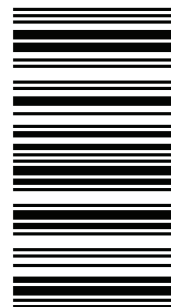


Dynamiques sociolinguistiques et didactiques en Afrique

Cet ouvrage analyse les dynamiques sociolinguistiques et didactiques à partir de l'expérience de quelques pays africains. Les dynamiques sociolinguistiques révèlent la nécessité d'harmonisation des alphabets de certaines langues transfrontalières comme le yoruba parlé aussi bien au Bénin, au Nigeria qu'au Togo. Les paysages linguistiques des pays africains manifestent plusieurs dynamiques : entre les langues africaines d'une part, et les rapports entre les langues africaines et les langues de colonisation d'autre part. La deuxième partie de l'ouvrage analyse les dynamiques didactiques qui traitent des particularités lexicales dans les manuels de lecture, des rapports que la culture orale africaine entretient avec la production écrite en langue française, des sens et des origines des noms de personnages dans les œuvres littéraires produites en Afrique. Cette situation évoque des questions relatives à l'enseignement / apprentissage du français qui prennent en compte les réalités socioculturelles de l'apprenant. Des activités d'apprentissage du français à partir des contes africains méritent d'être proposées dans le but de consolider ses valeurs culturelles endogènes.

Moufoutaou ADJERAN est titulaire d'un doctorat unique en sociolinguistique (Université d'Abomey-Calavi) et Maître-Assistant des Universités (CAMES). Il est enseignant-chercheur de sociolinguistique à l'université d'Abomey-Calavi en République du Bénin. Il travaille sur les variétés du français, les contacts de langues, les phénomènes variationnels.



978-3-8381-7035-0

Dynamiques sociolinguistiques

Adjeran (Éd.)

Moufoutaou Adjeran (Éd.)

Dynamiques sociolinguistiques et didactiques en Afrique

Herméneutique de la situation de quelques pays



presses
académiques
francophones

Moufoutaou Adjera (Éd.)

**Dynamiques sociolinguistiques et
didactiques en Afrique**

Herméneutique de la situation de quelques pays

Presses Académiques Francophones

Imprint

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher:

Presses Académiques Francophones

is a trademark of

International Book Market Service Ltd., member of OmniScriptum Publishing Group

17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page

ISBN: 978-3-8381-7035-0

Copyright ©

Copyright © 2017 International Book Market Service Ltd., member of OmniScriptum Publishing Group

All rights reserved. Beau Bassin 2017

THE USE OF CORPUS LINGUISTICS FOR INFORMING ENGLISH LANGUAGE TEACHING IN EFL SETTINGS

**Jean Cosme N. DOSSOU,
Estelle BANKOLE-MINAFLINOU,
Etienne K. IWIKOTAN**
Université d'Abomey-Calavi (Bénin)

Abstract

The advent of corpus linguistics revolutionized the world of linguistics and language teaching. This paper aims at exploring some theoretical considerations about corpus linguistics and explaining the relevance of corpus linguistics to the teaching of English as a foreign language. Even though the first computer corpus was generated only in the 1960s, a lot of corpora such as the ARCHER Corpus, the Bank of English and the British National Corpus are available online today. These corpora enable to study grammatical and vocabulary items across many registers and to find out how those items are used by native speakers in different contexts. For instance, Melrose (1983) found that in oral communication, while Americans use “have to” more frequently, the British prefer “must”. Furthermore, the use of corpus linguistics has significantly improved the teaching of English as a foreign language. That is the reason why Richards (2001) encouraged EFL teachers to familiarize themselves with computer technology so as to be able to carry out corpus research for their classroom practices. The use of corpora enables teachers to get authentic materials for their lessons and avoid the contrived language used in most EFL textbooks. Finally, corpus linguistics has enabled lexicographers to publish corpus-based dictionaries such as The Longman Dictionary of Contemporary English which often provide the most common words or constructions with which the words explained occur. These dictionaries have improved the teaching of vocabulary, especially in EFL settings.

Key words: Corpus Linguistics, Communicative Language Teaching, authentic communication.

Résumé

L'avènement de la linguistique de corpus a révolutionné le monde de la linguistique et de l'enseignement des langues. Le présent article vise alors à explorer certaines considérations théoriques relatives à la linguistique de corpus et à expliquer la pertinence de la linguistique de corpus pour l'enseignement de l'anglais langue étrangère. Quoique le premier corpus informatique ait été créé dans les années soixante seulement, beaucoup de corpus comme ARCHER Corpus, Bank of English et British National Corpus sont actuellement disponible en ligne. Ces corpus permettent d'étudier des éléments de grammaire et de vocabulaire à travers plusieurs registres et de voir comment ces éléments sont employés par les natifs de la langue. Par exemple, Melrose (1983) s'est rendu compte que dans la communication orale, les Américains utilisent "have to" plus fréquemment alors que les Britanniques préfèrent "must". En outre, l'utilisation de la linguistique de corpus a amélioré de manière significative l'enseignement de l'anglais langue étrangère. C'est pour cela que Richards (2001) a encouragé les enseignants à se familiariser avec la technologie informatique afin de faire de la recherche sur les corpus pour leurs pratiques pédagogiques. L'utilisation des corpus permet aux enseignants d'obtenir du matériel authentique pour leurs cours et d'éviter le langage artificiel utilisé dans la plupart des livres d'anglais langue étrangère. Pour terminer, la linguistique de corpus a permis aux lexicographes de publier des dictionnaires à base de corpus tels que The Longman Dictionary of Contemporary English qui indiquent les mots et les expressions les plus courants avec lesquels on retrouve les mots expliqués. Ces dictionnaires ont amélioré l'enseignement du vocabulaire, notamment dans les milieux où l'anglais est enseigné comme langue étrangère.

