Corpus Linguistics and Language Pedagogy

Fiyinfolu Olubunmi Idowu
London Brunel Int'l College
Brunel University
Uxbridge. United Kingdom

Abstract

There is no doubt that corpus linguistics will revolutionize grammar teaching in fundamental ways in the future (Conrad 2000) and there is no doubt that corpora can be used for various ends in language pedagogy. The suggestion of using corpus analysis in language teaching has resulted from rapid advancement in computer technology and corpus linguistics in recent years, which has shown unprecedented potential for language learning and teaching.

Conrad (2000) argues that corpus linguistics offer us a new way of describing grammar, for example, on the basic level of frequency and register variation (varieties of language which are used for different situations). The use of linking adverbials such as however, therefore, and in other words that explicitly connect two units of discourse is one aspect of English grammar that demonstrates register variation.

'Corpus-based research has consistently shown that grammatical patterns differ systematically across register' (ibid: 549), for example, the use of the above linking adverbials, as stance markers, varies across registers (academic prose, conversation, newspaper/press and academic writing) and the implication of this corpus research for pedagogy is that grammatical study needs to take place within the context of a register or by comparing the patterns of use across registers (Hyland 2008).

In view of this reality, this paper will discuss the debate on corpus linguistics and highlight some of the insights derived from corpus linguistics which may have some potential effect on pedagogic grammar. Finally, it will discuss how these insights can be applied to teaching grammar in Nigerian secondary schools.

Keywords: *Corpus Linguistics, corpora, language pedagogy, register variation*

The Debate on Corpus Linguistics

A Corpus consists of a database of language stored in a computer which is thereby available for analysis by linguists and other researchers (Kennedy 1998 in Ranalli 2003; Kennedy 2014)**.** The compilation of the Collins Birmingham International Database (COBUILD) in the 1980s initiated and organised by John Sinclair, led to the first corpus-based dictionary and in 1990s COBUILD was expanded to form the Bank of English. The main focus of Corpus linguistics is to discover patterns of authentic language use through analysis of actual usage or *real world* text (Krieger 2003).

Krieger (2003:1) maintains that if findings from corpus linguistics are incorporated into grammar pedagogy, then register variation becomes a crucial aspect of grammar tasks and materials. Register consists of varieties of language that are used for various situations. Language is made up of many registers, which include fiction, academic writing, newspapers and conversation. Corpus-based research has shown that the use of linking adverbials varies across registers and the implication of this corpus research for pedagogy is that grammatical study needs to take place within the context of a register or by comparing the patterns of use across registers (Conrad 2000, Hyland 2008).

Carter (1998) compares the *real* data collected from the CANCODE spoken corpus and dialogue from textbooks and shows that textbook dialogues lacked core spoken features such as discourse markers, vague language, ellipsis and hedges. He also refers to the lack of three-part-exchanges in question and answer sequences in textbook dialogues. Carter (1998) points out that the two-part question and answer sequences often appearing in textbook dialogues are not the norm in real conversations. For example, replies to such questions are regularly followed up by some fixed phrases such as *really*, *I thought so*, *that's interesting*, *that's nice*, *I guessed as much,* in real conversation. He notes that this third part is particularly interactive and affective.

Traditional grammatical description would simply note that in indirect speech reporting, reference is been made to simple past tense forms such as *said* or *told* as in *They told me you had to get a cab*. However, corpus-based analysis opens up the possibility of a use of the past continuous tense of the reporting verb. For example, McCarthy and Carter (1995) in their CANCODE corpus discovered the use of past continuous tense of the reporting verb in an indirect speech. For instance, '*Tony "was saying" they should have the heating system on by about Wednesday'*. This feature, according to McCarthy & Carter (1995), offers interlocutors with a grammatical choice which has less to do with tense and also provides speakers opportunities to give more emphasis to the overall idea of the message than the certainty with which it was expressed.

