Small-corpora Concordancing in ESL Teaching and Learning

Bruce Ka-Cheung Ma

City Polytechnic of Hong Kong

Abstract

For many years, computerised concordancing has been the domain of computational linguists, corpus linguists, lexicographers and dictionary compilers, working with large corpora of millions of words. The use of small-corpora concordancing in ESL settings is a relatively new application and has sparked keen interest among many researchers and teachers since the mid-80s. This paper discusses the use of small-corpora concordancing in the three domains of ESL: 1. syllabus design and evaluation, 2. classroom teaching, and 3. test construction. In particular, the classroom concordancing approach as an evolving ESL methodology is discussed with reference to its rationale, its potentials, its current applications and its impact. The paper concludes with some critical comments on what has been achieved so far with small-corpora concordancing and points out some directions for the future.

1.0 Introduction

This paper reviews from the perspective of ESL teaching and learning the roles and applications of small-corpora concordancing (SCC), focusing mainly on SCC for the classroom, and touches briefly, as far as present literature shows, on SCC in syllabus design and evaluation, and in test design.

1.1 What is concordancing?

The term concordancing originates from what has been known as concordances. The COBUILD dictionary defines a concordance as "an alphabetical list of the words in a book or a set of books which also says where each word can be found and often how it is used". Tribble (1990a) refers to a concordance as "a reference work designed to assist in the exegesis of biblical and other socially valued text". Tribble and Jones (1990) point out that concordances have been produced since the Middle Ages on popular works of well-known writers, such as the works of Shakespeare, and most of these have been undertaken manually, and, as one can imagine, painstakingly.

With the advent of the computer, concordances can be generated with the "speed and reliability" (Tribble and Jones, 1990) that perhaps manual concordancing could never match. As Sinclair (1991) puts it :

> Thirty years ago, ... it was considered impossible to process texts of several million words in length. Twenty years ago it was considered quite possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular. (p.1)

Tribble (1990b) also remarks, "the effort involved in such a task [concordancing], when taken manually, was intimidatingly large, and ... was more than most individuals would ever want to take on." This could be true even with any text more than a few hundred words, let alone texts measuring up to the millions. As Foulds (1991) observes, "the time required to do such a thing [text processing] on a regular basis for texts more than a few hundred words long would have been so great as to render the value, if there were any, totally uneconomic".

So the whole idea of computerised concordancing lies in making feasible what people might have always wished to do but have avoided doing because of the labour and time involved; and as computerised concordancing popularises, more and more people have come to realise its potential and subsequently embarked on various concordance-related projects in linguistic research and ESL applications.

The term concordancing is, however, generally used in the literature relating to ESL teaching and learning without a very clear definition. It is generally understood to refer to a way of analysing texts. Tribble and Jones (1990) describe concordancing as "locating all the occurrences of a particular word and listing the contexts" (p. 7), while Levy (1990) defines a concordance as "a collection of all the occurrences of a word, each in its own textual environment[,] together with references and word frequencies" (p. 178).

As computerised concordancing developed, manual concordancing disappeared as a matter of course and the term concordancing has become understood to be computer-based rather than manually performed, whenever it is used. In a review paper discussing the MSDOS concordancers, Higgins (1991) provides the following definition:

> A concordance of a word is a set of citations or line references, allowing every occurrence of that word within a corpus of text to be retrieved. (p.92)

What the computer does in concordancing is to display all the contexts in which a certain word or string appears in a text or collection of texts, called a corpus. Software employed to achieve this end is thus called concordancing software. Sometimes, computer programs are referred to as concordance generators or concordancers (Tribble and Jones, 1990) and sometimes a concordancing facility may be included as one of the functions of a set of programs for text analysis. COMPAID is one example of this (see Fang, 1991). Utility programs attached to a computer's operating system, such as FIND.EXE in the MS-DOS environment (Higgins, 1991), or home-made macros to be run under more sophisticated word processing packages, like WordPerfect and Microsoft WORD, can also serve the purpose. (see Tribble & Jones, 1990, pp. 84-89)

The ways in which a computer can display the context of a search word or key word may vary depending on the software used and the operation selected. The sentence concordance displays the sentences in which the search word is used, and paragraph concordance displays the paragraph (Johns, 1988). KWIC (key-word-in-context) concordances, by far the most widely used among researchers and teachers, display the search word in the middle, with as much context as will fit into the line which is truncated at either side (Tribble & Jones, 1990; Higgins, 1991). Concordances thus generated by the computer can be sent either to screen, or to printer as hard copy, or to file for future manipulation. An example of a KWIC concordance output is given in Appendix A.

1.2 Large-corpora concordancing

The use of concordancing software for text analysis has, for many years, been limited to the domain of computational linguistics and corpus linguistics, both being relatively new areas in the study of language, made possible by the advent and availability of the computer. These analyses have been carried out mainly with mainframe computers on very large corpora running into tens of millions of words. The interest in these analyses has stemmed mainly from the desire to provide objective descriptions of how the language really works, involving people like lexicographers and dictionary compilers.

