

Corpus Linguistics: Applications, Implications and Limitations

BENRABAH, Boumediène

Université Abu-Bakr BELKAID, Tlemcen

Résumé

The present article discusses the way a corpus can help materials developers, syllabus designers and even teachers in presenting a representative sample of a language or some parts of it. The issue is of recent increasing interest to corpus linguistics. Therefore, easily accessible and accurate data provided by corpora is largely indicated along with arguments for their usefulness in descriptive, theoretical and applied linguistics. However, a corpus is always designed for a particular purpose and when applying linguistics to a particular task, we need to preserve the norms specific to that given context or register. So, language variation should constitute the turbine engine in selecting the type of corpus data needed as repositories of actual instances of language use. Yet, one problem in employing a corpus in translating is to know how to describe the relevant frequencies in corpus. So, if the course designer collects pieces of authentic language that contain instances of the most frequent words, then that collection (corpus) will exemplify what the learner needs to know. Having argued for the benefits of corpora to the study of language, it is as well to consider also the limitations of a corpus.

1. Introduction

The use and uses of corpora in applied linguistics have expanded rapidly over the past 20 years, in part because of the advent of improved, more accessible systems of electronic storage and analysis, and also because of an ever-growing appreciation of the huge potential of corpus work. The relative merits of corpora designed as collections of texts and as collections of language samples, in turn reflecting how and for what purposes corpora are investigated; hence, defining some of the basic issues surrounding the construction and evaluation

of corpora. These issues concern the size, content, representativeness and stability in corpora over time. Most applications of corpus-based research is profitable in a variety of domains- most centrally in the design of syllabi and materials for language teaching: in English for specific purposes(ESP), language testing, but also in dictionary work, the study of ideology, culture, stylistics, forensic linguistics, etc.

Corpora in Applied Linguistics should be of great value to language teachers, syllabus designers, materials writers and language testers, as well as to applied linguists. However, it is no exaggeration to say that corpora and the study of corpora have revolutionised the study of language and of the applications of language, over the last few decades. The improved accessibility of computers has changed language study from a subject for specialists only to something that is open to all. The main issues of application are namely the effect of corpus studies upon theories of language and how languages should be described. Corpora allow researchers not only to count categories in traditional approaches to language but also to observe categories and phenomena that have not been noticed before. Therefore, corpus findings can be seductive, and it is important to be aware of the possible pitfalls in their production. Our specific interest as EFL teachers is how language, more specifically English, works, and how knowledge about language can be applied in certain life-contexts.

2. Language Applications

The field of Applied Linguistics itself has undergone something of an evolution over the last few decades. Once, it was almost synonymous with language teaching but now it covers any application of language to the solution of real-life problems.

As has often been said (eg. Widdowson 1979; 2000), the difference between linguistics and applied linguistics is not simply that one deals with theory and the other with applications of those theories. Rather, applied linguistics has tended to develop language theories of its own, ones that are more relevant to the questions applied linguistics seeks to answer than are those developed by theoretical linguistics. Increasingly, corpora are adding to the development of those applied views of language.

3. The Design of Content

Though there is an enormous literature on syllabus design, it remains a difficult task to be fulfilled. Hence, before designing the model that best meets the requirements of given situation, one needs to start from a theoretical basis finding out what is exactly meant by a syllabus.

3.1 Definition of Content

The syllabus is composed of the subjects, the language structures and tasks that should be studied in a particular course. This is clearly stated by Hutchinson and Waters (1987) who define the language content as: *.....a document which says what will (or at least what should) be learnt.*'' (Hutchinson and Waters 1987:80). Moreover, Yalden (1987) posits: *The content is now seen as an instrument by which the teacher, with the help of the syllabus designer can achieve a certain coincidence between the needs and aims of the learners and the activities that will take place in the classroom.* (Yalden 1987: 86). For Yalden, the syllabus is an organization of the learners' needs and aims into a number of activities which will constitute the content of courses. Furthermore, Corder (1937) notes that: *A finished syllabus is the overall plan of the learning process. It, too, must specify what components or learning items must be*

available or learned by a certain time and what is the most efficient sequence in which they are learned.(Corder1973:296)

Hence, the syllabus is a plan of work which acts as a guideline for the teacher. It determines the content of the course, i.e. language structures, subjects and tasks, that is to be taught and the way in which such a content has to be organized. In sum, one of the main purposes of the syllabus is to make explicit the linguistic content of teaching in a particular situation and to break down the mass of knowledge to be learned into manageable units.