They also observed that spoken English makes frequent use of the interrogatives, tails, adjacency pairs and the usage of tag question, which Carter interprets as dialogue facilitators, across speaking turns. A general clustering of tags, positive tags was discovered in the CANCODE Corpus. For example, *'yeah it does, doesn't it?'* They say the overall effect is to focus on how something is being said 'so that distinctly personal and interpersonal inflections are conveyed' (ibid: 212), and it enables speakers to develop the dialogue in a context where meanings are subject to negotiation and renegotiation.

They also discovered the use of *tend to* in the feature of spoken grammar. For instance, '...we *don't tend to go there very often'*. According to traditional grammar, "tend to" is a minor and insignificant form of modality but in their spoken corpus is one of the common features. They say its uses allow speakers to express point of view.

One common feature of the Nottingham corpus data is the numerous formulaic and fixed phrases used in spoken discourse. These findings in CANCODE support the view expressed in Lewis (1993 in Carter 1998:48) that since language is made up of lexical chunks then language teaching and learning should give more systematic attention to such features of the language. However, Carter (1998) argues that most of the fixed expressions are culture-bound and for non-native speakers, they may constitute an obstacle to acquisition of the language feature.

Corpus linguistics findings also show that idiomaticity makes language to be more *real*. Fox urges teachers to raise the student's awareness of 'how native speakers use English out there in the real world' (Fox 1998:43 in Prodromou 2003). However, authentic native speaker discourse might be quite inappropriate for speakers of English in other parts of the world, that is, what is authentic in one context might be inappropriate in the other (Kramsch & Sullivan 1996).

The native speaker's idiomaticity as found in the corpus may not be relevant for pedagogic grammar since most students in Nigerian secondary schools are learning English to become grammatically and linguistically competent and to be able to communicate effectively in English to other non-native speakers in the country and also to pass their examinations. For example, an expression such as *quid* for a pound as in *these trainers cost twenty quid* is common in UK English conversation and known to confuse second language speakers of English. In the Nigeria context, speakers say *pound* rather than *quid*. The high use of native speakers' idiomaticity as found in the corpus may appear to cause communication breakdown in Nigeria. However as a teacher, one can tell the students that when they (students) are interacting with native speakers, they need to be aware of their interlocutor's use of idiomaticity (Seidlhofer 2002).

McCarthy& Carter (1995) also discovered in CANCODE Corpus ellipted forms of language. Examples of such usage are: "*First class please*", "*Right, send that first class, please*." In my view, I would prefer to teach the full forms or structure because you cannot omit features until you know and have first practised the full forms from which the reductions can be made, that is to say, once the forms are learnt then students can now pick up or play with the language by producing the reduced form depending on the contexts of usage. For example, in Nigeria context, most of the English language teachers learnt English as their second language and a lot of them never came into contact with *real* English spoken grammar which is the inner circle English.

So how will they teach these forms? If English in the inner circle is *real*, then what about English as spoken in the outer circle such as Nigeria? For most learners of English, interactions with native speakers in the inner circle will be rare, so an imposition to expect learners to acquire naturalistic, real, native speaker English when they simply do not need it is unrealistic in our context (Prodromou 1990; Rampton 1990).

Some Corpus linguists (e.g. Sinclair 1991, Francis and Sinclair 1994) have continued to argue for a description of spoken grammar and vocabulary and pedagogic materials that is based only on insights from corpus findings. However, Biber, Conrad & Reppen (1998) acknowledge the relevance of corpus linguistics to pedagogy but argue that corpus findings should be seen as a complementary approach to more traditional approaches rather than a replacement. Cook also argues and supports Carter's view that pedagogic materials should be influenced by corpus findings but rejects the form of corpus driven approach. Cook (1998:58) argues that:

> '…where pedagogy is concerned, corpus statistics say nothing immeasurable but crucial factors such as students' and teachers' attitudes and expectations…. consequently, computer corpora – while impressive and interesting records of certain aspects of language use – can never be more than a contribution to our understanding of effective language teaching'.

Owen (1993) also supports the view that corpus findings should be seen as a complement and not as a form of replacement to traditional approach. He concluded by saying:

> 'There is no doubt that computer-assisted corpus- linguistics does reach some parts of language other grammars fail to reach. The basic insight that grammar and lexis are closely integrated is important linguistically and pedagogically, and the grammar provides evidence to support it. But overall, the description is disappointing…' (ibid: 184).