Among the major projects, the most well-known include the COBUILD project carried out at the University of Birmingham (Sinclair, 1987), from which quite a number of dictionaries and reference works have been completed and marketed

commercially (Sinclair, et al., 1987, 1990). Other examples include the Brown University project on its Corpus of Present-day Edited American English (quoted in Yang, 1985) and the Lancaster-Oslo-Bergen (LOB) project at the University of Lancaster (quoted in Levy, 1990), and the JDEST project on English for Science and Technology (Yang, 1985) at the Shanghai Jiao Tong University.

1.3 Small-corpora concordancing (SCC)

Apart from concordancing with large corpora, there has also been a growing interest in the use of small corpora analysable with microcomputers. This growing interest coincides with the surge of interest in computer-assisted language learning (CALL) and is catalysed by an era when microcomputers are becoming more and more accessible to ESL teachers and researchers.

This interest in small scale corpora concordancing began in the mid-80's, most notably with the work of Higgins and Johns (1984), and Johns (1986, 1988), which stirred up a movement in SCC. The result of the movement is that computerised text analysis has been brought much more closely to teachers, course designers, materials developers and learners alike, and SCC as a tool for text analysis or as a pedagogic activity is increasingly brought to test and experimentation in various places all over the world where one or more microcomputers are available.

2.0 SCC for syllabus design and evaluation

As early as 1988, Sinclair and Renouf put forward the idea of designing a general English syllabus based on "the common uses of common words" as identified by the computer-generated frequency lists of the COBUILD corpus (Sinclair & Renouf, 1988). Using data from the same corpus, Willis and Willis (1988) further developed the idea and completed a general English course while Willis completed designing his lexis-based syllabus, called the lexical syllabus (Willis, 1990).

As far as SCC is concerned, Flowerdew took the lead in its application in syllabus and course design. Flowerdew (1991) used concordanced-based word counts to establish the relative importance of vocabulary items and provided criteria for syllabus selection and grading. Using a specialist corpora of transcription of Biology lectures, he compared the word frequencies with those in the COBUILD general corpora and observes some overall similarity and some significant differences. It is argued that these observations could form a basis for course design in ESP contexts. Flowerdew

suggests that SCC can be employed to identify useful items to teach, reveal syntactic patterns in which certain words occur and locate functional and notional areas which might be included in a syllabus (Flowerdew, 1991, pp. 38-39)

Ma (1993a), in his concordanced-based analysis of the genre of direct mail sales letters, discovered the attachment of certain mood and modality to distinct sequenced moves exhibited in his 50-letter corpus. Imperatives are found to abound in both the opening and action-getting moves but in the former they are never used with the polite marker *please*. *Can*, *will* and *may* appear in large numbers in the product-description move while *must*, *ought to*, and *should* hardly exist. Thematised purpose clauses with *For* or *To* belong to the overwhelming majority of the action-getting move. It is suggested that these observations should contribute valuable references for the design of syllabuses and materials of business writing courses where students need to write this kind of sales letter.

Apart from designing syllabuses, SCC can also be used for evaluating an existing course or programme and its materials (Flowerdew, 1991). In Flowerdew's corpus, connectors like *then* are found to appear between the subject and verb, rather then sentence-initial as taught in many published materials (p.38). The defining function is seen to be expressed almost entirely by the word *call* while commercially available materials tend to focus on the word *define*. Published materials are also found to have overlooked the intervening adverbials in many of the passive constructions (p.40). In Ma's corpus, on the other hand, the postscript component in a sales letter, shunned in most published materials as being a sign of poor planning, is shown to be the rule rather than the exception (Ma, 1993a). The language of refutation, which receives heavy emphasis in an EAP course, is refuted, quite ironically, by Pickard (1992) with reference to corpus evidence.

3.0 SCC for the classroom (Classroom concordancing)

The idea of using SCC in the classroom for the teaching of ESL, generally known as classroom concordancing (CC), is strongly supported by a number of researchers and applied linguists (notably Stevens, Johns, and Tribble & Jones).

3.1 Why: SCC as methodology

In the ESL classroom, concordancing is seen more as an approach to teaching or learning than as a way of text analysis. The rationale for the CC approach is one of

authenticity and discovery. Johns (1986) describes this concordance-based approach as data-driven learning (DDL). As the name suggests, this approach is characterised by language data taking on a primary role in language learning. Johns suggests that concordances provide "intake", (after Corder, 1967) i.e. the part of input that is actually helpful, to the language learner, which strikes a healthy balance between the "highly-organised, graded and idealised language of the typical coursebook" and the "potentially confusing but far richer and more revealing authentic communication" (Johns, 1986).

With regard to authenticity, Stevens (1988) points out the "realism and relevance" that CC can offer. While teacher-invented exercises for vocabulary can often contain inadvertently interjected artificiality, concordance-based materials "assures that contexts will always be real ones" and "relevance is achieved when the corpus of text used is appropriate to the language learners for whom the exercise is being prepared".

Johns (1988) further breaks down the idea of authenticity into three aspects: authenticity of script, of purpose and of activity. He believes that in CC, the teacher takes the role of an authentic text presenter rather than the traditional text preparator. Authenticity of purpose is achieved by concordancing texts that "students are having to work with on their courses or in their research" and authenticity of activity is achieved when what is done with the text is transferable to real world situations.

