Perspectives on Corpus Linguistics: The Methodological Synergy in Second Language Pedagogy and Research

Ibrahim Bashir
Centre of English Language Studies, Faculty of Languages and Communication Universiti Sultan Zainal Abidin, Terengganu, Malaysia

Kamariah Yunus
Centre of English Language Studies, Faculty of Languages and Communication Universiti Sultan Zainal Abidin, Terengganu, Malaysia

Bashir Ibrahim
Centre of English Language Studies, Faculty of Languages and Communication Universiti Sultan Zainal Abidin, Terengganu, Malaysia

Abstract
Corpus linguistics is a new branch of linguistics but its status is still debatable - either as a theory or a methodology. This article aims to give an overview of the different approaches and perspectives of corpus linguistics. The neo-Firthians contend that corpus linguistics is a method, while other prominent corpus linguists claim that it is a theory. Other corpus linguists believe that corpus linguistics can be both a methodology as well as a theory depending on the extent and purposes it is used for. The applicability of corpus linguistics as a methodology is observed in English Language Teaching and Learning (ELT). Learner corpora are used extensively in second language pedagogy and research as either direct approach involving Data-Driven Learning (DDL) where students participated as researchers and worked directly with corpora or indirect approach where corpus is used as a research method in producing dictionaries, syllabuses, textbooks, and teaching materials.
Keywords: Corpus linguistics, corpus-based study, corpus-driven study, corpus-informed study, Data-Driven Learning (DDL), learner corpora, second language acquisition (SLA)

1. Introduction

Language is a system that people employs to communicate among themselves for multiple reasons, including persuasion, information or entertainment (Ziggiotto, 2016). Halliday (2007) advocates for 'taking language seriously' not only for the appraisal roles it plays in our 'construal experience' but for its status as an enactment of interpersonal relations. making of meanings is the unique and distinctive feature of any language. From the common sense view, based on Firthian (Firth, 1968) and Hallidayan (1993; 2014) approaches language is a 'system of meaning' and that a grammar of a language is "the study of how meanings are built up through the use of words and other linguistic forms such as tones and emphasis" (Bloor & Bloor, 1995, p., 2). However, linguists have varying views on language as a system of meanings in a direct way as Firth and his students Halliday and other neo-Firthian linguists. From other perspectives language is viewed from its segmental structure (words and phrases and sentences) and how these segments work to form meanings. The notion of meaning is central to the study of language and most pervasive and intricate with diverse views among linguists and numerous presuppositions essentially centered to words and range of references that are inherent in or related to or be part of them. Words always play great roles in language studies of meaning variably as primary factors or secondary segments in more abstraction as a metaphoric or idiomatic expression. Philip (2011) considers the word as a container of meaning that one has to delve into to pull out the desired meaning.

With the advent of the computer and enamours internet resources applied in teaching and learning, the ways we study and analyse language have received quite a lot of changes and transformation in the academic (Alkhataba, Abdul-Hamid & Bashir, 2018). The interest of modern language researchers is attracted and focussed on the meaning of lexical patterning and pragmatic discourse which are well established and documented under the corpus linguistic analysis. Hence, corpus linguistics is the fast growing area of language studies that concern the study of meaning from analysing the word list, frequencies and words concordances (Hoffmann, Evert, Smith, Lee, & Berglund-Prytz, 2008). The title of this article might capture at glance the essential presuppositions in corpus linguistics as a new trend of linguistics field that poses a lot of theoretical issues. Despite, having a long history as a field, corpus linguists have still contended with various issues related to the status of corpus linguistics as a method or theory. Many attempts are made to resolve such issues in the last three decades. Some corpus linguists argue that corpus linguistics should be seen as a methodology while others believe it should be considered as a theory of language on itself; some behold a neutral position between the two claims. This article presents these presuppositions of corpus linguistics and their implications in second language pedagogy and research.

Linguists have long age interest in studying on how language is acquired by either native speakers of that language (L1) or second language learners (L2). Many theories, perspectives, paradigms and analytical and research approaches have been grounded, tested and employed to account for that purpose. The evolution of computational linguistics and corpus linguistics in the computer and internet era have revolutionised the language analysis in both L1 and L2 contexts.

