

Learning strategies in ESP classroom concordancing: an initial investigation into data-driven learning

*Bruce K C Ma
City Polytechnic of Hong Kong*

Introduction

This paper reports findings from an application of classroom concordancing, or data-driven learning, in an ESP setting. It describes the learner strategies of a group of 18 subjects in using a small specialist corpus to help them cope with the language for an ESP writing task.

Classroom concordancing as ESL methodology

Computerized concordancing has long been a way of text analysis and its use has been limited to the domain of computational linguistics and corpus linguistics. These analyses with concordancing have been carried out mainly with mainframe computers on large corpora of millions of words. In recent years, there has also been a growing interest in the use of small corpora analyzable with microcomputers. This interest coincides with the surge of interest in computer-assisted language learning (CALL) and is catalyzed by an era when microcomputers are becoming more and more accessible to the whole population of ESL teachers and researchers.

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

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). 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.

As small-corpora concordancing penetrates the ESL classroom, it is developing into an ESL methodology and is strongly supported by a number of researchers and applied linguists, notably Stevens (1988, 1990, 1991a, 1991b), Johns (1986, 1988, 1991a, 1991b), and Tribble & Jones (1990). The rationale for this methodology, called classroom concordancing (CC), is one of authenticity and discovery. Johns (1986) names this concordance-based approach 'data-driven learning' (DDL). As the name suggests, this approach is characterized by language data taking on a primary role in language learning. In Johns' view, 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-organized, graded and idealized language of the typical coursebook" and the "potentially confusing but far richer and more revealing authentic communication" (Johns, 1986).

Previous applications of CC

CC applications reported fall into either one of the two categories, pre-classroom or classroom use. Pre-classroom use of CC refers to the transformation of concordance output into teaching materials in the form of either overhead transparencies or paper-based classroom tasks or exercises. Classroom applications of CC, on the other hand, represent the interactive use of concordancing. 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. Most reported work relating to CC in an ESL setting is concerned with the teaching and learning of grammar and vocabulary. For details, see Ma (1993).

Little work has so far been done to look into ways that learners adopt to facilitate learning in concordancing. The present study follows Levy's (1990) suggestion of integrating on-line concordancing with a word-processor to give a fully integrated word processing environment. Levy views concordances as an explanatory device, useful for learners using the computer as an electronic writing tool. It is said that when a concordancer is integrated with a word processor with a full dictionary and a thesaurus, the entire system can serve to answer a student's query about a word or phrase better than a dictionary, a concordance or a thesaurus alone. As most tertiary students nowadays use some kind of word processing regularly for their writing, it is the researcher's belief that Levy's suggestion has great relevance to students learning to write in the institution where the researcher teaches. This study aimed to provide for students an electronic writing environment very similar to what Levy suggests,

though not exactly the same, so that the learning strategies they would use could be investigated.

Learning strategies: definitions and typology

A lot of research and discussion has taken place on learning strategies employed by learners in learning or acquiring a second language. Most of the learning strategies investigation relates to their application to the four language skills (see Wenden and Rubin, 1987; Oxford, 1990; Chamot and O'Malley, 1990).

Oxford (1992) defines language learning strategies as "specific actions, behaviors, steps, or techniques that students use to improve their progress in developing L2 skills". Looking at strategies used by learners from a broader perspective, Rubin (1987), defines learner strategies as "any set of operations, steps, plans, routines used by the learner to facilitate the obtaining, storage, retrieval and use of information, that is, what learners do to learn and do to regulate their learning". In Rubin's terms, learner strategies are further broken down into learning strategies, communication strategies and social strategies. Of these, learning strategies, ones that are directly related to learning, can be either cognitive or metacognitive. Chamot (1987) thinks that it is most useful to organize L2 learning strategies into three main categories: metacognitive, cognitive and social-affective (p.80).

Since the use of the concordancer as a learning tool is conducive to highly focused activities rather different from those going on in a typical ESL classroom, it is suggested that, for analytical purposes, a different typology of learning strategies should be developed with respect to applications of CC in the ESL classroom.