Levy (1990) thinks that concordances "present the facts of the language in a precise way" as they are based upon "actual usage". Concordance users are thus consulting "the source, the original instances of a word's use" rather than trying to peep at its usage via an intermediary, e.g. a dictionary. As Johns (1991b) states:

> What distinguishes the DDL (CC) approach is the attempt to cut out
> the middleman as far as possible and to give the learner direct access to
> the data, ... (p.30)

Johns (1991b) also sees CC as an attempt to contextualise and demythologise language. By looking at natural language in use, SCC "dispels the myths and distortions that have arisen from reliance on 'armchair' linguistics" and it also dispels the need for the language teacher to answer learners' queries by resorting to intuition alone.

As far as discovery is concerned, Johns (1988) points out that SCC is in line with the assumption that effective language learning is a form of linguistic research. He

believes that the teacher is potentially most effective when he or she is most at risk, and thus when the teacher is placed alongside the learners in attempting to solve communication problems, made possible by concordancing subject-related texts, the teacher is then able to gain valuable insights which might be otherwise inaccessible. (Johns, 1988)

In relation to the concepts of authenticity and discovery, Tribble and Jones (1990) point out that the real value of concordancing lies in the question of visibility. Concordancing software enables the user to visualise text features in ways that have never been possible. Tribble (1990a) describes the use of CC as "making the invisible visible" and he comments that CC is a "very new approach to the very old task of teaching and learning a language". Taking this visibility dimension of concordancing further, Rundell and Stock (1992) remark, "Perhaps the single most striking thing about corpus evidence ... is the *inescapability* of the information it presents." Though Rundell and Stock are speaking from a lexicographer's point of view, their comments are certainly applicable to learners using the concordancer as a language learning tool since learners also assume a role very similar to that of a linguistic researcher.

On the learner's road to discovery, the role of the computer and the concordancer is described as a special type of informant, giving the learner access to linguistic data (Johns, 1991a). Johns (1991a) describes this approach as a break away from the rule-based approach into the data-driven approach and identifies it as a kind of inductive learning where it differs from the traditional approach in that data replaces the teacher as the basis. It is believed that the CC approach can build learners' competence by giving them access to the actuality of linguistic performance.

3.2 How: getting a concordancer to work

There are two prerequisites for classroom concordancing. First, there must be the computer hardware and software which operate the concordancing and second, there must be a corpus for the computer to work on.

3.2.1 Software selection

Tribble and Jones (1990) make a distinction between three different types of concordancing software: streaming concordancers, text-indexing software and in-memory text consulters. Streaming concordancers read a text one line after another and produce concordances as they work through the texts. Text-indexers are those that

create an index of the text in one operation and then allow for different types of text retrieval activities, including concordancing. One example of these is WordCruncher. The last type, in-memory text concordancers, reads the whole text into the computer's working memory and then operates on it to show different types of information as desired by the user. Longman Mini-concordancer is an example. (Tribble and Jones 1990, p.13)

Tribble and Jones (1990) recommend, though rather implicitly, using in-memory concordancers for classroom concordancing. They point out that this type of software is limited by the memory size of the computer but has the advantage of a variety of text-handling capabilities once a file, or set of files, has been loaded. Streaming concordancers are seen as too slow to justify classroom applications while text-indexers are viewed as too sophisticated and should be left only to large-scale researchers. (p.14)

Higgins (1991), in his review of MSDOS concordancers, makes a distinction between three types of concordancers: dedicated research concordancers, dedicated classroom concordancers, and text utilities. Dedicated classroom concordancers are characterized by their "rapid results and clear displays", and are what he thinks to be appropriate tools for the ESL teacher in the classroom.

3.2.2 Corpus creation

While most people talk about concordancing with a corpus of some kind, it is worth pointing out that concordancing can actually be done with individual texts. Tribble and Jones (1990) point out that individual texts could be the target for concordancing if the objective is to analyse the language of that text. (p.15)

In corpus creation, a distinction is generally made between a general corpus and a specialist corpus, the choice depending obviously on the needs of the learners.

Tribble and Jones (1990) specify the following criteria for the creation of a general corpus for classroom use:

1. Use authentic, natural language
2. Use contemporary texts
3. Exclude archaic forms
4. Exclude dialect

5. Stick to prose

6. Exclude technical material (p.18)

To create a suitable corpus for the general English classroom, Tribble and Jones (1990) describe the following possible methods. Similar methods are also advocated by Sinclair (1991), who works mainly with mainframes.

1. Keyboarding
2. Optical scanning
3. Adaptation from ready-made text files, including word-processed documents, READ.ME files accompanying software packages, sources of text by access to a network or to colleagues. (Tribble and Jones, 1990, pp. 19-21)

As optical scanning facilities become more and more popular, with prices of high-technology products falling all the time, it can be expected that more and more people will take advantage of this convenient means of input for corpus creation, both in research and classroom applications, rather than relying on manual keyboarding. It must be pointed out, however, that the margin of error with most optical scanning hardware and software today is still disappointingly large, which makes them less than an ideal means of input. Any heavy reliance on machine-read operations must be offset by a sufficiently large corpus to make the database a useful and dependable one.