2. Second Language Acquisition

The term second language acquisition (SLA) refers to a field of inquiry in linguistics studies established in the 1960s (Ellis & Barkhuizen, 2009; Gass & Selinker, 2007). However, it would

be hard-pressed to state a beginning date of second language acquisition studies, but based on VanPatten and Williams (2015) perspective view, it is probably fair to say that the study of SLA has expanded and developed in the past 40 to 45 years. The notion of second language acquisition is frequently labeled in its twofold senses as either object of inquiry as learning of another language rather than ones mother tongue or as a subject of inquiry as of the process of learning a second language. It is traditionally seen as an offspring of general linguistic theory (Richards, 1992), which enacted presumably from the practical orientation to language learning associated with two periods of linguistics studies, behaviourism, and structuralism. Therefore, we can say that SLA is a constituted field of knowledge which drawn its theoretical inspiration from the philosophical underpinnings of other associated fields. Ellis and Barkhuizen (2009) mention some of the fields associated with English as second language (ESL) including: 'linguistics', 'sociology', 'sociolinguistics', 'psychology', 'psycholinguistics', and 'education'. In general, the term second language acquisition (SLA), according to Richard and Schmidt (2014, p. 206), is "any language learned after one has learned one's native language".

Moreover, Kachru (1992) conceptualises the spread of English around the world into three main circles: the Inner Circle of English refers to countries where English is spoken as a Native Language (ENL) members under this circle included Britain, USA, and Australia. The Outer Circle of English refers to countries where English is spoken as a Second Language (ESL); members included Malaysia, India, Singapore, Nigeria, and Ghana. The Expanding Circle of English refers to countries where English is spoken as a Foreign Language (EFL), members included Saudi Arabia, Jordan, China, Russia, and Thailand (Gika, 1996; & James, 1998).

2.1 Analysing Learner Language

The term learner language in the words of Ellis and Barkhuizen (2009) refers to any form of language produced by learners through written or spoken and paralinguistic (e.g. gestures) modes of communication. The language produced through these modes of communication served as the primary source of data used for the study of L2 acquisition. From its basic features, learner language is 'not monolithic', but rather a highly manifold phenomenon comprising many variables that address the issues of how learner learn a particular language; the processes and techniques; problems and prospects involved. It is primarily concerned with data analysis as an integral part of the research process. "It is shaped by the purpose of the research and the theoretical principles that govern the chosen method of inquiry (Ellis & Barkhuizen, 2009, p. 3)". It aimed at describing competence and competence as the highest goal of linguistic knowledge in the L2 study, be it implicit (rule-based or formulaic) or explicit (Analysed or metalingual).

According to Ellis (2007) The better approach in second language study might be to find out what learners actually do, as opposed to what they think they do, this would be by collecting samples of learner language, that is the language that learners produce when they are called on to use an L2 in speech or writing- and analyse them carefully. There has been a research interest in the analysis of learners' language in the field of second language studies. Richard and Sampson (1974) believe that since language was viewed as a system, the notion of the second language could be considered as a juxtaposition of two systems or super-systems involving mixed features of the two or more systems or to inter-systemic interference.

However, it is worth to be noted that researchers in the field of linguistics have distinctively devised different perspective concerning the issues of learners language and deployed several analytical approaches to analyse the samples of data collected from L2 learners such as Contrastive Analysis (CA); Error Analysis (EA); obligatory occasion analysis; frequency analysis; functional analysis; computer-based analysis (such as corpus-based analysis) and a host of others. This paper limited its scope to corpus-based approach for its currency and the contended issues that surrounding it theoretical perspectives and the mixture of view among its scholastic founders.

3. Meaning and Etymology of Corpus linguistics

The term corpus linguistics was coined by Jan Aarts in the early (1980) who was hesitant in using the term, although many other linguists have been less than happy with it at that time   (Leech, 2011). Later it is conveniently seen as an umbrella term for linguistic research that depends on the use of corpora, as in the words of prominent linguist and corpus linguistics pioneer Sinclear (1991) as he asserts "corpus is a collection of naturally occurring language text chosen to characterise a state or variety of a change"(p.171). It can be best defined as a language study based on the samples from the real-life language used (McEnery & Wilson, 1996).

A corpus (the plural corpora) is a collection of electronic texts or "a text in the computer-readable form" (Wray & bloomer, 2012, p. 205), written or spoken which is usually stored on the computer. It was in the past associated with a body of work, let say of one author (O'keefee, McCarthy & Carter, 2007). It is a representative collection of language that can be used to make statements about language, it inextricably concerned with how people used language in the contexts (Crawford & Csmomay, 2016). What makes it peculiar is that it is not just a mare lists of words but a guided and principled collection of text of particular language, usually used for quantitative and qualitative analysis to provide language users with available data on how language is used at micro (lexical) and macro (sentential) linguistic levels and help them to enhance language knowledge at frequency levels (keywords analysis and number of collocations) and beyond (cluster analysis, concordance and semantic relation analyses) (Biber, Conrad & Reppen 1998; O'keefee, McCarthy & Carter 2007). It is a useful tool that enables researcher/user to 'make a meaningful comment about some aspect of the language' in one's data set (Wray & bloomer, 2012). Anthorny (2012) construes that corpus linguistics is inseparably associated to computer technology in more essential way than other applied linguistics field, with exceptions of perhaps CALL (Computer Assisted Language Learning).