Mots-clés: Linguistique de corpus, Approche communicative de l'enseignement des langues, Communication authentique.

Introduction

It has widely been observed that the English language that teachers teach learners in English as a Foreign Language classroom settings often differs from the English language that people use in the real world (Carter, 1998). The fact is that in foreign language teaching, the patterns of lexicon and grammar that people usually exploit to perform their communicative acts are not connected to the rationale behind the communication, the environment where it occurs, the participants involved, and the language type or mode of their communication.

English Language Teaching usually fails to incorporate into classroom methodologies and curricula the features that frequently emerge in naturally occurring language (Thornbury, 2004). The corollary of such an approach to language teaching is that despite several years of formal instructions in the classroom, learners remain deficient in the ability to use the English language for meaningful interactions and to comprehend how it is used in real world contexts (Widdowson, 1972). This failure of foreign language learners to use their target language as a means of communication, however, sharply contradicts the philosophy behind the emergence at the turn of the past century of the present-day Communicative Language Teaching whose ultimate goal demands that learners, especially those at the intermediate and advanced proficiency levels, be made better communicators before the end of the learning process at school (Brown, 2007b). How can learners become better communicators and use their target language meaningfully and appropriately in various circumstances if teachers' teaching practices and styles and the language input they derive from various curricula and syllabi remain to this day prescription-driven and have almost no relevance to the language in use?

In view of the inappropriate and inadequate language to which foreign learners are habitually exposed in their settings and the infrequent opportunities they are often presented for meaningful interactions in their target language (Fotos, 2002), Carter & Nunan propose that foreign language curricula "be carefully structured with extensive recycling of key target-language items...for the materials to provide the sort of rich and diverse linguistic input that ESL/[EFL] learners encounter in the world beyond the classroom" (Carter & Nunan, 2001, p.2). Their proposal appears to be of paramount importance primarily because most of the instructors who teach in EFL settings are non-native speakers with inappropriate and inadequate knowledge of the English language in use, and secondly because EFL learners often have infrequent

opportunities to interact meaningfully in their target language in and outside the classroom.

In the wake of the foregoing proposal by Carter & Nunan (2001), the present article seeks to explore some theoretical considerations about corpus linguistics and explain the relevance of corpus linguistics to the teaching of English as a foreign language.

1. Theoretical Considerations about Corpus Linguistics

For better insight into corpus linguistics, this section presents a historical background, defines the concept of corpus linguistics, and closes with instances of applications of corpus linguistics to the English language.

1.1 Historical Background of Corpus Linguistics:

Corpus linguistics has gone through a number of stages even though its historical contours are not clear-cut. It seems to have originated in Great Britain because Rajagopalan (2004) looks at it as a product of empiricism which is one of the major developments in the British thought. He further asserted that “the advent of generative linguistics in the 1950s actually put a damper on the work of corpus linguistics” (Rajagopalan, 2004, p. 411). It can be inferred from this statement that corpus linguistics had existed before generative linguistics or grammar and it might have started in the early 1950s or earlier. In the words of Rajagopalan (2004), “a corpus of data [was] manually assembled by a team of experts under the pioneering leadership of Sir Randolph Quirk” (p. 411) and it culminated in the publication of a book titled *A Comprehensive Grammar of the English Language* by Quirk et al. (1985). The publication of that book represented a major breakthrough for corpus-based linguistics as well as it confirms the idea that the first attempts to gather corpora date back to the 1950s.

As stated earlier, there was a lot of antagonism between corpus linguistics and generative linguistics which focuses mainly on the three adequacies

(observational, descriptive, and explanatory) and the minimalist theory (Chomsky, 1995; Meyer, 2004). Even though from first sight, one could say that it falls within the descriptive adequacy, corpus linguistics was severely criticized by the advocates of generative grammar. The criticism stems from the fact that the minimalist theory which is at the core of generative grammar deals with language as it is acquired by children “in advance of experience” (Chomsky, 1995, p. 4), not language as it is used in the real world, the raw material used by corpus linguists (Meyer, 2004). On the other hand, advocates of corpus linguistics blamed generative grammar for being too simplistic and for laying too much emphasis on explanation to the detriment of description (Chafe, 1994).

Despite the aforementioned antagonism, corpus linguistics continued to develop, especially with the advent of computer technology and its usage for linguistic purposes in the 1960s (McEnery and Gabrielatos, 2006; Meyer, 2004). In fact, the first computer corpus known as the Brown corpus was generated in 1964 (Meyer, 2004; Francis and Kučera, 1964). Following this first computer corpus, many other corpora such as the ARCHER Corpus, the Bank of English, the British National Corpus, the Freiburg-Brown Corpus of American English, the Freiburg-LOB Corpus of British English were also generated (McEnery and Gabrielatos, 2006). What is quite noticeable about these early corpora is that they cover only written English except for the British National Corpus which included a small percentage of spoken English. As it appears, the use of computers has modernized corpus linguistics to the extent that corpus linguistics is also known today as computer corpus linguistics (Leech, 1992). The scope of corpus linguistics has also increased considerably. Most corpora today contain millions of words and are used for linguistic description or analysis, lexicogrammar, the teaching of English as a foreign language, language change to mention but a few (McEnery and Gabrielatos, 2006; Stubbs, 2004; Meyer, 2004). Following this brief historical overview, it is now appropriate to define the concept of corpus linguistics.