Tribble and Jones (1990), advising on teacher-created corpora for classroom concordancing, suggest accumulating a number of specialist corpora to form a general corpus. While seeing this as an easier job than trying to assemble a large general corpus at one time, they point out that accumulation in this way also addresses the need to achieve "balance and variety" in a general corpus, (p.16) though one might wonder how this could avoid including technical material, one of the principles Tribble and Jones (1990) put forward for general corpus creation.

3.2.2.1 Corpora size

It is said that "small corpora can play a subsidiary role in investigating specialised varieties of texts that are neglected in large corpora or where the classification systems of the large corpora are insufficiently delicate to recover the information required" (Johns 1986, p.158). But how small should a small corpus be? According to Tribble and Jones (1990), it appears as a general rule that, even working with small

corpora, a bigger corpus gives richer, more interesting and more representative information while too small a corpus may result in distortion (pp.15-16).

Tribble and Jones suggest that a corpus of 50,000 words should be very useful for classroom purposes (p.14). The corpus Tribble and Jones used in their experimentation, the ELT Text Pack Corpus, consists of texts from both written and spoken English running into 45,000 words, which is not as large as one might have imagined necessary. The rationale behind this 50,000 word threshold is unclear, but a study of the size of the corpora used by some of the researchers mentioned in this paper, as given in Table 1, will give a rough idea of how small small corpora generally are, noting that some of them are not meant for classroom use.

Table 1. SCC in ESL: Corpora size

Researcher	Corpus	Number of words
King, 1989	Academic lectures and tutorials	155,000
King, 1989	Scientific & technical journals	11,400
Tribble & Jones, 1990	ELT Text Pack Corpus	45,000
Tribble, 1990a & 91	English Historical Review Corpus	104,555
Tribble, 1990a & 91	Longman Corpus of Learners' English	54,861
Mparutsa *et al.*, 1991	Economics corpus	20,749
Mparutsa *et al.*, 1991	Geology corpus	33,687
Mparutsa *et al.*, 1991	Philosophy corpus	6,854
Johns, 1988	Transportation & highway engineering corpus	100,000
Johns, 1988	Plant biology corpus	100,000
Johns, 1991a	New Scientist Corpus	760,000
Johns, 1991b	Byte Corpus	>1,000,000
Johns, 1991b	Corpus of academic papers	250,000
Roussel, 1991	New Scientist Corpus	760,000
Flowerdew, 1991	Biology lecture corpus	104,483
Pickard, 1992	Applied linguistics papers	>50,000
Ma, 1993a	Direct mail sales letters corpus	16,345
Ma, 1993b	Computer software user manuals	52,000

3.2.2.2 Corpora type

While general corpora are thought by many (e.g. Tribble and Jones, 1990) to be useful for ESL, specialist corpora with ESP texts certainly address the needs of a particular group of learners with "relevance" (Stevens, 1988) and have a definite value in ESP settings. Levy (1990) says:

> Concordances drawn from a specific subject area (e.g. scientific texts),
> a specific mode (e.g. journalism) or a specific medium (e.g. spoken

language) can provide very helpful data on the range of words and their particular patterns of usage within a given context or genre. (p.179)

Tribble (1991) demonstrated the need to achieve what he called "face validity" in the use of corpora. With an analysis of speech-related verbs in one learner corpus and three different specialised native speaker corpora, Tribble demonstrated the need to use "corpus resources appropriate to the domain with which the students were already familiar" as different corpora, apart from showing up different words, are shown to have sets of words used in dramatically and interestingly different ways.

Although most existing corpora are collections of well-formed authentic native speaker texts, there is also value in assembling a specialised corpus of ESL learner texts. Johns (1986) suggests that concordancing with learner texts provides an excellent tool for examining recurrent patterns of errors or successes, and also for studying "the ways in which they manage to avoid syntactic and lexical problems in the target language" (p.159). Tribble and Jones (1990) takes a step further to suggest that corpora of learner texts, besides helping to identify and analyse learners' problem areas in lexis, grammar and semantics, could shed light on how the native language influences the way English is learnt as a second language.

In fact, King (1989) used a corpus of learner texts from his students studying English for Science and Engineering to compare with a corpus from professional scientific and technical journals and was able to observe the differences in the use of sub-technical vocabulary and to point out implications for teaching. So the use of a student corpus can have its value in informing the teacher and in helping to devise strategies for teaching before the teacher enters the classroom.

3.3 What: potentials and applications

The 'what' of SCC includes what can be done and what has been done with SCC.

3.3.1 What can be done

A lot has been said about what can be done with classroom concordancing. Johns (1988) suggests the following six main uses:

1. CC can be used as "a resource for small scale on-the-ground research by the teacher in order to inform teaching decisions".
2. The teacher can use concordance output to prepare teaching materials.

3. The teacher can incorporate concordance output directly in teaching materials and "devise activities that get students to puzzle things out for themselves".

4. Concordances can be used for "serendipity learning", which is the kind of free-ranging and open-ended linguistic enquiry made possible by the rich information concordances provide.

5. Concordancing can be used interactively as a focus of classroom activity.

6. The concordancer can be used as "a sleeping resource", offering help when the need arises.

Tribble and Jones (1990) summarise their suggestions of uses of concordancing in the following ways.

