The computer-based nature of a corpus component can also favor ESL/EFL learning because it significantly increases opportunities for exposure to and contact with English language texts. That is, through such computer-based artifacts as the Internet and hypertexts, with their multiple links to other texts, students gain access to an almost limitless supply of target language texts (Conrad, 2000). Computational analysis, then, facilitates the selection of texts from whichever discursive domains students prefer to work within, as opposed to relying on more general texts. A student interested in chemical engineering, for example, can quickly locate and surf chemical engineering-based Internet sites and select those texts of greatest linguistic value relative to her or his needs. Then, too, web-based corpora are becoming increasingly

available for language learning and teaching purposes (Sun, 2000). For example, the computer provides access to the vast riches of the Bank of English, a storehouse of tens of millions of samples of authentic English collocations and other linguistic items based at the University of Birmingham in England, Gavioli and Aston (2001).

#### 4.1. Corpus-Based Approach and Lexicogrammatical Competence

In recent years, interest in teaching grammar has been revived as indicated by many publications on the issue, including those demonstrating the importance and benefits of grammar instruction on students' writing proficiency (N. Ellis, 2005; R. Ellis, 2001, 2002; Philp, 2003; Yuan & Ellis, 2003) and those suggesting new theories and approaches to grammar instruction (Conrad, 2000; R. Ellis, 1995; Francis, 1993; Hahn, 2000; Hinkel & Fotos, 2002; Hughes & McCarthy, 1998; Johns, 1994; Larsen-Freeman, 2002, 2003; Liu & Master, 2003; Thornbury, 2001). Among the suggested theories and paradigms, three advocate grammar teaching in discourse contexts, teaching grammar with a lexicogrammatical approach, and corpus or data-based teaching.

The idea of teaching grammar in contextual discourse is rooted in functional grammar and is based on the dogma that grammar is not limited to forms, but rather expands to include semantics (meaning) and pragmatics (context-appropriate use) (Celce-Murcia & Larsen-Freeman, 1999; Celce-Murcia & Olshtain, 2000; Halliday). The Functionalism trend considers grammar as a resource for language users in the process of meaning-making in a given social context.

In this respect Larsen-Freeman (2003) argues that:

language form, meaning, and use should be approached as an integrated whole. The three aspects of grammar are interwoven because a change in one will involve a change in another". English language students should learn not only how to use correct grammatical forms but also how to use them in a meaningful and appropriate way (p. 53).

To go deeper, Lexicogrammar views lexicon and grammar as two inherently connected parts of a single entity, challenging the traditional "wisdom of postulating separate domains of lexis and syntax" Sinclair, (1991, p.104). In this view, "a grammatical structure may be lexically restricted and, conversely, lexical items are often grammatical in nature, for the use of a lexical item often has grammatical implications" Francis, (1993, p. 142). Many corpus-based studies have exhibited this close lexical and grammatical connection (Biber et al., 1998; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Francis, Hunston, & Manning, 1996, 1998; Hunston & Francis). There also has been increasing evidence in applied linguistics showing the importance of contextual patterns in language use and learning (Hunston & Francis, 1998). In light of these findings, many scholars have argued for the use of a lexicogrammatical approach in language instruction (Aston, 2001; Clear, 2000; Schmitt, 2004, 2005; Sinclair, 1991). Aston claims that:

insofar as different words appear to have distinctive collocational, colligational, semantic, pragmatic and generic associations, . . . every word may have its own grammar in these respects, a grammar which can only be acquired through experience of its typical contextual patternings. (2001, p. 15).

The suggestion of using corpus analysis in grammar teaching has resulted from rapid advancement in computer technology and corpus linguistics in recent years, which has showed unprecedented potential for language learning and teaching. Corpus concordance not only makes accessible an enormous amount of authentic language input but also creates various inductive and deductive language learning opportunities not available in the past (Aston, 2001; Conrad, 2000; Francis, 1993; Hunston, 2002; Hunston & Francis, 1998; Johns, 1994; Sinclair, 2004; Stevens, 1995).

In terms of inductive learning, language learners observe grammar and vocabulary usages in concordance data, and then they discover and generalise findings about usage patterns and rules. In deductive learning, language learners use corpora either to test the rules and patterns they have learned or to classify concordance data by

applying the rules and patterns. It has been argued that such learning activities, especially the inductive type, motivate students and promote 'discovery learning'. The latter is said to be "particularly effective for the acquisition of grammar and vocabulary" (Hudson, 1992; Batstone, 1994), because it helps learners notice and retain lexico-grammatical usage patterns better by engaging them in a deeper language processing. Francis and Johns (1998) also contended that conducting concordance analyses of recurrent collocational and colligational patterns leads to acquisition of more useful general grammatical rules. Furthermore, corpus data offer contextualised language, which enables learners to better understand what Larsen-Freeman (2002) called "grammar of choice" in language use, a choice that native speakers often make according to the context.