2.2 understanding the Concept of Corpus Linguistics

Bernardinini defines a corpus as “a large collection or database of machine-readable texts involving natural discourse in diverse contexts” (Bernardinini, 2000 as cited in Huang, 2011, p.1) while Conrad (2005) defines it as “a large, principled collection of naturally occurring texts that is stored in electronic form (on a computer hard drive or server)” (p. 394). By naturally occurring discourse or texts, both linguists mean discourse or texts authentically produced by real language users for real communicative purposes as opposed to being created especially for the study or teaching of the language. The existing corpora available online include the *Corpus of Contemporary American English* (COCA) containing approximately 410 million words (<http://corpus.byu.edu/coca/>), the *British National Corpus* (BNC) beyond 100 million words (www.natcorp.ox.ac.uk/ or <http://corpus.byu.edu/bnc/>), Cambridge and Nottingham Corpus of Discourse in English (CANCODE), a multi-billion word corpus of naturally occurring English language (<http://www.cambridge.org/us/cambridgeen>) and the Michigan Corpus of Academic Spoken English (MICASE) featuring 1.8 million words of speech in various academic contexts (<http://micase.elicorpora.info>). Corpora are typically examined by means of computers. The function that the use of authentic, corpora-driven language plays appears to be significant in comprehending how language is used in authentic contexts, and such corpora constitute resourceful means that teachers and curriculum developers could exploit in English language teaching (Carter & Nunan, 2001). Below is what a corpus front page looks like online.

DISPLAY ?

LIST CHART KWIC COMPARE

SEARCH STRING ?

WORD(S) ?

COLLOCATES 4 4 ?

POS LIST -select ?

RANDOM SEARCH RESET ?

SECTIONS SHOW ?

1. IGNORE 2. IGNORE

SPOKEN
FICTION
MAGAZINE
NEWSPAPER
ACADEMIC

SPOKEN
FICTION
MAGAZINE
NEWSPAPER
ACADEMIC

SORTING AND LIMITS

FREQUENCY MUT IN ?

NIMUM MI FREQUENCY 10 10 ?

CLICK TO SEE OPTIONS ?

FREQ 100 KWIC

100

WORDS

Figure 1: COCA front page

Table 1: Variations of verb forms that follow the expression "It is time that"

EWS	N	Denver	A	B	C	Denver Post's article on Colorado's use of water was excellent. It is time that Colorado wakes up to where the water leaving the mountains really ends up.
EWS	N	Denver	A	B	C	of helping others but deny the commandments of Christ to His Church. It is time that evangelicals wake up. # Hartman seems to have the opinion of most Americans
CAD	A	World Affairs	A	B	C	they must first understand the religious teachings as they were originally intended. It is time that Muslims and Hindus are made aware that men and women were regarded as equal
SP		Fox_Cavuto	A	B	C	dividends goes down, I think that is

OK					highly likely. I think It is time that there is some kind of tax package or stimulus package that's put into effect	
OK	SP	CNN_ King	A	B	C	businesses and changing jobs out of the state. It is time,It is time that we change Gray Davis out of Sacramento! (END VIDEO CLIP) (
OK	SP	CNN_ Live From	A	B	C	more and more of our personal lives and our private lives. And It is time that we said, enough is enough. (CHEERING AND APPLAUSE) (END
EWS	N	Atlanta	A	B	C	the score up " on their opponents... I believe It is time that principals and athletics directors step in and reprimand coaches who do this on a
EWS	N	San Francisco	A	B	C	help the homeless, help our elderly, help in medical benefits. It is time that we stop the Napoleon complex of President Bush. He is not a good
CAD	A	Humanist	A	B	C	voice in what the ACGME does can nonetheless get involved. Indeed, It is time that all Americans stop becoming victims of abused and exhausted residents in teaching hospitals.
CAD	A	Humanist	A	B	C	stop becoming victims of abused and exhausted residents in teaching hospitals. And It is time that the government better regulate residency programs. # By Mimi H. Klaiman # #

Table 1 is a sample of an investigation carried out on the expression “*It is time that*” to see its frequency across the registers (News, Academic, spoken) and the verb form that native speakers and writers often attach to its use across the four registers of formality. The function that the use of authentic, corpora-driven language plays appears to be significant in comprehending how language is used in authentic contexts, and such corpora constitute resourceful means that teachers and curriculum developers could exploit in English language teaching (Carter and McCarthy, 2001). If what ‘a corpus’ is has unanimously been agreed upon in the spheres of research professionals, not all the researchers engaged in the use of corpus linguistics concur in what *corpus linguistics* actually means, and the issue has lent itself to hot debates in the midst of linguists and other researchers. On the one hand, researchers such as McEnery et al. (2006) define corpus linguistics as a methodology and a tool for enhancing research into linguistic disciplines such as lexicography, lexicology, grammar, discourse and pragmatics (Cheng, 2012). On the other hand, Tognini-Bonelli (2001) views corpus linguistics as a theory of language and a discipline in its own right. For

the purpose of this paper, *corpus linguistics* would simply be defined as “an approach to linguistic research that relies on computer analyses of language” (Brown, 2007a, p. 230). In line with this definition, Conrad (2005, p. 393) thinks of it as “an approach to research and teaching that makes use of computer-assisted analyses of language” while Cheng (2012) views it as the compilation and analysis of corpora.