A. Using concordance outputs for:

1. deducing the meaning of keyword from context
2. study of grammatical features of particular words and of general grammatical features
3. study of homonyms and synonyms
4. group work activities
5. gapfill exercises
6. matching exercises
7. remedial exercises based on learners' own writing (p.55)

B. Interactive uses:

1. learning about grammar
2. vocabulary development
3. English for specific purposes

Levy (1990) strongly recommends the use of CC for the teaching of collocations, which he views as one of the most frustrating features of the language for students and teachers at higher levels. He believes that "a set of examples as given in a concordance would give the students the correct sense of how a word is used". (p.178)

Levy (1990) also suggests using on-line concordancing, and integrating it with a word-processor to give a fully integrated word processing environment. Concordances are seen as an explanatory device, useful for learners using the computer as an electronic writing tool. When a concordancer is integrated with a word processor with

a full dictionary and a thesaurus, the entire system will serve to answer a student's query about a word or phrase better than a dictionary, a concordance or a thesaurus alone.

The concordance contributes in the following activities when used in combination with the dictionary or thesaurus.

1. checking meanings
2. checking general syntax
3. checking usage
4. exploring special lexis especially ESP vocabulary
5. checking derived forms
6. checking collocates of words
7. exploring set pieces, e.g. phrasal verbs, clichés

Figure 1 shows a diagrammatic representation of Levy's idea of an ideal electronic writing environment.

[Insert Figure 1 here.]

Figure 1. Concordancers and word-processing for language learners
(From Levy, 1990)

For concordances to be useful, Levy (1990) contends that flexible selection mechanisms are necessary. Students need to be acquainted with the search and retrieval techniques used in concordancing software. He further suggests that the success of any concordance program depends on "flexible and efficient user interface" as well as the "quality and relevance" of the text corpora.

Levy suggests concordancing with

1. adjacent words ordered alphabetically,
2. common words, and
3. small specific corpora.

According to Levy, the teacher will need to have at his disposal all the large and small, general and specific, corpora in order that students can refer to the most appropriate corpus of text for a relevant use of concordances. But one cannot help wondering whether it is practically feasible, and worthwhile, to do so.

More concerned with how concordancing can be carried out to benefit learning, Honeyfield (1989) develops a typology of exercises based on concordance-based material and suggests a four-step procedure for concordance-based teaching activities, as follows:

1. The student becomes aware of a need for data, for information about how the language is used. Such awareness may arise from a more communicative task, such as writing a report, or from a more language-oriented exercise, e.g. a vocabulary or grammar exercise.
2. The student consults relevant concordance material, either through direct access to a computer or by using concordance material supplied by the teacher.
3. The student analyses the data and draws conclusions.
4. The student applies the insights gained to the task in Step 1. (p.44)

Flowerdew (1992) suggests a process approach to the teaching of professional genres and believes that concordancing has a role to play in helping students discover specific features of a genre or compare features of two genres.

Going beyond ESL learners, Berry (1993) suggests using concordance printouts to help language teacher trainees to increase their awareness of the language, the rationale for which in fact does not differ very much from that applied to ESL learners at an advanced level.

3.3.2 What has been done

Experimentation with CC has been reported by quite a number of researchers, though what has been reported may represent only the tip of the iceberg.

Apart from being seen as an approach in teaching and learning, SCC is also seen as a pedagogic activity. Stevens (1990) sees concordancing as a form of text manipulation activity, which can be seen as parallel to other forms of text manipulation such as text reconstruction activities with jumbled sentences or paragraphs. Taking it a step further, some ESL teachers take SCC as a type of lesson, which could parallel

listening sessions or writing workshops, and SCC in the classroom has thus been called concordancing sessions (e.g. in Mparutsa et al., 1991).

What has been reported in the current literature about CC applications falls into either pre-classroom or classroom use. Pre-classroom use of CC refers to the transformation of concordance outputs into teaching materials in the form of either overhead transparencies (OHTs) or paper-based classroom tasks or exercises. Classroom applications of CC, on the other hand, represent the interactive use of concordancing, sometimes also called on-line concordancing (Levy, 1990). This is where either the teacher directs the learners to generate concordances for discovery-type study of language features or language use, or learners are allowed self-access to the corpora for carrying out student-initiated linguistic enquiry and research.

3.3.2.1 Grammar and vocabulary teaching

Most reported work relating to CC in an ESL setting is concerned with the teaching and learning of grammar and vocabulary. Table 2 gives an overview of the kind of work reported.

Table 2. Summary of CC applications in ESL

Researcher	Year	Students	Teaching area	Nature of application
Johns	1988	Post-experience MSc students	'to' as infinitive or preposition; differentiating 'therefore' and 'hence'; the use of articles	Paper-based exercises and interactive concordancing
Tribble	1990a	Advanced students in EAP programme	Use of prepositions and articles	From paper-based exercises to interactive concordancing
Isle	1991	Students studying investing and financing in a vocational training programme	Subject-specific vocabulary	Interactive searching and selecting activity
Johns	1991a	Postgraduate research students	Comparing 'convince' & 'persuade'; the use of 'should'	Paper-based exercises
Johns	1991b	Postgraduate research students	The use of 'should', 'recommend', that-clauses, 'such as', articles and 'have to'	Paper-based exercises