3.2 Variables in Designing Content

Language is a complex entity and cannot be taught in one go. Through syllabus design, course content is provided and arranged for an effective teaching/learning process. However, before the design process, a set of criteria have to be taken into consideration. In this sense, Allen (1984) reveals: *Since language is highly complex and cannot be taught all at the same time. Successful teaching requires that there should be a selection of material depending on the prior definition of objectives, proficiency level and duration of the course. This selection takes place at the syllabus planning stage.* (Allen 1984:65)

Thus, course objectives, learners' actual knowledge and time allotment are to be highly considered when designing a syllabus. Other methodological considerations are taken into account notably learners' learning styles and the pedagogical approach (Brumfit,1984). Besides, Dubin and Olsten (1986) mention a number of environmental factors, said to influence syllabus design. On the one hand, the syllabus design is greatly influenced by a set of environmental constraints, namely:

- The language setting, which concerns the status of the language in the community in which it is used (it can be the mother tongue, a second language or a foreign language).
- The areas of language use in society.
- The role of the language in the political life taking into account the country's economy and technology (Asworth, 1985).
- Learners' attitudes towards the language. This depends on their degree of awareness of the language.

On the other hand, the analysis of both the target situation and the learning situation provide answers and supply information about the teaching/learning situation. Miliani (1994), on his part, has summarized four essential points that have to be taken into account before designing a syllabus. These are: -Situation analysis.

- Setting aims and objectives.
- Generating Syllabus content.
- Assessment

Situation analysis deals with the educational institution and the means it provides in terms of finance, teaching materials and staff. It also provides a learners' profile including their motivation and attitudes towards the language course as well as the teachers' profile covering their attitudes towards the course and their professional 'know-how'.

Aims and objectives are set as a result of learners' needs analysis. The aims are defined at a broad level as general statements, while objectives refer to what the learners have to achieve at the end of their learning experience. Whereas, in generating syllabus content, four criteria are considered:

- Validity: the language of syllabus has to be authentic and acceptable.
- Significance: the content should arouse and sustain learners' interest.

- Learner ability: syllabus content should not go beyond learners' level of competence. Furthermore, the content of the syllabus has to be flexible enough to allow for eventual revisions as learners' needs are subject to change.
- Finally, assessment allows the evaluation of learners' degree of acquisition, course objectives, teaching techniques, and the syllabus itself.

Miliani (1984) notes that: *the relation between the studied factors should be cyclical*. A failure in the designed programme would, therefore, probably mean that a deficiency has occurred when dealing with one of the four points previously mentioned.

3.3 The Design Process

Such procedure needs deep consideration of a plethora of variables, all of which are to be included in the design process. Presumably, setting the objectives, selecting and grading content, the choice of the content, and some teacherly-related items should constitute the task of designing a process.

3.3.1 Setting Objectives

The general goals for which a given language programme is being taught or learned are generally embodied within the curriculum. They provide a rationale for the syllabus or course designer in the sense that they dictate the type of communicative and pedagogic objectives which are both appropriate and feasible for the educational system in question. In this sense, Dubin and Olsten (1986) maintain that: *A curriculum contains a broad description of general goals...A syllabus is a more detailed and operational statement of teaching and learning elements which translates the philosophy of the curriculum into a series of planned steps leading to more narrowly defined objectives.* (Dubin and Olshtain1986:34)

Therefore, the production of the syllabus involves the syllabus designer to determine the objectives from the curriculum generalised goals since, *goal statements are relatively imprecise. While they can act as general signpost, they need to be fleshed out in order to provide information for course and programme planners .This can be achieved through the specification of objectives.* (Nunan, 1988 : 61)

The design of a programme to a learning requirement begins by a careful determination of the learning objectives followed by the working of a relevant syllabus and the elaboration of a course based on it, in relation to learners' attainment and needs. In other words, setting objectives is an integral part of syllabus production. They can be useful, not only to guide the selection of content (structures, function, notion, tasks ...) but also to give a sharper focus for teachers to give learners a clear idea of what they can expect from a language programme and to help in developing means of assessment and evaluation.