Broadly defined, corpus linguistics reflects the firm belief that description of the language in use is more significant or essential to teach, study, and practice than the prescription of how language ought to be used. Therefore, corpus linguistics appears to be a descriptive approach to research and language teaching that examines and analyzes naturally occurring language from corpora in order to incorporate the linguistic features in current language use within the classroom styles and practices for communicative purposes. Its use in English Language Teaching conforms to the view held by *Contrastive Rhetoric* according to which language occurs not in isolated syntactic structures but rather in naturally occurring discourse, whether spoken or written. Of all the three definitions abovementioned, Conrad’s appears to be the sole explanation which directly points out the connection that exists between corpus linguistics and teaching in order to underscore its significance in English Language Teaching.

2.3 Major Findings Obtained Through the Application of Corpus Linguistics to the English Language

In order to tangibly show the significance and contribution of corpus linguistics in second and foreign language teaching, it is of necessity at this point that the instances of the findings at which a number of researchers have arrived through its application be adduced and highlighted as evidence.

Willis (1990), Tognini-Bonelli (2001), Mindt (1991) study the use of *any* in naturally occurring language in order to verify if this grammatical linguistic

feature functions in real-life discourse as the formal prescriptive grammar prescribes that it should be taught in the language classroom pedagogy. Contrary to the established prescription, they respectively found out that *any* is employed 42%, 46%, and 51% with positive structures across the registers in naturally occurring language. Biber et al (2002), on the other hand, found out that active-voice constructions, question-tags, get-passive, question-words, that-complement clauses and wh-complement clauses are features which exclusively remain specifically attached to the spoken mode while passive-voice constructions, adverbial/subordinate clauses, subject inversion, and the ordering of adjectives in premodification to the expository, written register. In addition, subsequent to in-depth investigations into the corpora of naturally occurring language, Biber et al (1999) found that almost 30% of the words in a corpus of conversation and about 20% in a corpus of academic prose are in three-or-four-word recurrent bundles (at least 10 times per million words, spread across at least five different speakers or writers). Melrose (1983), on the other hand, differentiates between *must* and *have to* since her investigations into corpora revealed that, unlike the input to which ESL/EFL learners are exposed in their textbooks, both linguistic features mark language varieties in conversational necessity-based contexts (Standard American English versus Standard British English) and register variations (spoken and written registers) and are not equivalent in semantic value (e.g., Excuse me, I have to go now (highly likely in everyday American spoken English versus Excuse me, I must go now (highly likely in everyday British spoken English)).

Much major research of the type mentioned above remains to be carried out to unlock the secret that the language in use still holds. It would therefore prove beneficial to highlight corpus linguistics to English Language Teaching practitioners and professionals in order that a wide and diverse variety of data about the English language in use might be obtained and incorporated into

ESL/EFL curricula and classroom methodologies to enable learners to produce authentic language in and beyond the classroom environments.

3. Corpus Linguistics and the Teaching of English as Foreign Language

3.1 The Rationale behind the Use of Corpus-based Descriptions for Language Teaching in Foreign Settings

Corpus-based descriptions for language teaching in foreign language settings would provide adequate and appropriate language to foreign language curricula and classroom discourse in order to keep learners abreast of language in use. In *New Perspectives on Grammar Teaching*, Celce-Murcia (2002) suggests that instructions be provided in context-embedded discourse which features authenticity and naturalness. Through the practice of using authentic corpora, EFL educators could compensate for the natural language that their linguistic environment lacks and “increase their learners’ levels of on-task behavior, concentration, and involvement in the target activity more than artificial materials” (Peacock, 1997; p. 152). Additionally, the use of corpus-based descriptions by EFL educators and students, especially those in higher education aspiring to become teachers, aims at encouraging critical and independent thinking if they carry out contextual analyses through corpus-driven research. The purpose of using contextual analyses through corpus-driven research is to test the validity of the structures that formal grammars prescribe in English Language Teaching pedagogy or to verify the accuracy of the established rules. Contextual analyses through corpus-driven research also provide answers to grammatical or lexical questions, and solve language problems for an in-depth comprehension of how the English language operates in real-life contexts. For instance, educators and student researchers might investigate the English language using at least 100 tokens of a target form or structure in a context-embedded discourse from COCA (Corpus of Contemporary American English) or from BNC (British National Corpus) to draw generalizations about the content where the target form occurs (or does not

occur), what it means, why it is used (or not used), the circumstances surrounding its occurrence and context of use. Research of the kind helps to differentiate between American English and British English at one level, or between the spoken register and the expository written mode at another level. Investigations through COCA (Corpus of Contemporary American English), for instance, reveals that *there is/there are* characterize expressions of description which are register-specific. The exploration of 1000 tokens for each of the expressions reveals that both language items are exclusively used for the academic register with only 40 tokens for Newspapers:

There is + singular noun/ There are + plural noun

However, *there's* + singular/ plural noun remains the unique alternative that the overwhelming majority of native speakers prefer to use in oral communication. The expression *there's* is hence perceived as a whole and indivisible unit that is not affected by number. The implication that such findings may have in second and foreign language teaching is that in the future, learners should be taught the rules that govern the register on which learning focuses attention to abide by the language in use.

Such a corpus-based research project provides evidence for how the English linguistic features *there's/there is/there are*, function in real spoken and written texts. This approach to language teaching helps to generate, or provide authentic data from, real language to show how the language is currently used across the four registers, namely spoken, fiction, academic, and newspaper. The findings of the research on the use of *there's*, *there is*, and *there are* confirm the status of existential *there* as subject in oral discourse to which Greenbaum (1990) alluded in *A Student's Grammar of the English language* on page 432: "Especially informally, *there* is treated like a singular subject where the 'notional' subject is plural'.