Using Rubin's (1987) and Chamot's (1987) classifications as a point of departure, Table 1 shows a typology which the researcher has developed by putting together information from Flowerdew (1993), Tribble (1989), and Tribble & Jones (1990) and the researcher's own experience and introspection in using the concordancer. The distinction between metacognitive and cognitive strategies is used as a basis and it is recognized that when enquiry-based concordancing is carried out, a concordancer user is typically going through the activities of formulating research questions, carrying out linguistic research, analyzing data, and finally drawing conclusions from the data displayed by the concordancer.

It must be emphasized that this typology is by no means a perfect one and should not be treated as definitive. It is of interest to the researcher to see which strategies they will develop, if any, and whether other strategies, which were not at first obvious to the researcher, will be developed.

[insert Table 1 here]

Research design and procedure

An ethnographic approach was adopted in this study, and the major instruments were observation, diary studies, questionnaires and interviews. The tasks that a student subject had to undertake in this project follows Honeyfield's (1989) four-step procedure (p.44).

Setting and subjects

This study was carried out at the College of Higher Vocational Studies, City Polytechnic of Hong Kong. The 18 students chosen to take part in the study, numbered S1 to S18, were a group of part-time evening students studying for a Higher Diploma in Computer Studies. The majority of the group were male students and there were only 3 females. They were between 20 and 35 years old, had a pass in English in the Hong Kong Certificate of Education Examination, and were studying their last year of a three-year course. The module these students were working on was 'English Communication Skills for Computing III', which extended over a period of 10 weeks, with students meeting for a consecutive 3-hour period every week. One of the objectives of this module was to develop the students' skills in writing and producing a piece of computer-related user documentation.

Tasks and environment

With the help of a concordancer, the students in this study worked on a corpus created by the researcher specifically to help with the language they were going to use in the writing task. A total of at least 10 hours' time was allocated to this concordancing work where they were asked to write one or more chapters of a user manual with the help of the concordancer if necessary.

Their work was carried out at one of the terminal rooms at the Computer Center of City Polytechnic of Hong Kong. They were provided with computer terminals linked on an institute network where a variety of word processors are available. These terminals were made up of IBM-compatible machines with 80386 central processing units and a random access memory (RAM) of 4 megabytes. Two printers were available in the room for production of hard copies. The *Longman Mini-Concordancer (LMC)*, which generates keyword-in-context (KWIC) concordances, was provided on floppy disks together with the text files making up the corpora. For most of the time, the students used a word processor as they wrote and were free to

choose one that they felt comfortable with. Some used *WordPerfect 5.1*, some *WordStar 5* or *6*, and some others working under *Windows*, *Windows Write* or *MicroSoft Word for Windows 2*. Most of these word processors come with a dictionary and a thesaurus, and provide an electronic writing environment very closely resembling what Levy (1990) suggests. Some were even able to toggle between their word processing program and the *LMC*.

The initial corpus used in this study contained about 52,000 words. It was created specifically for its purpose. As students were required to write part of a user manual for a computer software program or package as their target writing task, the corpus was assembled from a variety of software user manuals written for 14 different commercially available software packages. This corpus will be called corpus A.

After reviewing a variety of these manuals, the decision was made to choose one of the most common chapters, the start-up and/or installation chapter, to form the basis of Corpus A. This chapter typically describes the contents of the software package when it is sold, the requirements on the computer system for it to run, and procedures for installing the software on the system of the user.

After students had had 5 hours' work with the concordancer, another corpus of about the same size was additionally provided for students to work with. This was done in response to some students' demand for a bigger and more revealing corpus. This second corpus, which will be called Corpus B, was assembled from ready-made text files which come along with a number of shareware programs as user documentation files. In terms of the scope of the user manual, Corpus B, though on a relatively smaller scale, was more extensive than Corpus A. Unlike Corpus A, which was created with only selected chapters of user manuals, these on-disk manuals were meant to be complete manuals for the particular shareware programs.