As this level, it is worth mentioning that, in the real world, if one looks at the language programmes in operation, it is not impossible to find many situations where an operational guide on the syllabus level does not contain a statement of more general goals, (i.e. we may have a syllabus without a curriculum). However, absence of concrete statements about policy does not indicate that goals are totally missing. More likely, it may mean that the general goals are presented by the beliefs and attitudes of teachers and administrators in the programme and learners' needs; even though there is no written curriculum. This is generally the case in ESP teaching / learning situations.

3.3.2 Selecting and Grading Content

Setting objectives paves the way for the selection and production of the necessary content. Such content covers the

language forms and functions needed by the learners to cater for the requirements of the target situation, and will be presented within suitable activities. This means that there should be congruence between syllabus objectives, learning content and activities, in order to reach course effectiveness. In this sense, Nunan (1988) argues: *There is a desirability of relating classroom activities to syllabus goals and objectives so that courses and programmes derived from such syllabus have an overall coherence of purpose. Failure to provide links between goals, content and activities can lead to a situation in which the desired outcomes of a programme are contradicted at classroom level.* (Ibid: 67)

As well as listing what should be learned, a syllabus can also state the order in which it is to be learned (Hutchinson and Waters, 1987). This means that the content selected to meet course objectives has to be organized in a gradual manner in order to be more efficient. In the same vein, Nunan (1988) adds: *Grading ...is one of the central steps in syllabus construction...any proposal failing to offer criteria for grading and sequencing can hardly claim to be a syllabus at all.*

Moreover, activities which can be exploited at different levels of difficulty are more useful than those which are only suitable for a single proficiency level.

3.3.3. The Choice of the Content

A syllabus is normally organized according to the principles of a given pedagogical approach. However, the existence of various approaches results in the existence of different types of syllabus, three of which are most frequently used. They are:

- The structural, grammatical or linguistic syllabus: centred on the teaching of structures and grammatical items.
- The functional/notional syllabus: based on the teaching of functions and notions.

- The situational syllabus: based on the teaching of tasks and functions through specific situations.

These are the most widely used types of syllabus. Nevertheless, other kinds of syllabus, based on other different approaches, exist as well.

Course designers who carefully consider the various approaches to syllabus design emphasise the fact that a set of syllabuses need to be combined and integrated in an eclectic manner in order to bring about positive teaching and learning results. In this respect, Swan (1985) says: *The real issue is not which syllabus to put first: it is how to integrate eight or so syllabuses...into a sensible teaching programme*'''. (Swan, 1985 cited in Hutchinson and Waters 1987:89).

Therefore, any teaching material must, in reality, operate several syllabuses at the same time. One of them will, probably, be used as the principle organizing feature, while the others are still there even if they are not so apparent. Otherwise, a syllabus framed only on one aspect; be it structures, functions, situations, tasks or whatever, will in all probability miss the opportunity to develop the acknowledged elements efficiently.

Designing a syllabus is not an easy task to fulfil. Besides, the role of the syllabus is a complex one, but it clearly satisfies a lot of needs. Its importance stems from its multi-functional purpose. Hence, one needs to be crucially aware of the different functions the syllabus fulfils so that it can be designed and used most appropriately. The syllabus is also an important document in the teaching learning process in the sense that it provides a set of criteria for materials production, teaching and testing. In like manner, it should be flexible (not rigid), open-ended (not closed); and not static but subject to constant revision as a result of feedback from the classroom and in order to maximize the aims and process of learning.