In short, examined closely together, research findings have shown the need for a contextualized lexico-grammatical approach to grammar instruction and the useful role that corpora can play in such teaching. In other words, the integration of corpus use, lexicogrammar, and contextualisation in grammar teaching is not a random proposition, but one motivated by the inherent connection and interdependency found among the three practices. However, although there have been quite a few publications introducing the use of corpora in language teaching (Aston, 2001; Flowerdew, 1996; Hunston & Francis, 1998; O'Keeffe and Farr, 2003; Tribble & Jones, 1997), little empirical research has been conducted on the effectiveness of these novel theories and practices. The present study, therefore, was conducted to examine whether and to what extent these new theories and practices are applicable and effective when used as a unified approach in the classroom.

## 4.2. Corpus-Based Approach, Collocation, and L2 Writing Pedagogy

Collocation is one of the most useful assets that corpus research can provide for language description. Collocation refers to strings of words that conventionally go together, which can be more easily understood as "wording" or "word combinations." The term 'collocation' was first used by Firth in its modern linguistic sense, but it was popularised by Sinclair later (McEnery & Wilson, 2001; Partington, 1998). As Sinclair (1991, p. 57) observes, "collocation is originally confined to lexical association patterns, but it often expands to include the association with grammatical items". Many scholars claim that collocation is at the heart of lexicogrammar that looks at words in lexical as well as grammatical surrounding contexts of occurrence.

To further elaborate the notion of collocation, Kennedy (1998) argues that the concept of collocation is considered the place "where grammar and lexis meet in the phrase" (p. 289). Aijmer and Altenberg (1991) also assert that "collocations ... represent the intersection of lexicon and grammar, an area which can be fruitfully studied in corpora" (p. 4).

Whereas Chomskyan linguistics views the irregularity of collocation as a challenge to the rule-governed generative grammar, many scholars claim that human beings' language use is not so much creative or generative and that the conventional and idiomatic use of language comprises an integral part of language (Hopper, 1998; Howarth, 1983).

In their groundbreaking article "Two puzzles for linguistic theory: native-like selection and native-like fluency," Pawley and Syder (1983) claimed that native-like usage is much more restricted and predictable than is often assumed. They argued:

fluent and idiomatic control of a language rests to a considerable extent on knowledge of a body of 'sentence stems' which are 'institutionalized' or 'lexicalized'. In order to solve the puzzle of native speakers' ability to select natural and conventional usage among the wide possibility of grammatically correct sentences, they drew upon a 'lexical sentence stem, a unit of clause length or longer whose grammatical form and lexical content is wholly or largely fixed. (p. 191).

Pawley and Syder's work is congruent with the corpus linguistics approach to language. From the corpus-based approach viewpoint, native speakers' language use is not so much inventive as conventional, as is evident in fixed collocational patterns (Hill, 2000). The collocational field has recently gained great attention from those



interested in L2 acquisition and pedagogy. Hill (2000), who emphasises the importance of collocational knowledge in L2 pedagogy, addresses a fundamental question of what it means to know a language. He claims that “students do not really ‘know’ or ‘own’ a word unless they also know how that word is used, which means knowing something about its collocational field” (p. 60). That is, knowing a language really means knowing a variety of collocation patterns and learning “word grammar.” Consequently, this view of language calls for a change in the traditional approach to teaching a second language. It asks for the combination of lexis and grammar rather than teaching grammatical rules and vocabulary separately.

Collocational knowledge has become widely recognised as an important component of L2 writing proficiency (Lewis, 2000). One of the most difficult challenges for language learners is to acquire the natural, idiomatic word combinations that are commonly used in the language. While there is still a lack of a clear understanding of how L2 learners (even with sufficient grammatical knowledge) acquire native-like fluency that is not fully explained by rule-based formal system, the corpus-based approach can contribute to solving a chronic puzzle in L2 pedagogy.