He rejects corpus-driven language teaching and argues that 'the grammarian and the language teacher need the corpus as servant, not as master' (ibid: 185).

Application of Corpus linguistics to English teaching in Nigeria

The description of language forms in pedagogic grammar has generally been based on the grammarian's intuition (traditional approach), but recent research in corpus linguistics has suggested that our intuitions about language are too unreliable to be used as a basis for prescription (Sinclair 1991). From my point of view and considering my teaching context, insights derived from corpus linguistics can be applied to pedagogic grammar but must serve as a complement, not a replacement. This view is consistent with Biber, Conrad & Reppen (1998), Cook (1998), Carter (1998) and Owen (1996).

A teacher can compile and analyse a local learner corpora created from the collections of students' output in the form of an essay in the classroom into a database by using corpus linguistic software like WordSmith Tools. By compiling and analysing the learner's corpora, a teacher may assess the students' correctness and find the most typical in the use of grammar and vocabulary that can be addressed during the lesson, which can be in the form of drills, dialogues or explanations. Mukherjee (nd) argues that '...by compiling a local learner corpus, teachers are provided with a powerful resource for systematic error analysis.'(p.19)

Corpora can improve students' writing skills by introducing different stylistics figures or discourse markers. For example, most Nigerian Secondary School Students fail the West African Examination because of the difficulties they experience with the writing skills. A Corpus-based pedagogy that discusses collocation and cohesion with the use of corpus concordance will help individual student to draw from a reservoir of words that are often used together to express themselves. In writing an essay or composition, cohesion should play a vital role because writers and readers need to be aware of the connection that holds chunks of texts together and their contribution to the text as a unit of meaning (Mahlberg 2006). Furthermore, Cook (1989:127 in Mahlberg 2006) argues that cohesion does not receive adequate attention in traditional approach to language teaching; he notices that 'cohesion between sentences is too easily seen as an aspect of language use to be developed after the ability to handle grammar and words within sentences.' Textbooks in traditional language often discuss lexical and grammatical cohesion since grammar and vocabulary are seen as a separate entity.

Also, discourse markers and collocations as found in corpus findings will be relevant for pedagogic grammar. For example, Hyland (2008) identifies the presence of collocation introduced as *lexical bundles* or *clusters*. For example, *as a result of* and *it should be noted that*. He argues that the use of corpus concordance will enable the student to identify a text as belonging to an academic register.

Hyland (2008) confirms that the British National Corpus Baby edition tells us that *as a result of* is a frequent collocation in academic writing but *as an outcome of* is never used. The absence of these lexical bundles in academic writing might show lack of fluency of an inexperienced writer, but its presence helps us to shape meanings and contributes to our sense of coherence in a text (ibid).

A concordance also include information on other frequent routines in language use, for example, semantic prosodies, that is, the tendency of a word to occur in positive or negative contexts (Stubbs 1995). Corpus findings enable us to teach the context of use; for example, *provide* is used in positive contexts and *affect* in negative contexts. Also, the structure *set in* is used in a negative context. In other words, it will be contextually correct to say that *the Plague has set in,* but it will be inappropriate to say that *the summer season has set in*. This is because *set in* according to corpus concordance occurs in negative contexts alone and the *summer season* is a pleasant period that should occur in a positive context.

In conclusion, corpus-based pedagogy grammar can be based on discover patterns of authentic language use through analysis of actual usage because ESL teachers can not cover everything in a grammar class (Conrad 2000). However, it should not be a replacement but a complement to teaching in an ESL grammar class.

