Enhancing the Language Awareness of Hong Kong Teachers through Corpus Data: The TeleNex Experience

QUENTIN G. ALLAN
TELEC, Department of Curriculum Studies
The University of Hong Kong
Pokfulam Road, Hong Kong
qgallon@hkucc.hku.hk

Recent developments in corpus linguistics, in particular the relatively easy access to large corpora and the development of powerful concordancing tools have had an encouraging impact on ESP and EAP. After ten years, the development of classroom concordancing or data-driven learning has been well documented, yet there seems to have been relatively little written about applications for teacher education, in particular, the education of nonnative teachers of English. In this paper, I argue that the use of corpus data—and concordance lines in particular—has a unique and powerful role to play in raising the language awareness of English teachers in Hong Kong secondary schools. I outline the approach we have adopted at TELEC, (Teachers of English Language Education Center—which links some 300 English teachers in Hong Kong, through a dedicated computer network based in the Department of Curriculum Studies at Hong Kong University), and explore some of the implications arising from our decision to include corpus data both in on-line grammar files, and in our responses to queries from teachers about particular language points.

This paper argues that the use of corpus data—and concordance lines in particular—has a unique and powerful role to play in raising the language awareness of English teachers in Hong Kong secondary schools. As a
prelude to the discussion, I briefly introduce the concept of concordancing with reference to some of the more recent applications, including the use of concordances for language awareness purposes in teacher training to date. In order to place the TeleNex approach in context, it is first necessary to offer a reason to be for TELEC, and to outline the key functions of the TeleNex computer network. I then offer a rationale for incorporating concordance lines into messages and files, drawing on relevant examples as appropriate. The presentation of corpus data in an electronic environment poses a number of special problems, as does the processing of corpus data for language awareness purposes. These problems are examined, and our responses evaluated. The paper concludes by looking forward to future developments, in particular, the move towards helping teachers to initiate their own corpus searches.

WHAT IS CONCORDANCING? WHY USE CORPUS DATA?

A concordancer is a computer program which, among other things, allows a user to search a collection of texts (called a corpus) in order to locate all instances of a particular language item; the language item can be as small as a punctuation mark or morpheme, or may be a multiword phrase. The program displays the results of a search in the form of a list (or concordance) with every instance of the search item appearing in the center of the screen or page, surrounded by its immediate context. Corpus linguistics studies using computer-based corpora can provide insights into language which were simply not possible previously and some of the findings from these studies are challenging many of the assumptions linguists have held about different aspects of language, suggesting amongst other things, a redrawing of word-class boundaries (see for example Sinclair’s discussion as to whether of should be considered a preposition, 1991, p. 82; and Francis’ analysis of the behavior of what she terms “the troublesome little word worth,” 1994, p. 231).

Part of the spin-off from large corpus-based projects such as COBUILD is that the concordancing software packages which are available commercially nowadays are powerful, flexible, and above all, user-friendly. An obvious example is Tim Johns’ MicroConcord published by OUP, which we use at TELEC. This package is powerful by virtue of the amount of text which can be processed (up to 10 million words); flexible in the sense that corpora can be added, adapted and reorganized according to the needs of the user; and its user-friendliness derives from the fact that it was
designed by a language teacher, for use by language teachers and language
learners. One of the most useful applications of a concordancer is as a lan-
guage learning tool in what Johns refers to as a “data-driven learning” ap-
proach whereby the student “assumes the role of researcher” (Johns, 1986,
p. 160). Data-driven learning as developed by Johns (1986, 1988, 1991) has
had a significant impact in English for special purposes and English for ac-
ademic purposes; however, despite the tremendous interest of recent years
in promoting language awareness in teachers, (see for example, James &
Garrett, 1991; Wright & Bolitho, 1993), there seems to have been relatively
little written about the use of concordances for teacher training, in particu-
lar, the continuing education of nonnative teachers of English. Notable excep-