The first concordance developed in language study was presumably that of Vulgate Bible in the 13th century which consisted the 'list of almost every word in the bible and the point where it was used' compiled by 500 monks under the direction of Hugh of St. Cher, 'whose interest was in biblical exegeses. Probably, with the advent of the computer, Hugh can do this work single handily. Most corpus linguists are interested in the applications of the corpora rather than corpora per se, in the last century, in which Reference corpora and British National Corpus (BNC) laid the new foundation for creating dictionaries and for the translation purposes. Moreover, Sinclair (1991) was widely influenced work in applied as well as descriptive linguistics, leading to a view of lexis and structure as a continuum rather than an opposition. Corpus analysis provides an empirical basis or much contemporary research on language, employing data-based methods which emphasise statistical regularities rather than combinatory rules (Aston, 2011).

3.1 **Corpus Linguistics: A science or a methodology?**

Corpus linguistics is believed to be one of the fastest growing areas of language studies, more prominently in the last three decades. Nevertheless, corpus linguists heralded with sheer contended issues, recurrently concerning the status of corpus linguistics 'whether it constitutes a branch of linguistics, a method, a methodology, a discipline, an approach, or something else'. Therefore, there is a mixed reaction whenever the question about the status of corpus linguistics is brought before the corpus linguistics pioneers, who withhold different stands on that issues, for instance neo-Firthian scholars such as Tognini-Bonelli (2001) arguably postulates corpus linguistics on 'scientific panorama' and rejecting any view that foresaw it from 'methodological panorama' as she assertively posits that corpus linguistics attained theoretical status and defined set of rules and guiding principles, 'that distinguishes it from other branches of linguistic'(Sinclair, 2004), while a sheer number of linguists strongly deem it as a method no more no less. At this juncture it worth mentioning that some corpus linguists stayed at intermediary position believing that corpus can be seen as both science and methodology depending on who used it and the purpose it served while others strongly argued that the two terms used to describe corpus as "method" or "science" seem to be misleading and inappropriate.

Viana, Zyngier, and Barnbrook (2011) extensively provide some insights on the issues of the status of corpus linguistics in the series of interviews with professionals in the field of corpus linguistics composed and titled "*Perspectives on Corpus Linguistics".* To mention some among the contributors are Devies (2011), who arguably posits that corpus linguistics is a methodology rather than a science or even a separate field of linguistics. According to him corpus has implied set of laid down procedures based on frequency and functional approaches of linguistics that dealt with data at both micro and macro levels. Though he believes as a methodology it is always related to other important language approaches to complete a meaningful co-occurrence to give a meaningful result in language analysis. Th. Gries (2011), sees corpus linguistics as a methodology or methodological paradigm to him it is no more any less than that. He proposes some distinctions between corpus Linguistics and computational linguistics as the two overlapped.

Along the same Hyland and Tse (2005), considers corpus linguistics as methodology or a research tools used to analyses huge data which would be laborious or 'impossible to access by observational techniques'. Although, the question about status of Corpus Linguistics as either science or methodology seems to be deceptively tricky trick, since basically, the two differed as science is a systematic way knowledge condensed into rules and theories capable of producing a reliably-predictable outcomes, science implies perspective on reality while method implies perspective based on assumptions or evaluations of the observable instances.

Swales, (2006) concludes his stake as he views corpus linguistics as a methodology which he defines as a way of looking at large bodies of language data for wider varieties of purposes, (historical, critical, pedagogic, etc.). He rejects the notion of seeing it as the new branch of linguistics. Scholars like Aston (2011), Baker (2011), and Johansson (2011) in their interviews compiled by Viana et al. (2011) maintain a neutral stake about the status of corpus linguistics as either science or methodology For Aston corpus linguistic can only be realised as both science and methodology, although it is predominantly concerned with applications, that it is more of methodological per se, where its use is guided primarily by concerns of practical effectiveness and

theory from other fields. On the other hand, corpus linguistics is a science inasmuch as it has a particular object of study, namely language as it is actually used in naturally-co-occurring speech and writing, with texts as primary data which differ with other traditions in linguistics (Sinclair, 2004).

As for Baker (2011), Corpus Linguistics embodied the aspects of science such as gathering and organizing knowledge into testable laws and theories through observation, 'experimentation', 'measurement' and capable 'replicability' by another researcher. And it is at the same time be seen as methodology inasmuch as it contained a set of procedures and principles such as 'balance', 'representativeness' and 'sampling techniques in data collection as well as set processes in data analysis including annotations, wordlist, and concordances.