A number of scholars have addressed the role of collocation in second language acquisition (Cowie, 1992, 1996; Fillmore, 1979; Granger, 2001; Hakuta, 1974; Howarth 1996, 1998; Krashen & Scarcella, 1978; Lewis 1997, 2000; Nattinger & DeCarrico, 1992; Nesselhauf, 2003). Although they used different terminology to refer to collocation, they relied on a common framework for defining the characteristics of collocation. They often used different terms interchangeably: “recurrent or fixed combinations” (Bahns & Eldaw, 1993), “lexical phrases” (Nattinger & DeCarrico, 1992), “lexicalized sentence stem” (Pawley and Syder, 1983), “idiom principle” (Sinclair, 1991), “phraseological composite units” (Howarth, 1996, 1998), “prefabricated patterns” (Granger, 2001), “prefabricated routines” (Hakuta, 1974; Krashen & Scarcella, 1978), “formulaic expressions” (Fillmore, 1979), and “formulas, prefabricated or ready-made language, chunks, unanalyzed language or wholes, etc” (Weinert, 1995).

Despite their uses of different terms, the researchers seemed to agree that collocation should be understood as a continuum according to their restrictions of word combinations. Howarth (1996, 1998) and Nesselhauf (2003) presented the most explicit continuum model. Both of them focused on “verb-noun combinations” and presented three major classifications of the collocation: “free combinations (both verb and noun are used unrestrictedly, e.g., want a car), restricted collocations (the verb is used restrictedly and combined with certain nouns, e.g., take a picture), and idioms (both verb and noun are used restrictedly and often seen as a chunk, e.g., foot the bill)”. They went on to point out that traditional English education has primarily focused on the idioms at the expense of collocations that are “of far greater significance and considered as potential problems for learners are less fixed in form than idioms” Howarth, (1996, p. 1). Then, as Farghal and Obiedat (1995) noted, what learners need first is “the awareness of collocational restrictions in English” (p. 327).

A body of research based on this framework exists in order to investigate learners’ use of collocation in L2 writing. Although they adopted different tasks for the study, most of them focused mainly on quantitative analysis of students’ one-time collocation use or final products of writing. “Those studies can be categorized into three areas: 1) test on students’ collocation knowledge, 2) error analysis in students’ actual writing, and 3) comparison of native speakers’ writing with learners’ writing” Bahns, 1993, p. 2).

The first area mostly used a cloze test or translation task to test learners’ L2 collocation knowledge, while focusing on a smaller unit of clause level. The second area examined learners’ writing that was produced through composition tasks and developed error analysis in their collocational uses, (Nesselhauf, 2003). Thus, the studies in the area often focused on the impact of students’ native language on their L2 collocation production and mostly presented the result of L1 interference. The last area often adopted corpus analysis to compare production data of native speakers with that of learners to identify differences in the use of collocation (Granger, 2001; Howarth, 1996, 1998).

## Conclusion

Due to its natural occurrence, the corpus-based approach is a supportive function when it comes to teaching L2 writing. The foremost goal of teaching writing is to help learners gain mastery over the written form of language. Therefore, applying such an approach in the classroom is very helpful. By doing so, learners will have the ability to make their writing coherent and readable. The analysis of text patterns is an available means that creates a room for dispelling some of the predicaments associated with the enigmatic process of writing.

## Bibliography

- [1]. -Aijmer, K. and Altenberg, B. (Eds.) 1991. *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- [2]. Aston, G. (Ed.) 2004. *Learning with Corpora*. Cambridge: Cambridge University Press.
- [3]. Atwell, E. (Eds.) 1993. *Corpus Based Computational Linguistics*. Amsterdam: Rodopi.
- [4]. Baker, M., Gill, F. and Tognini-Bonelli, E. (Eds.) 1993. *Text and Technology: In honour of John Sinclair*. Philadelphia: John Benjamins.
- [5]. BIBER, D., CONRAD, S. and REPPEN, R. 1998. *Corpus linguistics - Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- [6]. Boguraev, B. and Pustejvsky, J. (Eds.) 1996. *Corpus Processing for Lexical Acquisition*. Cambridge, Mass.: MIT Press.
- [7]. Botley, S.P., A.M. McEnery, and A. Wilson (Eds.) 2000. *Multilingual Corpora in Teaching and Research*. Amsterdam-Atlanta, GA: Rodopi. \*
- [8]. Butler, C.S. (Ed.) 1992. *Computers and Written Texts*. Oxford: Blackwell Publishers.
- [9]. Collins, P. and Blair, D. (Eds.) 1989. *Australian English*. St. Lucia: University of Queensland Press.
- [10]. Cuyckens, H. and Zawada, B. (Eds.) 2001. *Polysemy in Cognitive Linguistics*. Amsterdam/ Philadelphia: John Benjamins.
- [11]. Dash, N.S. and Arulmozi, S. 2018. *History, Features, and Typology of Language Corpora*. Singapore: Springer.
- [12]. Fries, U., Mü, V. and Schneider, P. (Eds.) 1997. *From Aelfric to the New York Times*. Amsterdam: Rodopi.
- [13]. Garside, R., Leech, G. and McEnery, A. (Eds.) 1997. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman.
- [14]. Ghadessy, M., Henry, A. and Roseberry, R.L. Eds. 2001. *Small Corpus Studies and ELT: Theory and Practice*. Amsterdam/ Philadelphia: John Benjamins
- [15]. Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London: Longman.
- [16]. Halteren, H.V. (Ed.) 1999. *Syntactic Word Class Tagging*. Dordrecht: Kluwer Academic Press.
- [17]. Hickey, R. and Stanislaw, P. (Eds.) 1996. *Language History and Linguistic Modelling: A Festschrift for Jacek Fisiak*. Vol. 2. Berlin: Mouton de Gruyter.
- [18]. Hofland, K. and Johansson, S. 1982. *Word Frequencies in British and American English*. Bergen: Norway Computing Centre for the Humanities.
- [19]. Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- [20]. Jespersen, O. 1909-1949. *Modern English Grammar on Historical Principles*. 7 Vols. London: Allen and Unwin.