References

Biber, D, S, Conrad &Reppen (1998) *Corpus Linguistics Investigating Language Structure and Use.* Cambridge: CUP

Carter, R (1998) 'Orders of reality: CANCODE, communication, and culture,' *ELT Journal 52/1: 43-26*

Carter, R (1998) 'Reply to Guy Cook', *ELT Journal 52/1: 64*

Conrad, S (2000) 'Will Corpus Linguistics Revolutionalise Grammar Teaching in the 21st century?' *TESOL QUARTERLY 34/3: 548- 560*

Cook, G (1989) *Discourse*. Oxford: Oxford University Press

Cook, G (1998) "The uses of reality: a reply to Ronald Carter" *ELT Journal* 52/1: 57-63

Fox, G (1998) 'Using corpus data in the classroom', In Tomlinson, B (ed.) *Materials Development in Language Teaching*. Cambridge: CUP, Pp.25-43

Francis, G and Sinclair, J (1994) "I Bet He Drinks Carling Black Label": A Riposte to Owen on Corpus Grammar" *Applied Linguistics* 15/2: 190-200

Hyland, K (2008) 'Academic clusters: text patterning in published and post graduate', *International Journal of applied linguistics. Vol18. No1 41-62*

Kennedy, G (1998) *An Introduction to Corpus Linguistics*. Harlow: Longman

Kennedy, G (2014) *An Introduction to Corpus Linguistics*. London: Routledge.

Kramsch, C & P, Sullivan (1996) 'Appropriate Pedagogy', *ELT Journal 50/3: 199-212*

Krieger, D (2003) 'Corpus Linguistics: What is it and how it can be applied to teaching,' *The internet TESEL Journal* 1x/3 Available at http://iteslj.org/Articles/Krieger-Corpus.html

Lewis, M (1993) *The Lexical Approach*. Hove: Language Teaching Publication

Mahlberg, M (2006) 'Lexical Cohesion: Corpus Linguistic theory and its application in English language teaching', *International Journal of Corpus Linguistics* 11/3: 363-383

McCarthy, M & R, Carter (1995) 'spoken grammar: what is it and how can we teach it?' *ELT Journal* 49/3: 207-218

Mukherjee, J (no date) 'Corpus Linguistics and Language Pedagogy: The state of the art and beyond'. Retrieved on 27/10/11 Available at http://www.uni-giessen.de/anglistik/ling/Staff/mukherjee/pdfs/Corpus%20Linguistics%20and%20Language%20Pedagogy.pdf

Owen, C (1993) 'Corpus- Based Grammar and the Heineken Effect: Lexico- Grammatical Description for Language Learners,' *Applied Linguistics, 14/2: 167-187*

Owen, C (1996) 'Do Concordances require to be consulted?' *ELT Journal 50/3:219-224*

Prodromou, L. 1990. 'English as cultural action' in R. Rossner and R. Bolitho (eds.). *Currents of Change in English Language Teaching.* Oxford: Oxford University Press.

Prodromou, L (2003) 'Idiomaticity and non- native speaker', *English Today 74* 19/2: 42-48

Rampton, M. B. 1990. 'Displacing the native speaker: expertise, affiliation, and inheritance'. *ELT Journal* 44/2:97-106. Republished in T. Hedge and N. Whitney. 1996. *Power, Pedagogy, and Practice.* Oxford: Oxford University Press.

Ranalli, J.M (2003) 'ELT Course books in the age of corpus linguistics: constraints and possibilities' Retrieved on 12/11/11 Available at http://www.google.co.uk/#sclient=psy-ab&hl=en&source=hp&q=corpus+linguistics+by+Ranalli+james+2003&pbx=1&oq=corpus+linguistics+by+Ranalli+james+2003&aq=f&aqi=q-n1&aql=&gs_sm=e&gs_upl=15248l41816l0l42248l49l38l4l6l6l5l1520l2031811.2.3.7.5.12.3.3.1l47l0&bav=on.2,or.r_gc.r_pw.,cf.osb&fp=8de86bc119b3361f&biw=986&bih=593

Seidlhofer, B. (2002): 'Pedagogy and local learner corpora: working with learning-driven data', *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, Ed. Granger, S, Hung, J & Petch-Tyson, S. Amsterdam: John Benjamins. 213-234

Sinclair, J.M (1991) *Corpus, Concordance, Collocation*. Oxford: OUP

Stubbs, M (1995) 'Collocations and Semantic Profile: on the cause of the trouble with quantitative studies', *Functions of Language 2/1:23-55*