It is believed that both corpora exhibit what Johns' (1988) calls authenticity of script, of purpose and of activity, as students graduating from the course under study will be expected to be engaged in writing software user documentation at one time or another in their career. In fact, a few students reported that this was part of what they had to do in their present jobs.

Research instruments

Students were asked to fill in a concordance diary, which is a form for recording their research questions, search strings, information obtained and conclusions drawn. Students were instructed that they should fill out the form in the greatest possible detail. It was emphasized on many occasions that they should record every search they made. Appendix A shows the diary kept by a student subject in one of the sessions.

The researcher and teacher in this study primarily played the role of an organizer of activities, a source of help, and a participant-observer. Observation was carried out throughout the ten sessions and important information was noted. In the first five hours of writing and concordancing, teacher marking was done. Errors which it appeared to the teacher would benefit from concordancing were marked, following a similar suggestion by Johns (1988, p.161), so that students could look to the corpus for a way of revision or improvement.

Two questionnaire surveys/interviews were carried out in this study. The first one was conducted when students had had five hours' work with the concordancer, as an interim analysis. The other one was conducted at the end of the study, that is, after students had had at least 10 hours' work on concordancing and writing. Each time, a questionnaire was prepared and given out to students. In addition, 5 students were randomly chosen to take part in an interview at the end of 5 hours' work, and 6 at the end of the study. These interviews had the objective of getting more in-depth information about their concordancing behavior, including their learning strategies and learning outcomes.

Learning strategies: description and discussion

An examination of the learning strategies students used should give information about how students learned. In this section, the typology developed earlier is used as a basis for analyzing the strategies the students used.

Research question formulation strategies

As far as research question formulation strategies are concerned, students generally showed a lack of maturity in formulating relevant and effective enquiries. When asked in the questionnaire or at the interviews how they came to make a certain search with

the concordancer, most failed to give a very definite answer. The following are some typical responses:

To find some words/sentences that can better describe my ideas. (S7)

I want to get information related to my writing in meanings or correct format of the words. (S10)

When I don't know how to write a sentence to express my idea. (S12)

When I am not sure how to use a word. (S17)

Others talk vaguely about the usage of words or the spelling of words. This feedback shows that the overwhelming majority of the research question formulation strategies were in the area of **word level contextual research** (see Table 1). Functional research questions were almost absent. The only research questions formulated which went beyond the word level were the following:

I want to search some information or formation about the introduction in user manual. (S8)

How to call diff[erent] type of user? (S18)

Even when contextual research is carried out on the word level, the research question formulated can be relevant or irrelevant to the corpus, or in Johns' (1988) terms, may be "concordance-ready" (p. 25) or not. An irrelevant research question will bring up no concordance lines simply because the word is not in the corpus. The following examples of searches all fell into this category, when the student concerned had questions to ask about words lying in a subject area quite remote from what the present corpus could cover:

<u>Student</u>	<u>Searches made</u>
S2	point, sales, stock
S5	sell, sold
S2	pl* (looking for the word 'plural')
S8	*thesaur* (looking for the word 'thesaurus')

S2, writing on an electronic point-of-sales system, searched for words like 'point', 'sales' and 'stock'. The same subject, S2, looked for the word 'cashier' in the corpus in her last concordance session in the project, which still appeared to be out of the range

of what should be expected from a corpus of software user manuals. There is no concrete evidence to show that students have become more competent in developing relevant enquiry-based searches on their own initiative.

Some research questions, though well and sensibly asked, were not rewarded with good enough evidence of usage from the corpus. Again, the question is one of relevance. A question directed to compare the use of the words 'chance' and 'opportunity', made by S18, could perhaps be well supported by a more general corpus. The specialist corpus used in this study only threw up 2 lines each, from which the student found it quite inadequate to draw any useful conclusion.