On the other hand, Johansson (2011) rejects the notion of describing corpus linguistics as either 'theoretical or applied', 'science or methodology' and according to him such restrictions to a particular user or application is uncalled for. He distinguishes three different uses of the corpus by different scholars: corpus-informed studies; corpus-based studies; and corpus-driven studies. McEnery and Hardie (2012) ascribe these categorisations views as the sources of diversity among corpus linguists.

On the other hand, in the interviews with another set of corpus linguists such as Conrad, Laviosa, Leech Loun, Sampson and Scott (Viana et al., 2011) argue that corpus linguistics is neither science nor a methodology, but described it using other terminologies. Conrad (2011) for example, refers to corpus linguistics as an approach to studying language all other terms seem inaccurate to it. The term approach is used in the sense that "all corpus works share certain general characteristics and a certain research philosophy" (Conrad, 2011, p. 49), and encompass great diversity in research purposes and methods. Therefore, referring to corpus linguistics as methodology or a separate field of science according to her is misleading since corpus coped with diversities at the essence of different linguists.

Henceforth, Laviosa (2011) advocates that as linguists we can reach consensus to the fact that 'corpus linguistics' is a new kind of research domain, an immensely important development of descriptive linguistics and a new approach to language studies. According to Leech (2011) corpus linguistics is neither a methodology pure or simple nor a pure science/scientific domain, but it is rather a methodologically-oriented branch of linguistics than the scientific domain. He adds that "using corpus has led to the development of growing collection of computer tools for searching, retrieving, annotating, and analysing, electronic text data: concordancers, parsers and so forth" (2011, p. 157). Louw (2011) supports the view of corpus linguistics as neither methodology nor science, but for him, it is best to be referred to as an instrument of science in the hand of Sinclairean.

Sampson (2011) argues that it is misleading to think of corpus linguistics as a branch of linguistics, like other branches of linguistics such as psycholinguistics, philology, sociolinguistics or historical linguistics. Scott (2011) chooses to use the terms tool and resource instead, referring to corpus linguistics. In the same vein, he does not recognise it as a separate science but more close

to the methodology. He asserts that the sheer power of tools and corpora have transformed and modelled the approaches of linguistics analysis both quantitative and qualitative.

3.2 Corpus-based approach and corpus-driven study

Corpus linguistics is a field of study that can cover all aspects of language through investigation of the casual use of a corpus to obtain a suitable authentic illustration language variation, linguistics elements as well as pragmatic and discourse-related phenomena. The different application of corpus analysis leads to the categorisations of corpora studies to general Corpus-based or corpus-driven models of how language works (Viana et al., 2011). Henceforth, linguists withhold diverse presuppositions concerning this issue most of them are often claim that their work to be as either corpus-driven or corpus-based. In addition to that, Francis (1993) is the first person to start using the terms corpus-driven and data-driven in which he emphasises the fundamental differences between corpus-based and corpus-driven approaches. As he postulates the term corpus in the latter approach is the main informant the only reliable authority, while in the former sense the term corpus is associated with basic principles and process of language based on the naturally occurring language. In the early day of corpus linguistics, it was certainly important to emphasise that things can be discovered in a corpus which can never be imagined by introspection. Almost twenty years later, Tognini-Bonelli (2001) revives the debate on the distinction between corpus-based and corpus-driven in her famous book '*Corpus linguistics at work (2011)*'. She explores different functions of the corpus as either theory or methodology in relation to corpus-based or corpus-driven analysis. In her conception corpus-driven concerned with the application of corpora for different pedagogical purposes.

Moreover, the distinction is intended to signal whether data are used (merely) to illustrate or test old categories of linguistic order, which have been taken from earlier linguistic theory (this is corpus-based analysis), or whether it is possible to induce new findings from sequences of raw textual data, and theory avoid assumptions and self-fulfilling prophecies (this is corpus-driven analysis). According to Hasselgård, Ebeling, and Ebeling, (2013) the corpus-driven concept is clearly related to the concept of induction, although in all the discussion about the corpus-driven approach, there is hardly any reference to the intensive debate about induction over the past 400 years or so. The concept of is usually attributed to Bacon in the 1600s, though in fact, it goes back much further. In the 1700s, Hume expresses scepticism of the concept, since what has happened in the past cannot guarantee what will happen in the future. From the 1930s onwards, this scepticism was expressed even more strongly by Popper (1975, cited in Hasselgård, Ebeling, & Ebeling, 2013) who argues that induction is simply a myth.