Corpus linguistics thus needs to be applied to EFL higher education programs and teacher training to enable students and teachers alike to carry out

research in order to gain and provide insights into the English language in actual use. Jack C. Richards (2001) exhorts teachers to familiarize themselves with technology, principally computer-based technology in order to be able to carry out corpus-based research for their classroom methodologies. Tomlinson (2001) and Breen (2001) disapprove of teachers being merely consumers of other people's materials and advise them to assume the role of syllabus planners, and curriculum and material developers.

In conformity with corpus linguists' view, the present article argues that pedagogical instructions should teach foreign language learners solely the language and the language items found in naturally occurring discourse. Naturally occurring discourse is mostly composed of recurrent patterns as McCarthy observed:

The creation of text and meaning in speech and writing ...works on the basis of the speaker/writer having at his/her disposal a large store of ready-made lexico-grammatical chunks. Syntax, far from being primary, is only brought into service occasionally, as a kind of 'glue' to cement the chunks together (McCarthy and Carter, 2004, p. 9).

The main goal that the research pursues in advocating the use of corpus linguistics in EFL pedagogy is thus to provide through the curricula and put at learners' disposal parts of the large store of the ready-made lexico-grammatical chunks that speakers and writers exploit to perform their everyday acts of communication. Any language or language item not included in the large store or extraneous to the English language in use could only be considered mythical and unrealistic, and viewed as such should not be suggested for learning in the foreign language classroom.

The rationale behind the use of corpus linguistics in foreign language teaching pedagogy is based on two principles:

Microscopic language teaching through the standard dogmatic prescriptive view, at best, only helps students learn at an arithmetic rate, i.e. 1→2→3→4→5→6→7→8 etc.

Macroscopic language teaching through a holistic descriptive approach like corpus linguistics, if well implemented, would develop learners' communicative competence at a geometric or exponential rate, i.e. 1→2→4→8→16→32 etc.

The argument against microscopic language teaching through the standard prescriptive view is that the move from 1 to 100 proves a difficult task, a long and arduous journey full of obstacles which, in most language learning cases, prevent the vast majority of learners, especially EFL learners, from attaining the final destination of the learning process, and exhaust their resources en route. Macroscopic language teaching through a holistic descriptive approach, on the other hand, exposes learners to authentic language in which they encounter several facets of language: pragmatics, grammar (structures), vocabulary (lexis), language functions, pronunciations (rhythm and intonation), collocations, colligations, lexico-grammatical patterns, discourse coherence and cohesion.

The rationale behind the former principle is that if the rate of the learning process is slow and tedious, it takes longer, perhaps forever, to complete, and thereby proves an unfruitful experience for learners. For instance, the formal classroom instruction that solely relies on developmental structures (e.g., the mastery of the negatives and interrogatives before the mastery of the tags) hardly produces any gains in spontaneous language use (Schumann, 1978; Rod Ellis, 1984; Kadia, 1988). On the contrary, if English Language Teaching abides by a holistic descriptive methodology like corpus linguistics and the learning process proceeds at a fast pace and does not induce any pains to learners, it is highly likely that before its completion the latter will achieve communication gains in the production of spontaneous language use. The standard prescriptive view of language learning operates at the microscopic level (chicken-nuggets methodology) and largely depends on intuitions which often prove unreliable or

inaccurate (Biber and Conrad, 2001; Biber and Reppen, 2002). Since it does not organically originate from naturally occurring language but emanates from an artificially standardized Latin model as applied to written language (Lyons, 1969), and seems not to provide a realistic account of the language it intends to describe, Pennington (2002) argues that it is not tailored to learners' communicative needs. A descriptive approach like corpus linguistics, on the other hand, operates at the macroscopic level (full-roasted-bird methodology) and heavily relies on empirical studies to draw conclusions which tend to represent generalized patterns. The assumption underlying the full-roasted-bird approach is that language occurs not in isolated syntactic structures but rather whole in naturally occurring discourses, whether the spoken register or the written mode.

3.2 The Contribution of Corpus Linguistics to Second and Foreign Language Teaching

The objectives that corpus linguistics seeks to achieve in English Language Teaching within the framework of Communicative Language Teaching (CLT) as a whole seem to focus more on “the big picture” in order to promote communication on a short-term and long-term basis. Lawson (2001) identified four areas of language in which corpus linguistics could provide significant insights. Firstly, investigations into corpora could provide information about the frequency of linguistic features in naturally occurring language. Moreover, investigations into corpora could provide information about register variations (how the use of particular language items varies across different contexts of formality, e.g., “and” used as a clause-level connector in conversations whereas in academic prose, it is typically used as a phrase-level connector (Biber et al. 1999, p. 228). In addition, investigations into corpora could feature the degree of reliability and scope in language use (low/high reliability or narrow/broad scope). From the perspective of language acquisition, language items with high reliability and broad scope are acquired before those with low reliability and

narrow scope (Hulstijn & De Graaff, 1994). In view of this hypothesis, Doughty & Williams (1998) suggest that language items with high reliability and broad scope be highlighted in language teaching prior to highlighting language items with low reliability and narrow scope. Likewise, investigations into corpora could provide information about the discourse properties of specific language items (collocations, lexico-grammatical associations). According to Kennedy (1998), such insights into language in use directly contribute to enhancing the content instruction of language teaching pedagogy and affect the selection of the items to teach in the language classroom. However, the chief and most significant objective that corpus linguistics aims to achieve is to provide or make readily available naturally occurring discourse created by users of the language for a communicative purpose in order to observe how language behaves through times.