Linguistic research strategies

Linguistic research strategies are those relating to how linguistic research is carried out. There are 3 major ways in which linguistic research can be done. They are studying concordance outputs, studying the wordlist, and studying the wider context of a particular concordance line.

Studying concordance outputs is by far the most frequently used of the linguistic research strategies. It is interesting to point out that this was done entirely on screen, rather than on hard copy. It can be understood that the limited availability of printers in the terminal room was a factor deterring production of hard copies, but the fact that no one ever attempted to get any printing done could reflect a perceived lack of need to print concordance information on the part of the students.

In fact, in the interview, the following remarks emerged:

It will waste of time and paper, I think. ... No need to do that. (S2)

It's not necessary. Usually I get the input from the screen, I think it's rather enough. (S3)

Students generally found it easy to read from the screen and felt comfortable about it.

Studying the wider context of a citation was reported to be a widely practised strategy. When students spotted a certain concordance line which bore significant relevance to their own writing, they chose to view the wider context of the line. The *LMC* provides a facility to do this in a very convenient way. When the user highlights a certain concordance line with the up or down arrow key, pressing <Enter> will show him or

her the wider context of the search word by displaying as much context as the screen may show. It was observed that some students liked to do this. Although the context which was revealed was no more than a screenful of words, some students slipped into this environment regularly to immerse themselves in the wider context of the corpus in the hope of getting a better picture of how certain words are used or how certain ideas could be expressed. Some even hoped to borrow larger chunks of relevant texts for use in their own writing. It could perhaps be argued that this is akin to plagiarism. One student honestly remarks in his questionnaire:

Sometime[s], we may depend on the concordance too much. The use of the ph[r]ases or sentence will just copy from the database of concordance without thinking th[o]roughly. (S8)

It can be argued that, if used properly, this is an effective strategy for trying to find out more about the contextual environment of certain search words. This overcomes the limits of decontextualization inherent in individual truncated concordance line on the screen or even in the hard-copy output of any KWIC concordance operation. It would certainly be hard to draw the line between a legitimate use of this view function and plagiarism. Given the many facets of differences between the target software students were writing on and the software on which the corpus was built, these viewing activities, it is believed, should be regarded as something beneficial and healthy to writing rather than as plagiarism. Therefore, viewing the wider context of a citation should be considered an effective strategy in the use of the concordancer.

The *Wordlist* command was seen to be very seldom used, except for one student who developed a strategy of research by starting with the wordlist. He says in his questionnaire, "I select those words that occur frequently in the text but I scarcely use." (S16) This could be a commendable strategy for students when they do not have a good idea of what to look for as a novice in concordancing. Apart from this, students rarely visited the wordlist, let alone took advantage of any information that might be contained there. When asked whether they used the wordlist for anything, most gave a definite negative answer.

Search strategies

Search strategies refer to ways of putting together a search string for the concordancer to act on. There are two commands with which students can make a search. They are *Concordance : New Word* and *Concordance : Add Word*. *Concordance : New Word* was reported to be the most frequently used command of the concordancer.

Table 2, from results of the questionnaires, gives a profile of how students made their searches with these two commands. The rating of 1 means that the strategy was never used, and 5, regularly used.

	Mean rating	Rank
whole word + *	4.20	1
whole word	4.13	2
part of a word	2.73	3
* + whole word	2.60	4
* + whole word + *	2.47	5
2 words	2.00	6
3 words	1.53	7
more than 3 words	1.40	8
whole sentence	1.13	9

Table 2 How searches are made

An examination of the search strings students recorded in their diaries showed results closely resembling those of Table 2, except that whole-clause or -sentence searches did not exist. Most in fact started out using whole word searches but soon realized the power and convenience offered by the wildcard and were able to use it in their searches. The use of the wildcard character can be viewed as an easy and natural development in students' searches. In this study, it was discovered that when the wildcard was used, only the '*', which can stand for one or more characters, was used. The other wildcard character '?', which stands for 1 character only, was not seen to have been used by any student, though it was introduced to them at the briefing and exploratory session. Moreover, when they used the '*', they mostly used it at the end of a search string and seldom used it at the beginning, or in the middle. The potential of the wildcard characters are obviously not taken full advantage of.