4. The Use of Learner Corpora in L2 Classroom

According to Vannestal and Lindquist (2007) linguists have been using corpora for pedagogical purposes for more than two decades. As a methodology based approach with a set of rules and principles, corpus linguistics is primarily concerned with electronic tools used to provide an authentic linguistic bank of data that could help in investigating language at the hand and promote pedagogical understandings in English Language Teaching (ELT) (Proctor 2012). Kennedy (2014) acknowledges the link between studies on corpus linguistics to second language pedagogy back to the early half of the twentieth century. This initially started with the use of corpora to generate vocabulary lists for foreign learners (Vannestal & Lindquist, 2007).

Learner corpora are relatively recent development in linguistic studies. Flowerdew (2013) defines learner corpora as "systematic collections of learner language data with aim of comparing learner usage with native, or expert, usage" (p. 171). Meanwhile, Granger (2003) defines corpora of learner language as "electronic collection of authentic texts produced by foreign or second language learners" (p.34).

The main importance of learner corpora is that they are more generally allied with empirical data' which makes possible for linguists to study the aspects of language objectively with the language as a central object (McEnery & Wilson, 1997). According to Flowerdew (2012) "most of the learner corpora are collected from learners of a particular background so that the distinctive interlanguage of those speakers (or writers) is represented and can be compared with expert usage or from other L2s" (171). For Gaviola and Aston (2001 cited in Baker 2012) 'using corpus findings to inform language teaching, and the actual use of corpora in language classroom has been a topic of much interest and discussion'. Tim Johns, was early pioneer in bringing the corpus into classrooms (Rappen, 2001 cited in Baker, 2012, p. 206), whom in his work 'data-driven learning' approach secured to exploit relatively small corpora to generate concordances from which learners could work out the linguistic regularities or themselves (Viana et al. 2011, p.1).

The development corpora of learner studies are always associated with application of computer for pedagogical and learning purposes, generally known as Computer Assisted Language Learning (CALL) as McEnery and Wilson (1996, p., 105) posit that "a computer system based upon a parsed corpus database is already being used in the teaching of English Grammar at the University of Nijmegen" and many subsequent studies used developed software either in teaching language to the learners or for analysing language produced by language learners. Moreover, Sinclair's (2004) work on how to use Corpora in Language Teaching immensely contributes to the development of corpora pedagogical purposes and help readers/researchers with quite a number of discussion and hands-on as well as practical activities covering a range uses of corpora in the classroom from various research findings.

Corpus is mostly used in Language Teaching (language pedagogy) for the availability of various electronic tools with the advent of computer and internet that can aid learners in studying the nature and different uses of language through frequency value, concordance hits, expended contexts, register types, collocations and syntactic patterns. Corpus-based language learning/teaching methods have revolutionised the traditional maxim about the method as they provide learners with the frequency value of lexical structures. At the same time, this approach forcibly redefined and refined the way and manner some language concepts are viewed or learned, for example, lexis and structure are learned as a correlated unit rather than piecemeal or separate linguistics units. John (1991) views of the leaner as researcher is complemented by a tendency for teaching materials to adopt syllabuses exploiting corpus-based research, privileging those features which appear most frequent in native speaker corpora, or most problematic in corpora o learner data.

Another considerable feature of pedagogical corpora is the size of the corpus. Based on the size corpus was classified as either small or large, most researchers advocate the use of small corpus in language pedagogy and learning. This is why the use of large corpus in the classroom is discouraged and ignored by language researchers (Proctor, 2012). However, there is no consensus

among the scholars in determining the size of the corpus that could be considered for pedagogical purposes (Sinclair, 2001). Most studies considered the use of corpus for specific language purposes which may consist from 20,000-words to 250,000-words (Ghadessy et al., 2001). Moreover, Flowerdew (2012) adds that in two decades or more ago 100,000-words is regarded as small corpus but nowadays even one million-words may be seen as small corpus not very large at all.

Corpus as a guided teaching of language from studying the linguistics pattern in more general passion rather than units, this will help ESL/EFL beginners to enhance their knowledge of structure of a particular language as one of their basic needs as beginners is to interact with natural occurring language data (Krashen's (1981) second language acquisition theory)', this can be provided through corpus-driven approach, "exposing to authentic English and producing native-like English through corpus are of significance for many EFL students as beginners or intermediate ones" (Proctor, 2012, p. 5). O'Keefee, McCarthy, and Carter (2007) add that the book of Schmitt and Schmitt (2005) can serve as a clear example of using corpora in the study of Academic English, each unit of the book was based on 'a set of target words taken from Coxhead's Academic Word List and all target words are presented explicitly at the beginning of each unit which enables the user to conduct a self-test.

Many studies learner language especially those done to investigate the EFL learners revealed that learners at almost all proficiency levels have difficulty in using the high frequency verbs such as '*make'*. And suggest that concordance-based exercises can help raise awareness of the complexity of high-frequency verb (O'Keefee et al., 2007). There are numerous studies conducted using learner corpora spoken or written to investigate language use by language learners.