Using Concordances for Teacher Training: The Story So Far

Coniam’s paper describes a study undertaken with trainee teachers of
English in Hong Kong; a central theme is the importance of challenging
teachers’ notions about some of the ‘facts’ they have long accepted about
the English language: facts which he suggests need to be very carefully re-
considered. Berry (1994) presents a reasoned discussion of the pros and
cons of using concordance printouts in language awareness training in
teacher education in general, (with examples) and argues that a successful
language awareness program should “equip teachers with the strategies
needed to acquire the knowledge on their own, and thereby help them to be-
come their own grammarians” (p. 195). Francis puts forward a convincing
argument for the use of large corpora in promoting grammatical awareness
in both language teachers and their students. She addresses the issue of met-
alanguage in such an approach, suggesting that in many cases it makes
more sense to focus on patterns rather than structures; she also raises some
interesting questions about traditional word-class boundaries. Hunston
(1995), adopting a similar approach to Berry, argues that “the ability to do
grammar should be seen as a set of skills rather than as a body of knowl-
dge” (p. 15), and proposes that a language awareness program should, in
contrast to traditional courses which teach a particular school of grammar,
focus on encouraging “observation, sensitivity to pattern, interpretation and
abstraction or generalization” (ibid.).

The thrust of the present paper is rather different: while I am in com-
plete sympathy with the rationale and objectives of the aforementioned
studies, the approach we have adopted at TELEC is necessarily different to
those outlined above, by virtue of our function as teacher educators on a computer network (*TeleNex*).

**OVERVIEW OF TELEC AND DESCRIPTION OF TELENEX**

TELEC is an acronym for Teachers of English Language Education Center, a teacher education facility based at Hong Kong University’s Department of Curriculum Studies. TELEC administers a local area computer network (*TeleNex*) linking some 300 secondary school English teachers, many of whom are not subject trained, and who, in addition to their own difficulties with fluency and accuracy, lack an explicit knowledge about the English language. Through a user-friendly front end, *TeleNex* provides teachers with access to information about English grammar in the form of files written in a hypertext format, with jumps leading to files on common misconceptions, Hong Kong students’ specific problems, and teaching implications. These files have been written especially for Hong Kong teachers and have links to graded teaching materials, which can be printed out ready for use in the classroom. In addition to the grammar and teaching ideas databases, a text bank allows teachers to prepare their own customized teaching materials using articles from the South China Morning Post, (one of the leading English language daily newspapers in Hong Kong); and a customized test generator enables teachers to prepare test items for school examinations. A key feature of *TeleNex* is a conferencing facility through which teachers are able to share their ideas, experiences, and problems about English language teaching with other teachers and TELEC staff members. Teachers are also able to send in any questions they have about the English language, and it is largely through the forum of the Grammar Corner that we are able to use concordances to clarify misconceptions, and to raise teachers’ language awareness through illustration of points of grammar and usage.

There are two main aspects to our use of the concordancer at TELEC:

1. The most common use to date has been in responding to teachers’ questions about grammar and usage in the grammar conferencing corner. The concordancer is invaluable as a source of useful information about the relative frequency of words, allowing us to show teachers how certain words and structures are typically used in a range of contexts. Where appropriate, we use concordance lines to illustrate explanations and in this respect, the concordancer furnishes a unique resource for enhancing teachers’ language awareness, complementing traditional resources such as dictionaries and grammar reference books.
2. When preparing grammar files, it provides us with a never-ending source of illustrative examples of native speaker usage; it also allows us to conduct research into areas in which students have recurrent difficulty.

In the main section of the paper, I shall elaborate on these two aspects, with reference to examples from TeleNex as appropriate. However, at this juncture it may be appropriate to first briefly outline our rationale for using corpus derived data.

**Rationale for Incorporating Concordance Lines into Messages and Files**

Language teaching in Hong Kong and elsewhere has traditionally distinguished between grammar and vocabulary: grammar reference books concentrate on syntactic patterning whereas dictionaries are preoccupied with vocabulary and issues of choice and substitution. Concordancing offers a synthesis, providing a unique view of language which can usefully be thought of in terms of two axes.

Starting with the vertical axis, suggested by the striking visual effect of the search word repeated down the center of the concordance, the focus is on the search word or structure. The keyword-in-context (KWIC) format places the keyword or search string in the center of the page or screen (see for example Figure 1) and so iconically, individual words or ‘chunks’ become the focal point and the point of departure for the analyst. On the horizontal axis we are able to focus on the relationship between words in sequence, and it is here that the study of concordance lines promotes an awareness of the lexico-grammatical patterning, (or *colligation* as it is sometimes called); the various patterns begin to emerge quite strikingly when we start to examine the typical structures associated with the keyword in its various parts of speech, or different inflexions.