It could also be seen that some tended to overuse the wildcard by putting it in every search. The disadvantage of this is the generation of an over-abundant number of citations that may render analysis difficult, particularly when analysis was done on screen. If the users were sufficiently adept, they would have narrowed down the search by getting rid of the wildcard, or made follow-up searches by using a more restrictive search string.

Another interesting point concerns the choice of the search string. A distinction can be made between a good choice and a bad one. A good choice constitutes effective and efficient use of the concordancer. S3, when told there was a problem with the expression 'into the appropriate disk drive for your computer system', looked for 'for' and then 'of' in his searches, each generating hundreds of lines. In doing so, he depended on his having a hunch for the word 'of', rather than on the revealing power of the corpus of data itself. A better choice of the search string would have been the word 'system', which would have generated fewer lines and made his data analysis easier and more effective. S11 searched for the word 'master' to see if 'master menu' was a good term to use. He obtained the response "The word is not in the corpus." from the concordancer. A teacher-prompted search with the word 'menu' directed him to get the collocation 'main menu', which he could then use confidently in his writing.

Data analysis strategies

Apart from enquiries about spelling, which require only examples of the word in question, all other forms of linguistic enquiry call for some kind of analysis of the data presented by the concordancing software. Since few of the students in this study used information from the wordlist, strategies relating to the analysis of wordlist information do not warrant any discussion in this paper. The following discussion will focus on strategies used to analyze information from KWIC concordance outputs.

Analyzing the information provided by the concordancer is by no means an easy task, and data analysis strategies are difficult to develop. Responses like the following came up as typical with students interviewed both during the interim and the final session.

The output maybe is useful but I can't take the information out from the output. (S3)

A lot of information come out and we have need to analysis the ... the text which part is useful to you. I can't ... maybe I can't use it very well in getting the meaning of the words. (S3)

I feel confused with the concordance output because always I don't know how to interpret the output. (S18)

A lack of data analysis strategies could render efforts in concordancing unrewarding. Sometimes, students were seen to have formulated very sound and sensible research

questions, but they were unable to analyze the data to draw any useful conclusion. For example, S18 intended to compare 'provide', 'offer' and 'give' and see which of these words could fit what he wished to say. Plenty of examples showed up. Unfortunately, he failed to exercise effective data analyses strategies and reported that he had had 'nothing special' to learn from the corpus.

Nevertheless, many students did develop certain strategies to analyze the concordance outputs. From the concordance diaries students kept, it can be seen that most students employed the strategy of **studying the context around** the search string.

They did this by **looking for appropriate words** either before or after their search string. They used the up and down arrow keys and the [**Page Up**] and [**Page Down**] keys on the computer keyboard to scan for words or phrases which might be useful.

Counting was another strategy used by many to decide whether certain words or combination of words are typical or used often enough to be safely used. For some, a single occurrence would be adequate, and this is true for spell-check concordancing. For some others, a lot more examples were needed. S15 reported in the end-of-project interview that he sometimes was not very confident if there were not a sufficient number of lines showing the same use. He said when he talked about his attempts to search for the right preposition,

Sometimes, I suspect about ... , not sure if the preposition is correct.
... the frequency of use is too low. (S15)

He said he would feel confident if about half the outputs showed the same pattern.

It was observed that the ***SORT*** and ***PROXIMITY*** commands were not often taken advantage of by the students. This is to say that they failed to develop data analysis strategies by sorting concordance lines or by seeking words in the proximity of the keyword.

Conclusion-drawing strategies

Conclusion-drawing strategies are difficult to uncover in this study because of two reasons. First, they were not obvious from concordance diaries. Second, even during interviews, students generally were not competent enough in their spoken English to verbalize what they did when they attempted to draw conclusions from the data

shown. However, a few strategies were identified. They are verification, inferencing, and identification.