Mparutsa et al.	1991	Undergraduate students from a teacher-dominated rule-based learning system	Subject-specific vocabulary	Interactive concordancing
Stevens	1991a	First-year science undergraduate	Subject-specific vocabulary	Interactive concordancing
Stevens	1991b	Undergraduates	Subject-specific vocabulary	Interactive concordancing
Taylor	1991	Learner teachers	Grammar in general	Edited concordance outputs presented on OHTs
Ma	1993b	Third-year students in a higher diploma course in computing	Use of the corpus to aid writing part of a software user manual	Interactive concordancing

As most discussion centers around the teaching of vocabulary, whether general or ESP, it is worth pointing out that Stevens (1988; 1991a; 1991b) puts forward a strong case for the teaching of vocabulary with classroom concordancing. Stevens (1991b) suggests selecting "the most revealing contexts for the same word" from concordance outputs for making gap-filling exercises with multiple contexts, which is argued to reduce the chances of error and increasing student confidence and improving performance (p.38). After students are familiar with how concordances can be generated, they can be directed to self-access vocabulary study by running what Stevens calls "exploratory concordances". An example of a concordanced-based gap-filling exercise, taken from Stevens (1991b), is shown in Appendix B.

An empirical study (Stevens, 1991a) comparing the traditional gap-fillers and the KWIC concordance-generated ones draws the conclusion that the latter can be seen as a viable alternative to the former. The pedagogical value of traditional gap-fill vocabulary exercises is questioned as an incorrect choice of word at the beginning could "compound the error" by taking away yet another contextual clue which might be needed for further decoding of the text. It is argued that, though not necessarily superior, concordance-based gap-fillers are more easily solved provided that students are given a brief familiarisation phase. Stevens claims that "the truncated demi-context typical of concordance output does not seem to be a hindrance to the their discerning the word missing from the contexts" and that the "multiple of disjunct contexts helps them more in settling on a correct word than do the clues inherent in a passage of discourse with the same words missing" (p.55).

3.3.2.2 Cross-Linguistic Parallel Concordances

Although CC does not seem at first sight to have any face validity for the teaching of pronunciation, Roussel (1991) advocates the use of Cross-Linguistic Parallel Concordances (CLPCs) for teaching tonic placement. Roussel carried out a study on transcribed speech of English and French and finds that CPLCs could be of help in teaching tonic placement related to auxiliary verbs in English. Roussel's experiment, using her own intuition about tonic placement with a largely written corpus, may be at fault. But the use of CPLCs-based exercises could indeed help heighten learners' awareness of the difference they have in the two languages they speak. And the opportunities for CLPCs to be used in the classroom for comparing two languages are no doubt open for more research and investigation, though for many pairs of languages, like English and Chinese, parallel concordancing is still far from being technically possible.

3.4 Impact of classroom concordancing

The impact of the CC approach is perhaps best summarised by Johns (1991a), who reports having used concordances in his teaching for four years with overseas postgraduate students. Johns claims that CC could have an impact on the process of learning, the role of the teacher and the place of grammar in ESL teaching. While John's first claim is supported by a number of practical applications of CC, his second and third claims remain unexplored and open to further research.

Johns claims that "concordances stimulate enquiry and speculation on the part of the learner", and help the learner "to develop the ability to see patterning in the target language and to form generalisations to account for that patterning." (p.2) He reports that by using interactive concordancing, his learners were able to provide more valid answers than the teacher could provide intuitively (Johns, 1991a).

This claim of Johns is supported by a number of researchers. Mparutsa et al. (1991) found that concordancing could help "develop students' learning skills with written text" as well as "promote independent and group learning". They also report changes in students' attitudes from the acceptance of the textbook as the supreme authority to having a more interactive and inquisitive approach to learning.

Taking it a step further, Taylor (1991) reports high transferability of discovery learning from concordance-based lessons when students showed better performance in subsequent text evaluation tasks. And Mparutsa et al. report cases where the student

was seen to "contribute his/her developing subject knowledge" and the teacher could "contribute knowledge of language functions", leading to an understanding of the text through joint-discovery (p.131).

In addition, a number of other researchers (e.g. Butler, 1991; Isle, 1991; Mparutsa et al., 1991) report boosted motivation with the new approach. Isle (1991) points out:

> The motivation is undoubtedly there: my students found the concordance program a fascinating piece of software and appreciated its potential for investigating and extracting information whether on facts and figures or linguistic questions. (p.107)

As regards the impact on the teacher, Johns concludes that the teacher's role is to have undergone a healthy change from the traditional roles to "a director and coordinator of student-initiated research". Syllabuses, teacher's key books and many traditional practices have to give way to the natural data of language and this role is a challenging one as there are a lot of new questions that remain to be answered.

And the third major impact observed by Johns is that the CC approach makes possible "a new style of grammatical consciousness-raising by placing the learner's own discovery of grammar at the centre of language learning". Johns theorises that "when grammatical description is the product of the learner's own engagement with evidence, that description may show a far greater degree of abstraction and subtlety than with a given description" and as a result the place of grammar in the ESL classroom has to be entirely re-evaluated.

4.0 SCC for test design

So far, references made to the use of SCC by teachers and researchers lie in a teaching-related context. Butler (1991) is perhaps the first person to use SCC in an ESL testing environment. Butler used SCC for test construction. He argues that SCC could improve the very popular cloze test in that the bias of the text content of just a single piece of text could be eliminated by concordance-generated tests of the gap-filling type where a test item appears in a set of different sentences drawn from a number of different texts in a corpus.