Corpus linguistics has also contributed significantly to improving the lexicography of the English language. In actual fact, it has helped lexicographers to publish corpus-based dictionaries which include descriptions of words as they are most commonly employed in the real world and provide instances from authentic corpora to illustrate word use. A number of these dictionaries even provide basic information about the frequency of words in different types of texts and about the most common words or constructions with which the words often occur. In the number of such major, modern-day and corpus-based English dictionaries, there are *Collins Cobuild Advanced Dictionary* and the *Longman Dictionary of Contemporary English*. As opposed to other dictionaries, they elucidate to ESL/EFL learners both concrete and abstract concepts within their original contexts, followed by additional illustrative information to facilitate comprehension of all the various aspects surrounding their uses. Not only do corpus-based dictionaries explain concepts and provide their definitions, but they also describe the syntactic, semantic, and pragmatic aspects surrounding their uses through the construction of thorough, plain, and meaningful language

accompanied with their contexts of use. In *Collins Cobuild Advanced Dictionary of American English*, for instance, verbs are often explained in an interactive, user-friendly fashion using a plain and simple language construction containing 'if/when clauses' and a main clause in which their syntactic, semantic, or pragmatic values emerge. The authentic contexts in which they appear are then provided. Here is an example selected from *Collins Cobuild Advanced Dictionary of American English* to illustrate the use of the verb "engrave"

Engrave /in'greiv/ (engraves, engraving, engraved) V-Tif you engrave something with a design or words, or if you engrave a design or words on it, you cut the design or words into its surface. Your wedding ring can be engraved with a personal inscription at no extra cost (CONV) (p. 431).

The constant practice of utilizing monolingual, corpus-based dictionaries similar to the abovementioned has highly been recommended to ESL/EFL learners by lexicographers and language experts because these dictionaries have been found to be excellent, interactive, and user-friendly tools very conducive to fostering receptive skills, which constitute the foundation necessary to trigger authentic communication in and outside language classroom settings.

The other major contribution that corpus linguistics has made to the field of language teaching and language learning is the most frequently used word lists which it has helped researchers to generate for 'General English' as well as 'English for Specific Purposes'. Such lists assist ESL/EFL learners in familiarizing themselves with everyday lexicon and prepare them for international entrance competitions such as the TOEIC test, the TOEFL test and the GRE examination. The *Most Useful 2000 Words in English* (West, 1953), for instance, has been made available to help ESL/EFL learners to interact orally and communicate meaningfully in informal circumstances. These words are believed to be the most common 2000 words in English that they need for effective learning. Students are thought not to be able to achieve communication

gains without their exposure and use because they will meet the words very often in and out of classroom settings.

Apart from the generation of the most frequently used word-lists, corpus linguistics research and analysis have resulted in the design of various types of reference materials such as textbooks (e.g., Hong Kong English language textbooks/Finnish EFL textbooks), grammars and practice books to facilitate the learning process. In a review of recent research, Khojasteh and Shokrpour (2014) observed that textbook language models and dialogs often sound artificial and inappropriate for Communicative Language Teaching because of the idealistic situations that they describe and the simplistic language that they usually present for learning. In view of this inconsistency between real-world language and traditional textbooks' language, a growing body of research by linguists and language professionals has demonstrated how the use of corpus linguistics could contribute to the enhancement of L2 teaching by pointing out its significance and utility in language teaching. Almost unanimously, they advocated the use of corpus-based findings in order to inform material writers as to the quality of L2 teaching materials. According to Barbieri & Eckhardt (2007), corpus-based analysis represents the ideal tool that language teachers need in order to critically reappraise the order in which linguistic features are presented in textbooks and to determine which to prioritize in textbook presentations.

Conclusion

The use of corpus linguistics as a descriptive approach by foreign language teachers and teacher educators in English Language Teaching (ELT) at the advent of Communicative Language Teaching (CLT) in lieu of the dogmatic prescriptive methodology that Grammar Translation has firmly established in language teaching pedagogies would provide authentic context-embedded language to the grammar, vocabulary, and language functions of foreign language curricula and teaching styles, expose learners to the strategies and

processes typical of everyday conversations, eliminate what appears superfluous to teach and learn and prioritize only the frequently occurring linguistic items in order of importance, readjust or reposition grammar and vocabulary into their own registers of use, highlight how they function across the registers in authentic discourse, and display in context and through authentic discourse the functions that they play for the creation of meaning. In a word, the use of corpus linguistics in EFL pedagogy will unlock the secret of language in use across the various registers of formality.

The present article aims at highlighting corpus linguistics to foreign language educators in order to raise awareness over the issue of language authenticity in the context of foreign language teaching. If curriculum designers, material writers, and EFL educators eventually embrace a descriptive approach as a substitute for the existing standard prescriptive view and apply corpus linguistics methodology to the design of the classroom activities as advocated by Frazier (2003) and Romer (2010), authentic communication will eventually emerge in and beyond EFL classroom settings. The yawning gap which has often been found to exist between school and the real world in terms of authentic communication will narrow significantly and frequent exposure to real language followed by constant practice and production in the classrooms will enable learners to keep abreast of language in use. Contextualized and realistic pedagogical instructions firmly rooted in language use entail following a planned descriptive procedure patterned on First Language Acquisition. Carter and McCarthy (1996) believe that, to achieve such realistic pedagogical instructions in the second or foreign language classroom settings, the real challenge which language educators and curriculum designers face is to “provide descriptions and to develop materials which serve the needs of teachers in all situations, whether they be native or non-native, so that they can decide how best to make such hitherto unrecorded aspects of English more widely accessible” (p 370). The application of a descriptive approach to language

teaching in ELT pedagogy, however, only requires sound knowledge of general linguistics and of corpus linguistics in particular. It is of paramount necessity then that they be trained so that they might adjust their teaching styles to corpus-based practices.