Verification

The strategy of verification was mostly used by students to confirm their hunch about the spelling of words. As mentioned earlier, a single instance of the word searched would be adequate for the student to draw his conclusion. This strategy was also used to confirm the use of prepositions, derived forms of words and the existence of a certain word or piece of jargon. The following are some examples of the use of the strategy of verification, as shown in students' concordance diaries:

<u>Subject no.</u>	<u>Search string</u>	<u>Verification</u>
S4	sop*	Spelling: sophisticated
S4	act*	Spelling: activate (not activite)
S9	download	
	down load	Word form: download (1 word, not 2)
S9	newly	Existence of the word: newly
S12	sch*	Spelling: schedule
S15	destin*	Spelling: destination
S19	hard*	Word form: hard disk (2 words, not 1)
S19	copy-prot*	Computer jargon: copy-protected

Inferencing

Inferences were often made on account of the frequency of occurrence of words, but there is a potential danger in this strategy. For example, S9 decided that 'both X' was more common than 'both of X' as he saw more citations of 'both X' in the corpus. S3 decided that 'of the system' was better and gave up 'for the system' because the former occurred 4 times and the latter 3. He also decided to use 'produce a hard copy', instead of 'generate a hard copy', since he obtained 3 lines from a concordance of the word 'produce', including 'produce a report', and none from 'generate'. S2 gave up using 'key in' as a verb because he could not see it used in the corpus. S5 concordanced for 'document', her research question being "Can I use 'document' as a verb?". The output contained 78 lines of the word all used as nouns. So her conclusion to the question was a simple and straightforward "No." S4 wanted to see if he could talk about a 'large-sized document' and searched for the word 'document'. When she was unable to get any example of the term used, he decided to use simply 'a large document' instead.

These examples demonstrate that it could be dangerous relying on the occurrence of the search word in the corpus when making inferences.

Johns (1988) has warned of this, "... there is the danger of forgetting the limitations of the data on which it is based, a 'possible tendency' being identified as a 'definite tendency' and a 'definite tendency' being interpreted as a 'definite rule'." (p. 25)

Some successful use of inferencing strategies include the following:

<u>Subject no.</u>	<u>Search string</u>	<u>Inferencing</u>
S3	software	'software' does not take the indefinite article
S5	most	'most' takes a definite article in comparison
S17	fie*	'field' can be used in the plural
S17	information	'information' is always singular

Identification

Identification strategies were used to establish certain collocational patterns, use of plural endings, use of articles and to identify word class.

The following are some examples from students' concordance diaries:

<u>Subject no.</u>	<u>Search string</u>	<u>Identification</u>
S1	mak*	Verb pattern: make something easier
S1	requirement	
	requirements	Use of plural in 'System requirements'
S1	operation	Collocation: perform an operation (not 'do an operation')
S2	menu*	Use of hyphen: 'menu-bar'
S3	attach	Use of preposition: attached to your computer (not 'with')
S9	versatile	used as an adjective
S11	menu	Collocation: 'main menu' (rather than 'master menu')
S16	def*	Use of article: the default

Strategies to deal with confusing outputs

Many students reported that they sometimes felt confused by the output generated on screen. This confusion, or even frustration, arose when there were too many lines in

the output, making it very confusing to the eye. As a result, students were unable to get any useful information from it, let alone draw any meaningful conclusion.

In the course of their concordancing, students also developed strategies to deal with confusing output on screen. With confusing output of more than a screenful, it is not easy to read it from the screen. Since printing facilities were only barely available in the terminal room where students were working, printing of concordance lines was not considered by any one in the group to be convenient or efficient. Instead, students developed strategies to get round the problem by making alternative searches, usually with the use of the wildcard character or with a different search string, to get a neater and more specific output, or by using the ***SORT*** and ***PROXIMITY*** commands, to rearrange the output in ways more ready for scrutiny.

The following remark from a student's questionnaire reflects what was going on.