Reppen (2010) provides a step-by-step procedure on how to use corpora on to the classroom and for pedagogical purposes, which involved identifying the 'structure or language features' for instruction. The student needs to practice. It should be a goal oriented that inform the teacher about the student level of proficiency. Flowerdew (2012) construes that learner corpora can be used in L2 error analysis, as a process which involved annotating features of learner interlanguage (errors), as such a comparison of the learner and expert corpora can reveal inappropriate use and also over-and under-use of given features on the part of the learners.

5. Corpus as a research paradigm

According to Proctor (2012) corpus linguistics is "a methodology based on electronically collected which opens up a new dimension language research" (p. iii). The use of language corpora in research has revolutionised the process of dictionary-making, language translation studies, traditional stylistic analysis, lexicographic and grammatical studies (O'keefee, McCarthy, & Carter, 2007), a corpus-based Sinclair project (1987-2003) COBUILD has great impact on studying grammar based on lexico-grammatical research, Hoey (2012) lexical priming Louw (1993) semantic prosody. Rappen (2010) postulates that:

> Research from corpus linguistics can provide a great deal of useful information to language teachers/researchers that can be used to inform course planning. For example, information from corpus linguistic research can provide insights as to what features of spoken language

students will frequently hear outside the classroom, or what grammatical features students will encounter in the different types of texts that they will be reading or writing. (p.10)

Corpora could be used as a meaningful source of data in an empirical language study. McEnery and Wilson (1996) note the importance of corpora in language teaching as believe that they can be closely allied to the important means of generating empirical data corpora help linguists to be more objective and language centred in their view about language, rather than being subjective depending on individual's perception on language. Corpora could be important tools in language studies and played more generous roles in collecting samples of language varieties, dialects, and for translations purposes. Corpus enables us to collect a broad sample of speech or huge bank of written texts across varieties of genres from which we can make an analysis using co-variables of the status of the learners like age, sex, class, learning a language as second or as foreign. In the case of lexicography corpora studies have changed the way in which lexicographers- and other linguists interested in the lexicon –can look at language as it provides up-to-date information about language in more organised and meaningful way of analysis both quantitatively and qualitatively.

Flowerdew (2012) posits that to understand how corpus linguistics is applied in language pedagogy distinctions can be made between direct and indirect approaches to corpus studies. Indirect applications are in the sense of using corpus as a research method in the process of dictionaries, syllabuses, textbooks, and teaching materials productions. For example the production of Collin COULD English Dictionary; Collins English Grammar textbook under the guidelines of Sinclair (1991); Macmillan Dictionary of English for advanced learners with Hoey as a chief adviser; the Longman Grammar of Spoken and Written English (Biber, et al., 1996) Cambridge Grammar of English (Carter & McCarthy, 2006). Indirect applications also involve the use of corpus in English for specific purposes (ESP), English for Academic Purposes (EAP). On the other side, the direct approaches of corpus linguistics to pedagogy can be sum up of what is referred to Data-Driven Learning (DDL) where students participated as researchers and worked directly with corpora with teachers in the' help in the interpretation of corpus data (Flowerdew, 2012).

6. Conclusion

It can be recalled that many presuppositions have emerged about corpus linguistics as discussed somewhere above in this article. Thus, corpus linguists have different views about the status of the corpus as a field of study or as a methodology or tool used in the analysis of language. Some of the perceived as a method or methodological panorama while some scholars believed that it could be a science on its own, some corpus linguists reluctantly partook an intermediary position and considered it to be partially a methodology and partially a field of enquiry on its own; consequently others rejected both of the notions referring to corpus linguistics as methodology or as science, they believed that the two terms are uncalled for and seem to be misleading. Corpus linguists observed that corpus linguistics is not directly a study about a particular aspect of language. Rather it is an area of which focuses upon a set of procedures, or methods, for studying a language or as characterised by others as a field of inquiry. Moreover, corpus linguistics is heterogeneous field not monolithic as such it focuses upon a group of methods and procedures for studying language, nevertheless, these methods and procedures themselves are still developing and remain an

unclearly delineated set-though some of them, such as analysis of concordance lines, are well established and regarded as central to the approach.

Based on this critique discussion we posited our intermediary stake concerning the corpus, as such by revisiting the definition as a collection of naturally occurring language text chosen to characterise a state or variety of a change corpus linguistic sound clearly as method, notwithstanding, the collection of samples of language is not just mare generating of word lists or concordances, but a guided and principled collection of text of particular language, with applied parameters, such as gathering and organizing knowledge into testable laws and theories through observation, experimentation, measurement and capable replicability by another researcher, by this definition, it is sound more of having scientific basis. Hence, corpus linguistics can be seen as both science and methodology depending on who used it and the purpose it served.