References

- Barbieri, F. & Eckhardt, S. (2007). Applying Corpus-based Findings to Form-focused Instruction: The Case of Reported Speech. *Language Teaching Research*, 1(3), 319–346.
- Bernardinini, S. (2000). Systematizing serendipity: proposals for concordancing large corpora with language learners. In Burnard, L. & McEnery, T. (eds.). *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt am Main: Peter Lang.
- Biber, D. & Conrad, S. (2001). Quantitative Corpus-based research: Much more than bean counting. *TESOL Quarterly*, 35, 331-336.
- Biber, D. & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24, 199-208.
- Biber, D., Conrad, S. & Leech, G. (2002). *Longman Student Grammar of Spoken and Written English*. Essex, England: Longman.
- Breen, M. P. (2001). Syllabus design. In Carter, R. & Nunan, D. *The Cambridge Guide to Teaching English to Speakers of Other Languages* (Eds.). Cambridge, UK: Cambridge University Press.
- Brown, H. D. (2007b). *Teaching by Principles: An interactive approach to language pedagogy*. White Plains, NY: Pearson Education.
- Carter, R. A. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52(1), 43-56.
- Carter, R. & McCarthy, M. (1996). Correspondence. *ELT Journal*, 50, 369-371.

Carter, R. & Nunan, D. (Eds.). (2001). *The Cambridge Guide to Teaching English to Speakers of Other Languages*. Cambridge, UK: Cambridge University Press.

Carter, R. A. (1998). Orders of reality: CANCODE, communication, and culture'. *ELT Journal*, 52/1: 43-56.

Celce-Murcia, M. 2002. Why It Makes Sense to Teach Grammar in Context and Through Discourse. In Hinkel, E. & Fotos, S. *New Perspectives on Grammar Teaching in Second Language Classrooms*. New York and London: Routledge.

Chafe, W. (1994). *Discourse, consciousness, and time*: Chicago: University of Chicago Press.

Cheng, W. (2012). *Exploring Corpus Linguistics*. London and New York: Routledge.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press
Collins Cobuild Advanced Dictionary of American English (2007), Thomson Heinle.

Conrad, S. (2005). Corpus linguistics and L2 teaching. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cullen, R. & Kuo, I. (2007). Spoken Grammar and ELT course materials. *TESOL Quarterly*, Vol 41.

Davies, M. (2005). BYU-BNC (an online interface with the British National Corpus: World Edition). Provo, UT: Brigham Young University. Retrieved from <http://corpus.byu.edu/bnc/x.Asp>.

Davies, M. (2010). The corpus of contemporary American English. Provo, UT: Brigham Young University. Retrieved from <http://www.Americancorpus.org/>.

Doughty, C., & Williams, J. (Eds.) (1998). *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.

- Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon.
- Fotos, S. (2002). Structure-Based Interactive Tasks for the EFL Grammar learner. In Hinkel, E. & Fotos, S. *New Perspectives on Grammar Teaching in Second Language Classrooms*. New York and London: Routledge.
- Francis, W. N. and Kucera, H. (1964). *A standard corpus of present-day edited American English*. Providence: Brown University.
- Frazier, S. (2003). A Corpus Analysis of *Would*-clauses without Adjacent *If*-clauses. *TESOL Quarterly*, (3), 443–46.
- Greenbaum, S. & Quirk, R. (1990). *A student's grammar of the English language*. Harlow, Essex: Longman.
- Huang, L. (2011). Corpus-aided language learning. *ELT Journal*, 10.1093/elt/ccro31
- Hulstijn, J. H., & de Graaff, R. (1994). Under What Conditions Does Explicit Knowledge of a Second Language Facilitate the Acquisition of Implicit Knowledge? A research proposal. *AILA Review*, 11, 97–112.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Kadia, K. (1988). The effect of formal instruction on monitored and spontaneous interlanguage performance. *TESOL Quarterly*, 22, 509-15
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Harlow, UK: Addison Wesley Longman
- Khojasteh, L. & Shokrpour, N. (2014). Corpus Linguistics and English Language Teaching materials: A Review of Recent Research. *Khazar Journal of Humanities and Social Sciences*, Vol 17, N° 3.
- Lawson, A. (2001). Rethinking French Grammar for Pedagogy: The Contribution of French Corpora in Simpson, R.C. & Swales, J.M., (Eds.) *Corpus Linguistics in North America. Selections from the 1999 Symposium* (179–194).