... because some output from concordance are useless or make me confuse, I need to search other word or ph(r)ase from that outputs. (S8)

Strategies to deal with nil outputs

In this study, 18.2 % of the searches made did not generate any output. This is not a high percentage, but for individual students, for example S1, the percentage could come to be as high as 50 %. There are several causes of a nil output.

First, the search string could be wrongly spelt. For example, S1 searched for 'inexperient', when in fact the word he wanted to use was 'inexperienced'. Second, the search was an irrelevant one. This usually happened when students looked for some highly specific vocabulary in another subject area. For example, S2 searched for 'sales' in an attempt to get something about the point-of-sales systems. Third, the word in the corpus was used in a slightly different form from that of the search string. For example, S1 searched for 'requirement' and it did not show up at all, whereas a follow-up search with 'requirements' showed a revealing set of examples talking about system or hardware requirements.

Obviously, there are strategies to avoid generating and to deal with nil outputs. Students generally were able to make better and more relevant searches as they gained more experience in their concordancing work. For example, S8 searched for 'Before to perform' in an attempt to find out what was wrong with the expression which he

wrote. A nil output was the inescapable result as the expression was in itself erroneous. His follow-up search was simply the word 'Before' and he was able to see that after 'Before', an -ing form of the verb should be used instead.

Social-affective strategies

In most learning situations, the use of social-affective strategies like seeking peer assistance or cooperation can be of value to the language learner and students are sometimes encouraged to develop cooperative and questioning strategies to obtain help from peers (See Chamot, 1987; Oxford, 1990.).

In this study, however, peer assistance was seen to be only of minimal importance. All students interviewed reported that they never or seldom resorted to help from or cooperative work with their peers. Requests for assistance from the teacher were also minimal.

Unused strategies

There are quite a number of strategies in our typology that appear to be never or nearly-never used. This may be for the reason that students could not handle the strategies concerned, or that there was no need to use such strategies.

First, metacognitive strategies, in this study research formulation strategies, were poorly mastered. Students rarely planned what to look for in the corpus. Most students reported in the interviews that they had difficulty in deciding what to look for in the corpus. No one in the group really used the corpora to find out the language characteristics of the corpus before attempting to write on their own. The end-of-project questionnaires also show that students in general thought that concordancing was not very useful in the planning stage of their writing, but was more useful in the rewriting and editing stage. Punctuation search strategies are also rarely seen.

Conclusions and recommendations

From the results of this study, some recommendations can be made about how CC can be improved, as far as development of learning strategies is concerned, and what future research can be carried out.

Encouragement of certain learning strategies

Irrelevant searches wasted time and rendered the writing process inefficient. It is therefore necessary that students should learn to develop enquiry-based, relevant and successful searches, or that this development should be speeded up in some way by the teacher.

Faced with multiple screenfuls of concordance lines and not being able to get helpful information from it could be very frustrating. To make good use of a concordancer, it is essential to develop effective data analysis strategies. The model of typology put forward in this paper could be a basis for teachers to be involved. It is believed that if students are led to know about these strategies in advance, they may be able to use some of them successfully. One way to do this is by designing activities based on concordance printouts.

Pre-writing use of the corpus is believed to be helpful for student writers to get a global view of the corpus, and hence get some ideas about the kind of language appropriate to the genre before starting to write. Though the present study does not prove that this will lead to better writing, it at least serves to propose that students when left alone will not develop pre-writing strategies to use the concordancer and must be instructed, guided or advised to do so.

Viewing the wider context of a concordance line is a linguistic research strategy which should be encouraged. More context is thrown up with this operation and more information is available for drawing conclusions. The inherent shortcoming of decontextualization arising from the truncated concordance lines, on which most paper-based exercises are developed, is also overcome.