The approach to corpus linguistics can be direct or indirect. Where students participated as researchers and worked directly with corpora. Indirect applications are in the sense of using corpus as a research method in the process of dictionaries, syllabuses, textbooks, and teaching materials productions. Findings from learner corpus analysis can be applied to dictionaries, grammars and the design of the syllabuses. Learner corpus is an important tool for testing student performance, and relevance of a particular target language features to students' background. It helps teachers to know whether a particular target feature is difficult or not. It is more motivating and relevant for teachers to develop their own corpora from regular classes for immediate pedagogical use. Finally, we need to add that the term corpus in applied linguistics is generally concerned with the collection of naturally occurring texts that is the language that was generated by real speakers and writers rather than the language that was invented solely for the purposes linguistic analysis and argumentation.

About the authors:
Ibrahim Bashir: is a PhD candidate of English Language Studies specializing in Corpus Linguistics at Faculty of Languages and Communication, Universiti Sultan Zainal Abidin, Malaysia. He holds a master degree in TESL from the same university graduated in 2016. His areas of research interest include, applied linguistics, pragmatics and corpus linguistics. ORCID: https://orcid.org/0000-0003-2360-8468

Dr. Kamariah Yusus: BA (El Paso), MA (Newcastle) PhD. University Malaya, was a senior lecturer and deputy dean academic and graduate affairs at Faculty of Languages and Communication of the Universiti Sultan Zainal Abidin. Her areas of expertise include: linguistics, applied linguistics and corpus linguistics.
Bashir Ibrahim: is a M.A. English Language Studies student at Faculty of Languages and Communication, Universiti Sultan Zainal Abidin (UniSZA), Malaysia. He holds B.A. English from Al-Qalam University Katsina, Nigeria. His areas of research including, semantics, cognitive linguistics and applied linguistics.

References
Alkhataba, E. H. A., Abdul-Hamid, S., & Bashir, I. (2018). Technology-supported online

writing: An Overview of six major Web 2.0 Tools for collaborative-online writing. *Arab World English Journal, 9* (1), 433-446.DOI: https://dx.doi.org/10.24093/awej/vol9no1.30

Anthony, L. (2012). Issues in the design and development of software tools for corpus studies: the case for collaboration. In Baker P., (2012). *Contemporary corpus linguistics*. New York: Continuum International Publishing Group.

Aston, G. (2011). Applied corpus linguistics and the learning experience. *Perspectives on corpus linguistics*, *1*(2), 1-16.

Baker P., (2012). *Contemporary corpus linguistics*. New York: Continuum International Publishing Group.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.

Bloor, T. & Bloor, M. (1995). *The functional analysis of English* London. New York: Arnold.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*. Cambridge, UK: Cambridge University Press.

Conrad, S(2011). Variation in corpora and its pedagogical implication. *Perspectives on corpus linguistics*, *1*(2), 47-62.

Crawford, W., & Csomay, E. (2016). *Doing corpus linguistics*. New York: Routledge Tailor and Francis Group.

Devies M. (2011). Synchronic and diachronic uses of corpora. *Perspectives on corpus linguistics*, *1*(2), 63-80.

Ellis, R. & Barkhuizen, G. (2009), *Analysing learner language*. China: Oxford University Press.

Ellis, R. (2012). *The study of second language acquisition*. China: Oxford University Press.

Flowerdew, J. (2012). *Discourse in English language education*. New York: Routledge Tailor and Francis Group.

Firth, J. R. (1968). *Selected papers of JR Firth, 1952-59*. Bloomington, Indiana, USA: Indiana University Press.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. *Text and technology: In honour of John Sinclair*, *1*, 137-156.

Gass, S.M. & Selinker, L. (2007). *Theories of second language acquisition: an introductory course*. Abingdon, London, UK: Routledge.

Gika, S. (1996). Foreign in English language teaching: meaning, role, and appropriateness. *IATEFL Newsletter*, *130*(1), 14-17.

Granger, S. (2003). The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. *Corpus-based approaches to contrastive linguistics and translation studies*, *1*(1), *20*, 17.

Groom, N. & Littlemore, J. (2011*). Doing applied linguistics: a guide for students*, New York: Routledge.

Ghadessy, M., Henry, A., Roseberry, R.L. (Eds.), 2001. *Small corpus studies and ELT*: *Theory and practice*. Amsterdam, Philadelphia: John Benjamins Publishing.

Gries, S. (2011). Methodological and interdisciplinary stance in corpus linguistics. *Perspectives on corpus linguistics*, *1*(2), 81-98.