4. What is a Corpus?

A corpus is defined in terms of both its form and its purpose. Linguists have always used the word ‘corpus’ to describe a collection of naturally occurring examples of language consisting of anything from a few sentences to a set of written texts or tape-recordings which have been collected from linguistic study. More recently, the word has been reserved for collections of texts (or parts of a text) that are stored and accessed electronically. Because computers can hold and process large amounts of information, electronic corpora are usually larger than the small, paper-based collection previously used to study aspects of language. A corpus is planned, though chance may play a part in the text collection, and it is designed for some linguistic purpose. The specific purpose of the design determines the selection of the texts, and the aim is other than to preserve the texts themselves because they have intrinsic value. This differentiates a corpus from a library or an electronic archive. The corpus is stored in such a way that it can be studied non-linearly, and both quantitatively and qualitatively. The purpose is not simply to access the texts in order to read them, which again distinguishes the corpus from the library and the archive.

4.1 What a Corpus Can Do

Strictly speaking, a corpus by itself can do nothing at all, being nothing other than a store of used language. Corpus access software, however, can re-arrange that store so that observations of various kinds can be made. If a corpus represents, very roughly and partially, a speaker’s experience of language, the access software re-orders that experience so that it can be examined in ways that are usually impossible. A corpus does not contain new information about language, but the software offers us a new perspective on the familiar. Most readily available software packages process data from a corpus

in three ways: showing frequency, phraseology, and collocation.

4.1.1 Frequency

The words in a corpus can be arranged in order of their frequency in that corpus. This is more important when corpora are compared in terms of their frequency. For instance, in a survey conducted within the Bank of English Corpus compared with comparable corpus of politics and with a corpus of materials science, grammar words are more frequent than lexical words. Another finding is that in materials science the vocabulary is less wide than in politics.

Frequency lists from corpora can be useful for identifying possible differences between the corpora that can be studied in more detail. For that, another approach using frequency where words are compared across corpora. A best example of such, a comparison is made between a corpus of books, a corpus of the Times newspaper and a corpus of spoken British English about the use of the words “*must*” and “*have to*”. The corpus findings provide the following: the books corpus and the Times corpus use “*must*” in preference to “*have to*”, however, the spoken corpus shows the reverse trend; suggesting that “*have to*” is less formal than “*must*”. Other works on comparative frequencies deal with registers (conversation, fiction, news and academic) where proportions of the use of tenses are found.

4.1.1. Phraseology

Most people access a corpus through a concordancing programme. Concordance lines bring together many instances of use of a word or phrase, allowing the user to observe regularities in use that tend to remain unobserved when the same words or phrases are met in their normal contexts. It is through concordances, then, that phraseology is observed. One point of interest is the way that phraseology can be used as an

alternative view of phenomena that teachers of English are frequently called upon to explain. For example, learners often confuse adjectives such as '*interested*' and '*interesting*' and find that explanations of the different meanings do not make the choice more accessible in spontaneous speech. (Such a research has been undertaken through the Bank of English which has increased in size since its inception in the 1980s to 400 million word corpus). What the findings show is that, overwhelmingly, '*interested*' is used in the phrase '*interested in*' and the pattern '*someone is interested in*' is exceptionally frequent. By contrast, '*interesting*' is nearly always used before a noun. A similar approach is taken by Kennedy (1991) in his study of '*between*' and '*through*'. After pointing out that reference books have difficulty in expressing the differences between these words, Kennedy adopts a phraseological approach. He notes that *between* is frequently found after nouns such as: difference, distinction, gap, contrast, conflict and quarrel, as well as relationship, agreement, comparison, meeting, contact and correlation; whereas, *through* is more frequently found after verbs such as: go, pass, go, run, fall and lead. These and other observations enable him to provide a profile of each word that relates each aspect of meaning to typical phraseologies. Logically it helps to assign frequencies to the different meanings, or semantic functions. As a matter of fact, a quarter of the instances of *between* in the LOB(Lancaster-Oslo-Bergen) corpus have a location meaning, whilst about the same proportions of the instances of *through* have an instrumental meaning.