Developing awareness of the potential and limitations of a corpus

It can be seen from this study that if students are to reap more benefits from concordancing activities in an ESP classroom setting, they should be made aware of the potential and limitations of the corpus. Teachers should inform their students about how the corpus was created and discuss with them what potential uses there are in relation to their classroom task and what limitations there are to their learning. It is expected that this would reduce the likelihood of students making too many irrelevant searches, taking up time and causing unnecessary frustration.

Future research directions

The present study was carried out in a very specific second language learning setting. The writing tasks involved centered around a very specific genre of language. Whether similar things will happen to settings with other ESP orientations remains to be explored.

The feasibility of self-access learning with concordancing has been claimed by some CC enthusiasts (e.g. Johns, 1988) but the results of this study suggest that self-access learning should be deferred to a much later stage when students have demonstrated mastery of some of the essential learning strategies, or should be run with students at a greater proficiency level. The stage of development at which self-access concordancing would best be introduced is certainly a good area for further exploration. Again, whether students of a greater proficiency level will find it easier to develop effective learning strategies in the use of the concordancer is another question which is worth serious consideration.

In relation to the development of learner strategies, it would be interesting to see how the teacher could play a better role. The marking of errors tried out in this study has proved to play a constructive role. What other more specific teaching strategies the teacher could develop in order to cultivate effective learner strategies remains to be further explored.

While it is very difficult to see in this study whether or not students more mature in the use of learning strategies are better learners or better writers, it is certainly worth exploring whether this is true. It could be valuable if comparisons of the quality of students' writing could be made between groups of students working with and without concordancing.

It is the sincere hope of the researcher of this study that future research will address some of these areas mentioned and that classroom concordancing, being a useful way of assisting students with their writing, will one day be carried out effectively in ESP writing classes around the world.

References

- Corder, S. P. (1967) 'On the significance of learners' errors.' *International Review of Applied Linguistics in Language Learning* 5: 161-170.
- Chamot, A.U. (1987) 'The learning strategies of ESL students.' In Wenden, A. & Rubin, J. (Eds.) *Learner strategies in language learning* (pp.71-83). Englewood Cliffs, NJ: Prentice Hall.
- Chamot, A.U., & J.M. O'Malley. (1990) *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Flowerdew, J. (1993) 'Variation across speech and writing in Biology.' Public lecture given at City Polytechnic of Hong Kong on 11 January, 1993.
- Higgins, J. & T. Johns. (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
- 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.' *English Language Research Journal (New Series)* 4: 1-16.
- Johns, T. (1991b) 'From printout to handout: grammar and vocabulary teaching in the context of data-driven learning.' *English Language Research Journal (New Series)* 4: 27-46.
- Levy, M. (1990) 'Concordances and their integration into a word-processing environment for language learners.' *System* 18(2): 177-188.
- Ma, K.C. (1993) 'Small-corpora concordancing in ESL teaching and learning.' *Hong Kong Papers in Linguistics and Language Teaching* 16: 11-30.
- Oxford, R. (1990) *Language learning strategies: What every teacher should know*. New York: Newbury House.
- Oxford, R. (1992) 'Language learning strategies in a nutshell: update and ESL suggestions.' *TESOL Journal*. 2(2):18-22.
- Rubin, J. (1987) 'Learner strategies: Theoretical assumptions, research history and typology.' In Wenden, A. & Rubin, J. (Eds.) *Learner strategies in language learning* (pp. 15-30). Englewood Cliffs, NJ: Prentice Hall.
- 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?' *Computer-Assisted English Language Journal* 1(2): 5-8.
- Stevens, V. (1991a) 'Concordance-based vocabulary exercises: a viable alternative to gap-fillers.' *English Language Research 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.
- Tribble, C. (1989) *Longman Mini-Concordancer Handbook*. Harlow: Longman.
- Tribble, C. & G. Jones. (1990) *Concordancing in the Classroom*. Harlow: Longman.
- Tribble, C. (1990) 'Concordancing and an EAP writing programme.' *Computer-Assisted English Language Journal* 1(2): 10-15.
- Wenden, A. & J. Rubin. (Eds.) (1987) *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice Hall.