Hasselgard,H. et al. (2013). *Corpus perspectives on patterns of lexis: studies in corpus*

linguistics 57, John Benjamins Publishing Company, Amsterdam/Philadelphia

Halliday, M. A. K. (2007). *Language and education* (Vol. 9). London and New York, USA: Continuum.

Halliday, M. A. (1993). Towards a language-based theory of learning. *Linguistics and education*, *5*(2), 93-116.

Halliday, M. A. K., Matthiessen, C., & Halliday, M. (2014). *An introduction to functional grammar*. New York, USA: Routledge.

Hasselgård, H., Ebeling, J., & Ebeling, S. O. (Eds.). (2013). *Corpus perspectives on patterns of lexis: studies in corpus linguistics 57*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund-Prytz, Y. (2008). *Corpus linguistics with BNCweb-a practical guide*(Vol. 6). *Frankfurt*, New York, USA: Peter Lang.

Hoey, M. (2012). *Lexical priming: A new theory of words and language*. New York: Routledge Tailor and Francis Group.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for specific purposes*, *24*(2), 123-139.

James, C. (1998). *Errors in language learning and use*. USA: Addison Wesley Longman.

John, P. D. (1991). A qualitative study of British student teachers' lesson planning perspectives. Journal of Education for Teaching, *17*(3), 310-320.

Johansson, S. (2011). A multilingual outlook of corpora. *Perspectives on corpus linguistics*, *1*(2), 115-130.

Kachru, B. B. (1992). World Englishes: Approaches, issues and resources. *Language teaching*, *25*(1), 1-14.

Kennedy, G. (2014). *An introduction to corpus linguistics*. London and New York: Routledge.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford, UK: Oxford University Press.

Laviosa, S. (2011). Corpus-based translation studies: Where does it come from? Where is it going. *Corpus-based translation studies. Research and applications*, *1*(2), 13-32.

Leech, G. (2011). Principles and applications of corpus linguistics. *Perspectives on corpus linguistics*, 1(2), 155-170.

Louw, B. (1993). Irony in the text or insincerity in the writer? — the diagnostic potential of semantic prosodies. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: in honour of John Sinclair* (pp. 157–176). Amsterdam, Philadelphia: John Benjamins Publishing Company.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. UK

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh, UK: Edinburgh University Press.

McEnery, T., & Wilson, A. (1997). Teaching and language corpora (TALC). *ReCALL*, *9*(1), 5-14.

O'keefee A. et al. (2007). *From corpus to classroom: Language use and language teaching*. Cambridg, UK: Cambridge University Press.

Philip, G. (2011). *Colouring meaning: Collocation and connotation in figurative language*. John Benjamins Publishing.

Proctor F. (2012*). Functional and corpus approaches of English grammar*. USA: Nyx Academics LLC.

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.

Richards, J. C. (1974). *Error analysis, perspectives on second language acquisition*. London, UK: Longman.

Richards, J. C. (1992). Content knowledge and instructional practice in second language teacher education. In J. E. Alatis (Ed.), *Georgetouln~UniversityR ound Table on Languages and Linguistics 1*(1), 76-99. Washington, DC: Georgetown University Press.

Richards, J. C., & Schmidt, R. W. (2014). *Longman dictionary of language teaching and applied linguistics*. New York, USA: Routledge.

Sampson, G. (2011). A two-way exchange between syntax and corpora. *Perspectives on corpus linguistics*, *16(1)*, 256.

Scott, M. (2011). Th technological aspect of corpus linguistics. *Perspectives on corpus linguistics*, *1*(2), 213-220.

Sinclair, J. (1991*) Corpus, Concordance, Collocation*: Describing English Language, Oxford, UK: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. New York, USA: Routledge.

Swales, J. M. (2006). Corpus linguistics and English for academic purposes. In *Information technology in languages for specific purposes*. *1*(2), 19-33. Springer, Boston, MA.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). Amsterdam, Philadelphia: John Benjamins Publishing.

VanPatten, B., & Williams, J. (2015). *Theories in second language acquisition: An introduction*. New York, USA: Routledge.

Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, *19*(3), 329-350.

Viana, V., Zyngier, S. & Barnbrook (2011). *Perspectives on corpus linguistics*: *studies in corpus linguistics 48*, John Benjamins Publishing Company, Amsterdam/Philadelphia

Wray A. & Bloomer A. (2012). *Project in linguistics and language studies: A practical guide to researching language (3rd ed.).* Great Britain: Hordder Arnold Publication.

Ziggiotto, A. (2016). Learning to write at university level in Italy: A longitudinal corpus-based study of interpersonal and textual metadiscourse. (Unpublished master dissertation). Università degli Studi di Padova.