4.1.2. Collocation

The final example of how the data in corpora can be manipulated is the calculation of collocation. Collocation is the statistical tendency of words to co-occur. A list of collocates of a given word can yield similar information to that provided by

concordance lines, with the difference that more information can be processed more accurately by the statistical operations of the computer than can be dealt by the human observer. For example, collocates of the word ‘*shed*’ include: light, tear/s, gardens, job, blood, cents, image, pounds, staff, skin and clothes. Only when it collocates with ‘*garden*’ is the word *shed* a noun; in all other cases it is a verb. Its meaning is something like ‘lose’ or ‘give’ but the precise meaning of each phrase depends on the collocate(s):

-shed light (on) means ‘*illuminate*’, usually metaphorically

-shed tears means ‘*cry*’ (literally) or ‘*be sorrowful*’ (crying metaphorical tears)

-shed blood means ‘*suffer*’ or ‘*die*’, either literally or metaphorically

-shed jobs or shed staff mean ‘*get rid of people*’

-shed pounds means ‘lose weight’

-in shed skin and shed clothes, shed means ‘*remove*’

A collocation can indicate pairs of lexical items, such as shed+tears, or the association between a lexical word and its frequent grammatical environment (colligation)

4.2. What corpora are used for?

Corpora nowadays have a diverse range of uses. For language teaching, corpora can give information about how a language works that may not be accessible to native speaker intuition, such as the detailed phraseology mentioned above. In addition, the relative frequency of different features can be calculated. According to Mindt (2000), for example, nearly all the future time reference in conversational English is indicated by **will** or other modals. The phrase **be going to** accounts for about 10% of future time reference, and the present progressive less than 5%. Information such as this is important for syllabus and materials design. Increasingly, language classroom teachers are encouraging students to explore corpora for

themselves (see, for example, Burnard and McEnery eds.2000), allowing them to observe nuances of usage and to make comparisons between languages.

Translators use comparable corpora to compare the use of apparent translation equivalents in two languages, and parallel corpora to see how words and phrases have been translated in the past. As an example, the English word **still** can translate or be translated by the French **toujours** or **encore**, or by expressions with **couramment** or the verb **continuer**. Sometimes when an English sentence includes the word **still** the parallel French sentence has no translation equivalent at all, but when **toujours** and **encore** are present in the French sentence, the English parallel sentence always contains **still**.

General corpora can be used to establish norms of frequency and usage against which individual texts can be measured. This has application for work in stylistics and in clinical and forensic linguistics.

Corpora are used also to investigate cultural attitudes expressed through language and as a resource for critical discourse studies.

4.3. Types of corpora

A corpus is always designed for a particular purpose, and the type of corpus will depend on its purpose. Here are some commonly used corpus types:

a- Specialised Corpus: A corpus of texts of a particular type, such as newspaper editorials, geography textbooks, academic articles in a particular subject, lectures, casual conversation, essays written by students etc. It aims to be representative of a given type of text. It is used to investigate a particular type of language. Researchers often collect their own specialised corpora to reflect the kind of language they want to investigate.

There is no limit to the degree of specialisation involved, but the parameters are to limit the kind of texts included. For instance, a corpus might be restricted to a time frame, consisting of texts from a particular century, or to a social setting, such as conversation taking place in a bookshop, or to a given topic, such as newspaper articles dealing with violence in stadiums.

b- General Corpus: A corpus of texts of many types. It may include written or spoken language, or both, and may include texts produced in one or many countries. It is unlikely to be representative of any particular ‘whole’, but will include as wide a spread of texts as possible. A general corpus is usually much larger than a specialised corpus. It may be used to produce reference materials for language learning or translation, and it is often used as a base-line in comparison with more specialised corpora. Because of this second function, it is also sometimes called a **reference corpus**. Well-known general corpora include the British National Corpus (100 million words) and the Bank of English (400 million words).

c- Comparable Corpora: Two (or more) corpora in different languages (e.g. English and French) or in different varieties of a language (e.g. Indian English and Canadian English). They are designed along the same lines; for example, they will contain the same proportions of newspaper texts, novels, casual conversation, and so on. Comparable corpora of varieties of the same language can be used to compare those varieties. Comparable corpora of different languages can be used by translators and by learners to identify differences and equivalences in each language.

d- Parallel Corpora: Two (or more) corpora of different languages, each containing texts that have been translated from one language into another (e.g. a novel in English that has been translated into French, and one in French that has been

translated into English) or texts that have been produced simultaneously in two or more languages (e.g. European Union regulations which are published in all the official languages_of the EU). They can be used by translators and by learners to find potential equivalent expressions in each language and to investigate differences between languages.