Drawing on Oller's (1979) idea that a cloze test "deals with contextually interrelated series of blanks", Butler (1991) believes that sentence concordance outputs can be

easily manipulated, with the use of word-processing software, to provide computer-enhanced cloze tests which, though not providing a complete discourse, meet Oller's criterion for a cloze test. What Butler did was to run a concordancer through a corpus and had it generate sentence concordances of certain selected words. The role of the test designer changes from that of selecting and/or modification of a text to selecting the test words and the appropriate citations. An example of the test Butler used is given in Appendix C.

Of course much of Butler's argument lies in whether one is convinced that a collection of sentence concordance outputs as such can be viewed as the equivalent of a "contextually interrelated series of blanks" suitable for the design of cloze tests and also whether one approves of the test being constructed without a complete discourse. The criteria for word selection which Butler used in his experiment remain unclear, and, although there was positive feedback from students (p.34), it remains doubtful whether the test so constructed was a valid and reliable one.

Anyway, Butler's reservations about the use of the current CBELT (computer-based English language testing) software programs, based on random deletion (Butler, 1991, p.33), are perhaps sensibly cast. The use of concordances and a corpus, supported by the expertise of the user, the teacher or the test designer, is obviously superior to just leaving the job to the machine and the extra time they spent could also be well justified.

5.0 Conclusion

As described above, SCC has been looked at with enthusiasm by most who believe in the use of authentic materials in second language teaching. Interestingly enough, even people who believe in having to exercise great control over educational texts may view concordancing positively and believe that concordancing with authentic texts can have a role to play. Foulds (1991) , for example, points out the value of concordancers in "monitoring and adjusting linguistic features" in pedagogic texts. (pp.47-53)

As with any application of new technology in the classroom or in research, both the researcher and the students are likely to get excited with it at the beginning. Whether it is going to stay there as a useful pedagogical tool will be subject to serious experimentation in different situations, using students of different backgrounds and levels.

SCC has stirred, and will no doubt continue to stir, a wave of excitement in the field of ESL teaching as more and more teachers try out SCC in their classrooms. SCC is now only in its infancy and it has been enthusiastically promoted by a number of people, especially Johns, Stevens, and Tribble.

However, not many of its applications are in fact revealed in the literature and not much of the learner feedback has been seriously examined. Most work on CC tends to slant towards the speculative rather than the evaluative end. Descriptions about learner responses tend to be observation-based rather than empirically studied. The influence of CC on the teacher and on the place of grammar has hardly been investigated. It remains doubtful whether teachers and learners can cope with the inherent technical problems of concordancing such as hardware operation, search techniques and output procedures so as to make concordancing sessions effective and worthwhile, without the lessons being turned into desperate attempts to get the hardware, software and database in the right place and the machines to work in the right way.

In particular, not many of the dangers of CC are ever cited, though obviously as work associated with a new technological tool, it could not be without any pitfalls. In corpus creation, for example, bias could be one, where owing to the inherent convenience of inputting texts in the written medium, the spoken aspect of the language could be easily neglected and this could result in learners having an unbalanced picture of the language. Overdependence on machine-reading type of text input and misuse of corpus creation criteria could well be other potential sources of danger.

So far, applications of CC seem to have been limited to students at the very advanced level and to the teaching of grammar and vocabulary. Much has still to be learnt about how it can be employed with students of a level much lower than those cited in the current literature, say with secondary school or primary school students. The value of CC in the teaching of macro aspects of the language, such as discourse level features, also remains unexplored. It should be obvious that CC cannot be the entirety of any ESL course and so the question remains as to how it can be integrated with other areas of a course so that CC can become most fruitful and rewarding. Materials developed from CC are not yet seen marketed for use by ESL population teachers (Johns is preparing to do this; see Johns, 1991a) and there is obviously a long road to drive before CC-conscious researchers will see CC popularized.

Other areas in SCC, like CLPCs, test design and the teaching of segmental or prosodic features in pronunciation are virtually virgin lands open for exploration and what SCC has in store for ESL is still waiting for teachers and researchers alike to put in more effort if the fruits of the technology are to be reaped for yet greater abundance.

References

Berry, R. S. (1993). Using concordance printouts for language awareness training. Paper presented at Second International Conference on Teacher Education in Second Language Teaching. 24 March, 1993.

Butler, J. (1991). Cloze procedures and concordances: the advantages of discourse level authenticity in testing expectancy grammar. *System 19(1/2)*: 29-38.

Corder, S. P. (1967). On the significance of learners' errors. *IRAL 5*: 161-170.

Fang, C.Y. (1991). COMPAID and its applications. In Milton J.C. & Tong, K. (Eds.), *Text analysis in computer assisted language learning* (pp.15-30) Hong Kong: The Hong Kong University of Science and Technology and the City Polytechnic of Hong Kong.

Flowerdew, J. (1991). Corpus-based course design. In Milton, J.C. & Tong, K. (Eds.), *Text analysis in computer assisted language learning* (pp. 31-43). Hong Kong: The Hong Kong University of Science and Technology and the City Polytechnic of Hong Kong.