- Longman Dictionary of Contemporary English*.(2001). Third Edition. Pearson Education: Harlow, Essex, England.
- Lyons, J. (1969). Intervention. In “Formal logic and natural language.” *Foundations of language*, 5, 269.
- McCarthy, M. & Carter, R. (2004). This, that and the other: Multi-word clusters in spoken English as visible patterns of interactions. *The Language Teacher*, Vol. 21.
- McEnery, T. and Gabrielatos C. (2006). English corpus linguistics. In Aarts, B. and McMahon, A. (Eds.). *The handbook of English linguistics*. Malden, MA: Blackwell Publishing Ltd.
- McEnery, T., Xiao, R. and Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Melrose, S. (1983). *Must and its periphrastic forms in American English usage*. Unpublished Master Thesis. University of California, Los Angeles.
- Meyer, C. F. (2004). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Peacock, M. (1997). The effect of authentic materials on the motivation of EFL learners. *ELT Journal*, 51(1).
- Pennington, M. C. (2002). Grammar and Communication: New Directions in Theory and Practice. In Hinkel, E. & Fotos, S. *New Perspectives on Grammar Teaching in second language classrooms*. Mahwah, NJ: Laurence Erlbaum.
- Quirk, R., Greenbaum, S. Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman
- Rajagopalan, K. (2004). The philosophy of applied linguistics. In Davis, A and Elder, C. (Eds.). *The handbook of applied linguistics*. Malden, MA: Blackwell Publishing Ltd.
- Richards, J. C. (2001). Postscript: The ideology of TESOL in Carter, R. & Nunan, D. *The Cambridge Guide to Teaching English to Speakers of Other Languages* (Eds). Cambridge, UK: Cambridge University Press.

Romer, U. (2010). Using General and Specialized Corpora in English Language Teaching: Past, Present and Future. In Campoy, M.C., Belles-Fortuno, B. & Gea-Valor, M. L. (Eds.), *Corpus-based Approaches to English Language Teaching*. London: Continuum.

Schumann, J. (1978). *The pidginization processes: A model for second language acquisition*. Rowley, MA: Newbury House.

Sinclair, J. McH. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Thornbury, S. (2004). *How to Teach Grammar*. Malaysia: Pearson Education Limited.

Stubbs, M. (2004). Language corpora. In Davis, A and Elder, C. (Eds.). *The handbook of applied linguistics*. Malden, MA: Blackwell Publishing Ltd.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins.

Tomlinson, B. (2001). Materials development. In Carter, R. & Nunan, D. *The Cambridge Guide to Teaching English to Speakers of Other Languages* (Eds.). Cambridge, UK: Cambridge University Press.

West, M. (1953). *A general service list of English words*. London: Longman, Green & Company.

Widdowson, H. G. (1972). The Teaching of English as Communication. *ELT Journal* 27/7: 15-19.

Widdowson, H. G. (1998). 'Context, community and authentic language'. *TESOL Quarterly* 32/4: 705-16.

Willis, D. (1990). *The lexical Syllabus: A New Approach to Language Teaching*. London: Collins.

TABLE DES MATIÈRES

Présentation	5
CONTRIBUTIONS	10
PREMIERE PARTIE : DYNAMIQUES SOCIOLINGUISTIQUES EN	
AFRIQUE.....	12
1. Deux alphabets pour une même et unique langue transfrontalière, le yorùbá: mythe ou réalité?, Moufoutaou ADJERAN, Université d'Abomey-Calavi (Bénin)	
	14
2. Mobilité linguistique au «marché cinquième» de Nouakchott: quelle articulation?, Mamadou LAM, Université de Nouakchott (Mauritanie) et Aminata GADIO, Université Montpellier III (France).....	
	23
3. Expressions figées en lokpa: critères de définition et portée sociolinguistique, Justine BASSABI SAMA CHRISTOPHE et Kègnidé Paulin YAI, Doctorants (LASODYLA-REYO), Université d'Abomey-Calavi (Bénin)	
	38
4. Barrière linguistique comme entorse à l'évolution des soins de santé administrés aux patients du centre hospitalier universitaire de Kara au Togo, Laré KANTCHOA et Mimboabe BAKPA, Université de Kara (Togo)	
	51
5. Aventure de la langue française parlée au Cameroun, Germaine Bienvenue NOUKIO, Centre Régional de la Recherche Scientifique (Cameroun)	
	63
6. Le festival international nuits d'Afrique à l'épreuve du plurilinguisme sporadique au Canada, Bangre Yamba PITROIPA, Université de Koudougou (Burkina Faso).....	
	76
7. Marché linguistique marocain: quels usages?, Amina SAOUSSANY, Université Ibn Zohr, Agadir (Maroc).....	
	93
8. Les universités du Burkina Faso: un univers linguistique francophone nouveau, Victor KABRE, Université de Koudougou (Burkina Faso).....	
	113
9. Les particularités lexicales dans les manuels de lecture conçus au Burkina Faso: représentations et perspectives didactiques, Issaka SAWADOGO, Université Ouaga I Pr Joseph Ki-Zerbo (Burkina Faso).....	
	129
DEUXIEME PARTIE : DYNAMIQUES DIDACTIQUES	
	143
10. Analyse ethnolinguistique et perspectives didactiques de pourquoi le bouc sent mauvais et autres contes du Bénin de Raouf Mama, Florentine AGBOTON, Université d'Abomey-Calavi (Bénin).....	
	145

11. Onomastic Analysis Of Ngugi Wa Thiong'o's Characters' Names: A Social Semiotic Perspective, Yémalo C. AMOUSSOU University of Abomey-Calavi, (Bénin).....	165
12. The Use Of Corpus Linguistics For Informing English Language Teaching In Efl Settings, Jean Cosme N. DOSSOU, Estelle BANKOLE-MINAFLINOU et Etienne K. IWIKOTAN, Université d'Abomey-Calavi (Bénin).....	198
13. Stage pratique des élèves maîtres en côte d'ivoire: quelles réalités socio-psycholinguistiques?, Koffi Ibrahim AMADOU, Université Félix Houphouët Boigny d'Abidjan-Cocody (Côte d'Ivoire).....	223
LISTE DES CONTRIBUTEURS.....	232