e- Learner Corpus: A collection of texts- essays, for example- produced by learners of a language. The purpose of this corpus is to identify in what respects learners differ from each other and from the language of native speakers, for which a comparable corpus of native-speaker texts is required. There are a number of learner corpora around the world, of which the best known is the International Corpus of Learner English (ICLE). This is in fact a collection of corpora of 20,000 words each, each one comprising essays written by learners of English from a particular language background (French, Swedish, German, etc.)

f- Pedagogic Corpus: A corpus consisting of all the language a learner has been exposed to. For most learners, their pedagogic corpus does not exist in physical form. If a teacher or researcher does decide to collect a pedagogic corpus, it can consist of all the coursebooks a learner has used, plus any tapes they have heard. The term ‘pedagogic corpus’ is used by D.Willis (1993). A pedagogic corpus can be used to collect together for the learner all instances of a word or phrase they have come across in different contexts, for the purpose of raising awareness. It can also be compared with a corpus of naturally occurring English to check that the learner is being presented with language that is natural-sounding and useful.

g- Historical or Diachronic Corpus: A corpus of texts from different periods of time. It is used to trace the development of aspects of a language over time. Perhaps the best known

historical corpus is the Helsinki Corpus which consists of texts from 700 to 1700 and comprises 1.5 million words.

h- Monitor Corpus: A corpus designed to track current changes in a language. A monitor corpus is added to annually, monthly or even daily, so it rapidly increases in size. However, the proportion of text types in the corpus remains constant, so that each year (or month or day) is directly comparable with every other.

5. Limitations of a Corpus

Having argued for the benefits of corpora to the study of language, it is as well to consider also the limitations of a corpus.

1- A corpus will not give information about whether something is possible or not, only whether it is frequent or not. Descriptions of English are moving towards a concentration on the typical and away from notions of well-formedness (Sinclair1991:17; Biber et al 1998:3), but questions of the type- 'Can I say this?' -still need to be answered. To give a specific example, is the verb **expire** used with the preposition of, by analogy with **die** of (e.g. died of heart disease)? The Bank of English shows us that **expire** of is rare: 5 lines out of 3,519 lines for **expire** (0.1%), compared with die of, which has 5,259 lines out of 85,511 lines for **die** (6%). It is also rare in comparison with **expire** from (12 lines). Furthermore, of the 5 lines of **expire** of, three are jokey, non-serious. All this is useful information, but it does not actually answer the question as to whether **expire** of is acceptable English. Native-speaker intuition has to answer that question.

2- A corpus can show nothing more than its own contents. Although it may (justifiably) claim to be representative, all attempts to draw generalisation from a corpus are in fact extrapolation. A statement about evidence in a corpus is a

statement about that corpus, not about the language or register of which the corpus is a sample. Thus conclusions about language drawn from a corpus have to be treated as deduction, not as facts.

3- A corpus can offer evidence but cannot give information. For example, what does the phrase *something of a* mean before a noun, as in *something of a surprise*? It might be supposed to be a ‘downtowner’: *something of a surprise is a small surprise*. Comparing **come** as something of a surprise with **come** as a surprise in the Bank of English affirms this supposition to some extent, in that adjectives such as total, complete and big occur before surprise but not in the phrase *something of a surprise*. Apart from that, however, there are no real clues as to the meaning of *something of a surprise* in the concordance lines. The corpus simply offers the researcher plenty of examples; only intuition can interpret them.

4- Perhaps most seriously a corpus presents language out of its context. The work of Kress and van Leeuwen (Kress and van Leeuwen 1994; Kress 1994), for example, depends upon a text being encountered in its visual and social context (or, more properly, the text consists not of the words alone but of the spatial context in which the words appear). A corpus (as corpora are currently conceived) cannot show this. Equally significant is the issue of spoken data, in that transcription can never represent intonation, kinesics (‘body language’), and other paralinguistic information entirely accurately. Even if the issue of visual and intonational features is ignored, it remains true that a corpus masks some of the features of the texts in it by presenting concordance lines, in which the structure of the original is lost. These factors all show the need for a corpus to be one tool among many in the study of language.