Flowerdew, J. (1992). An educational or process approach to the teaching of professional genres. Paper presented at International Conference on English for Professional Communication, City Polytechnic of Hong Kong, 23-26 March, 1992.

Foulds, D. (1991). Using computers for language control in educational texts. In Milton, J.C. & Tong, K. (Eds.), *Text analysis in computer assisted language learning* (pp. 44-54). Hong Kong: The Hong Kong University of Science and Technology and the City Polytechnic of Hong Kong.

Higgins, J. (1991). Which concordancers: a comparative review of MSDOS software. *System 19*(1/2): 91-100.

Higgins, J. & Johns, T. (1984) Computers and language learning. London: Collins.

Honeyfield, J. (1989). A typology of exercises based on computer-generated concordance material. *Guidelines, 11*(1):42-50

Ilse, W-R. (1991). Concordancing in vocational training. *ELR Journal (New Series) 4*: 103-114.

Johns, T. (1986). Micro-concord: a language-learners research tool. *System 14*(2): 151-162.

Johns, T. (1988). Whence and whither classroom concordancing? In Bongaerts, T., et al. (Eds.) *Computer applications in language learning* (pp. 9-33). Dordrecht: Foris.

Johns, T. (1991a). Should you be persuaded: two examples of data-driven learning. *ELR Journal (New Series) 4*: 1-16.

Johns, T. (1991b). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *ELR Journal (New Series) 4*: 27-46.

King, P. (1989). The uncommon core: some discourse features of student writing. *System 17*(1):13-20.

Levy, M. (1990). Concordances and their integration into a word-processing environment for language learners. *System 18*(2): 177- 188.

Ma, K.C. (1993a). Text analysis of direct mail sales letters. In Boswood, T., Hoffman, R. and Tung, P., (Eds.), *Prospectives on English for Professional Communication*. Hong Kong: City Polytechnic of Hong Kong.

Ma, K.C. (1993b). Classroom concordancing: what and how students learn. Unpublished MA dissertation, City Polytechnic of Hong Kong.

Mparutsa, C., Love, A. & Morrison, A. (1991). Bringing concord to the ESP classroom. *ELR Journal (New Series) 4*: 115-134.

Oller, J.W. (1979). *Language tests at school.* London: Longman.

Pickard, V. (1992). Should we be teaching refutation? Concordanced evidence from the field of applied linguistics. Paper presented at the Eighth ILE International Conference, 15-18 December, 1992.

Roussel, F. (1991). Parallel concordances and tonic auxiliaries. *ELR Journal (New Series) 4*: 71-102.

Rundell, M. & Stock, P. (1992). The corpus revolution. *English Today 31*: 21-31.

Sinclair, J.M. (1987). *Looking up: an account of the COBUILD project in lexical computing.* London & Glasgow: Collins.

Sinclair, J.M. & Renouf, A. (1988). A lexical syllabus for language learning. In Carter and McCarthy, (Eds.), *Vocabulary and language teaching* (pp. 140-160). London: Longman.

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

Sinclair, J.M. et al. (Eds.), (1987). *Collins COBUILD English Language Dictionary.* London & Glasgow: Collins.

Sinclair, J.M. et al. (Eds.), (1990). *Collins COBUILD English Grammar*. London & Glasgow: Collins.

Stevens, V. (1988). Studying vocabulary using concordances on microcomputers. *TESOL Newsletter 22*(3): 27.

Stevens, V. (1990). Text manipulation: What's wrong with it anyway? *CAELL Journal 1*(2): 5-8.

Stevens, V. (1991a). Concordance-based vocabulary exercises: a viable alternative to gap-fillers. *ELR Journal (New Series) 4*: 47-62.

Stevens, V. (1991b). Classroom concordancing: vocabulary materials derived from relevant, authentic text. *English for Special Purposes Journal. 10*(1): 35-46.

Taylor, M.B. (1991). Discovering concordancing. *CAELL Journal 2*(1): 25-26.

Tribble, C. & Jones, G. (1990). *Concordancing in the Classroom.* Harlow: Longman.

Tribble, C. (1990a). Concordancing and an EAP writing programme. *CAELL Journal 1*(2): 10-15.

Tribble, C. (1990b). Small scale corpora in ELT: an investigation into vocabulary use. *CAELL Journal, 1*(4): 13-17.

Tribble, C. (1991). Some uses of electronic text in English for academic purposes. In Milton, J.C. & Tong, K. (Eds.), *Text analysis in computer assisted language learning* (pp. 4-14). Hong Kong: The Hong Kong University of Science and Technology and the City Polytechnic of Hong Kong.

Willis, J.D. (1990). *The lexical syllabus.* London & Glasgow: Collins.

Willis, J.D. & Willis, J.R. (1988) Collins COBUILD English Course. London & Glasgow: Collins.

Yang, H.Z. (1985). The use of computers in English teaching and research in China. In Quirk, R. & Widdowson, H.G. (Eds.), *English in the world: teaching and the language and literatures* (pp. 86- 100). Cambridge: Cambridge University Press.

(from Ma, 1993)

Appendix B Example of concordance-based classroom exercise

(From Stevens, 1991b)

Appendix C Example of concordance-based cloze test

(From Butler, 1991)

Table of Contents