- [21]. Johansson, S. and Hofland, K. (Eds.) 1982. *Computer Corpora in English Language Research*. Bergen: Norwegian Computing Centre for the Humanities.
- [22]. Katamba, F. 1993. *Morphology*. London: Macmillan Press.
- [23]. Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. New York: Addison-Wesley Longman Inc.
- [24]. Kenny, A.J.P. 1982. *The Computation of Style*. Oxford: Pergamon Press.
- [25]. Kucera, H. and Francis, W.N. 1967. *Computational Analysis of Present Day American English*. Providence, RI: Brown University Press.
- [26]. Kytö, M., Ihalainen, O. and Rissanen, M. (Eds.) 1988. *Corpus Linguistics, hard and soft: Proceedings of the 8<sup>th</sup> International Conference on English Language Research on Computerised Corpora*. Amsterdam: Rodopi.
- [27]. Lancashire, I., E. Carol, and C.F. Meyer (Eds.) 1997. *Synchronic Corpus Linguistics*. Bergen, Norway: ICAME.
- [28]. Landau, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- [29]. Leech, G., G. Myers, and J. Thomas (Eds.) 1995. *Spoken English on Computer: Transcription, Markup and Applications*. Harlow: Longman.
- [30]. Levy, M. 1997. *Computer Assisted Language Learning*. Oxford: Oxford University Press.
- [31]. Ljung, M. (Ed.) 1997. *Corpus-Based Studies in English*. Papers from the 17<sup>th</sup> International Conference on English-Language Research Based on Computerized Corpora. Amsterdam: Rodopi.
- [32]. McArthur, T. 1981. *Longman Lexicon of Contemporary English*. London: Longman.
- [33]. McCarthy, J. 1982. *Formal Problems in Semitic Phonology and Morphology*. New York: Garland.
- [34]. McCarthy, M. 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University press.
- [35]. Peters, B.P., Collins, P. and Smith, A. (Eds.) 2002. *New Frontiers of Corpus Research*. Language and Computers. Amsterdam-Atlanta, GA: Rodopi.
- [36]. Ravin, Y. and Leacock, C. (Eds.) 2000. *Ploysemy: Theoretical and Computational Approaches*. New York: Oxford University Press Inc.
- [37]. Schütze, H. 1997. *Ambiguity Resolution in Language Learning: Computational and Cognitive Models*. Cambridge: Cambridge University Press.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- [38]. Souter, C. and Atwell, E. (Eds.) 1993. *Corpus Based Computational Linguistics*. Amsterdam: Rodopi. Sperberg-McQueen, C.M. and Burnard, L. 1994. *Guidelines for Electronic Text Encoding and Interchange*. Chicago and Oxford: ACH-ACL-ALLC Text Encoding Initiative.
- [39]. Svartvik, J. (Ed.) 1992. *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 -Stockholm, 4-8 August 1991*. Berlin, New York: Mouton De Gruyter.
- [40]. Tannen, D. (Ed.) 1982. *Spoken and Written Language: Exploring Orality and Literacy*. Norwood, New Jersey: Ablex Publishing Corporation.
- [41]. Thomas, J. and Short, M. (Eds.) 1996. *Using Corpora for Language Research: Studies in the Honour of Geoffrey Leech*. London and New York: Addison Wesley Longman.
- [42]. Véronis, J. (Ed.) 2000. *Parallel Text Processing: Alignment and Use of Translation Corpora*. Dordrecht: Kluwer Academic Publishers.
- [43]. Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (Eds.) 1997. *Teaching and Language Corpora*. London: Longman.