6. Challenges to the Use of Corpora

Although corpora are widely acknowledged to be a valuable resource in describing language, there are fewer consensuses on the value of corpus findings in the description of language for learners or on the use of corpus-based material in language classrooms. Among others, Widdowson (2000) and Cook (1998) have spoken against what they term an ‘extreme’ attitude towards using corpora in language teaching. At the risk of over-simplification, their arguments can be summarised thus:

1. A corpus is ‘real language’ only in a very limited sense. Language in a corpus is de-contextualised and must be re-contextualised in a pedagogic setting to make it real for learners. In Widdowson (2000) terms, a corpus comprises traces of texts, not discourse.
2. Teachers (and course book writers etc) should not accept corpus evidence uncritically, but should appraise it in the light of other sources of information about language such as introspection and elicitation. In particular, frequency should not be the only factor in deciding what to teach: how salient a language feature is should also be taken into account, as should how highly valued a language item is. Learners should be encouraged to be creative in their language use, and should not be restricted to clichéd utterances.
3. Corpora tend to comprise the language of native speakers only, whereas many learners will never communicate with a native speaker and/or are not interested in native speaker norms. In particular, the details of phraseology or collocations may be unimportant to a non-native speaker of English. Too strong a dependence on corpora of native-speaker English tends to de-value the language of non-native-speakers and to perpetuate colonialist attitudes towards English.

4. In a similar vein, learners should be allowed to approach language in a way they feel comfortable with. In many cases, this will be via grammatical rules and lists of lexical items. Learners should not be forced to approach English via 'lexical chunks' exclusively.

Some of these points can be taken as common ground. It would be very odd to suggest that language should not be contextualised within the classroom or that teachers should approach corpus evidence uncritically. We must stress the need for caution in extrapolating from a corpus to a language and the importance of careful thought in interpreting corpus evidence.

As Barlow (1996:2) comments, **using such powerful tools should not cause the researcher to become complacent and imagine that 'language' is now in the computer.**

REFERENCES

- BRUMFIT, C.J., 1985** Dictionaries, Lexicography and Language Learning. Pergamon Press Ltd and THE British Council, Oxford.
- DUBIN, F. & OLSHTAIN, E., 1986.** Course Design: Developing Programs and Materials for Language Learning. New Directions in Language Teaching. C.U.P. Cambridge.
- HUTCHINSON, T. & WATERS, C., 1989.** English for Specific Purposes. A Learning-Centred Approach. London: C.U.P.
- Miliani, M., 1992.** Teaching English as a Foreign Language. Algiers. O. P. U.
- BARLOW, M. 1996.** 'Corpora for theory and practice' International Journal of Corpus Linguistics 1:1-37.
- BERNADINI S. 2000.** Competence, Capacity, Corpora. Bologna: CLUEB.
- BIBER D., Conrad S. and Reppen R. 1998.** Corpus Linguistics: investigating language structure and use. Cambridge: CUP.
- BURNARD L. and McEnery, T. (eds.) 2000.** Rethinking Language Pedagogy formal Corpus Perspective. Frankfurt: peter Lang.
- KRESS, G. and Van LEEUWEN, T. 1994.** Reading Images: The design of visual communication. London: Rutledge.
- LEECH, G. 1992.** 'Corpora and theories of linguistic performance' in Svartvik (ed.) 105-122.
- LEECH, G 1997c.** 'Teaching and language corpora: a convergence' in Wichmann et al (eds.) 1-23.

- LONG, M. H. and CROOKES, G. 1992.** 'Three approaches to task-based syllabus design' *TESOL Quarterly* 26: 27-56.
- MINDT, D. 2000.** *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.
- SINCLAIR, J. M. and RENOUF, A. 1988.** 'A lexical syllabus for language learning' in R. Carter and M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman.
- WIDDOWSON H. G. 2000.** 'On the limitations of linguistics applied' *Applied Linguistics* 21:3-25.
- WILLIS D. 1990.** *The Lexical Syllabus: a new approach to language teaching*. London: HarperCollins.