Anyone using a concordancer becomes increasingly attuned to the differing syntactic patterns associated with, for example, the different tenses of a verb. What is even more striking for many users is the realization that making even a minor change to the keyword or search string leads not only to a change in the lexico-grammatical patterning, but can lead to quite marked changes in the semantic prosody, (the positive or negative connotations associated with a particular form or structure). For language teachers in Hong Kong, this awareness can have a positive effect on the way vocabulary is dealt with in their classrooms.
Using Concordances for Teacher Development: The TeleNex Approach

TELEC staff are able to use corpus data to complement information obtainable from traditional sources, and frequently make use of concordance lines in order to illustrate explanations about grammar and usage. For example, a teacher sent in a message to the Grammar conferencing corner asking if anyone could explain the difference between the adjectives content and contented.

In preparing a reply to this message, I looked first in a 5 million word native speaker collection, (including texts from Cobuild’s Bank of English, MicroConcord and the South China Morning Post), and found only one instance of contented as an adjective, compared with 196 examples of content—a ratio which I found rather surprising. Referring to a number of dictionaries, I found general agreement that contented means very happy, whereas content means willing to do something, or reasonably happy to accept something. However, something which the dictionaries didn’t stress, but which I found quite striking in the corpus is the negative semantic prosody associated with content. In a large number of cases, the immediate context suggests a clear idea of compromise, or settling for second best—in half of the citations in Figure 1, there is some form of modality and many of the citations contain either a negative, (as in # 3,4,6,10) or some suggestion of negative evaluation, (as in # 1,2,5,7, and possibly 9 and 11 too.)

```
1. including ourselves were just content to jog along the course. We we
2. Are you as fit, healthy, and content as you could be? Most people no
3. approved when Cousin May, not content with painting her own face, set
4. ced by Mell Brooks who - not content with that - plays many parts, o
5. m in your class you would be content to wrestle and scrape along in
6. of Case Law; we should not be content with a code of the brief and ab
7. t to Tessa, that you may feel content with that view, but this...we
8. u have heavy lorries that are content to stay at the heavy lorry-typ
9. s, that for your share you be content with 100 gold pieces”? The cohe
10. mmitments. <p> Attlee was not content simply to urge the need for eco
11. ss the people of Hongkong are content with the final outcome, it is u
```

**Figure 1. Concordance of “content” (TeleNex Corpus)**

In response to my first message, a teacher sent in a response pointing out that another way of distinguishing between content and contented is that while both are commonly used predicatively, the latter can also be used attributively while content is seldom used in this way. As our 5 million word corpus had only one instance of contented as an adjective, I checked in Cobuild’s 20 million word corpus from the Bank of English, and was able to illustrate the tendency as follows in Figure 2:
1. < sully, the ones remaining quietly contented as they happily see
2. looks silly but she had been totally contented with her music lesson
3. but whereas Emanuelle gave her a contented look, Vasco Miguel r
4. upon Mavis replied with a warm and contented smile, he went back
5. nger gave rise to whole platoons of contented soldiers. Songs, laughter
6. h unless this is aimed at producing contented and productive adults,
7. led tenderly at each other. Are you contented my darling? he asked
8. < peace and social stability. Here, contented peasants harvest their

Figure 2. Concordance of “contented” (Cobuild Direct)

I was also able to highlight the difference in semantic prosody between content and contented, illustrating in Figure 3 how many of the words which occur in the environment of contented have positive connotations:

1. < sully, the ones remaining quietly contented as they happily see
2. looks silly but she had been totally contented with her music lesson
3. but whereas Emanuelle gave her a contented look, Vasco Miguel r
4. upon Mavis replied with a warm and contented smile, he went back
5. nger gave rise to whole platoons of contented soldiers. Songs, laughter
6. h unless this is aimed at producing contented and productive adults,
7. led tenderly at each other. Are you contented my darling? he asked
8. < peace and social stability. Here, contented peasants harvest their

Figure 3. Concordance of “contented” (Cobuild Direct)

Part of our motivation for using concordance lines in response to queries in the conferencing corners is to enhance teachers’ understanding of and mastery of English. One of the main advantages of using concordances in this way is that the artificial concentration of contexts in the KWIC format represents a distillation of usage from a diverse range of registers and genres, which to some extent replicates the underlying intertextuality with which the keyword is imbued. This instinctive awareness of the various connotations and semantic prosody of a word is available to proficient users of the language, (albeit subliminally), but such an awareness is built up gradually, over time and through exposure to a wide range of language situations. Because of their relatively limited exposure to English, this “feeling” for the language may be only partially developed in learners. The majority of TeleNex teachers are not native speakers of English, and part of the rationale for exposing them to edited concordance lines is to help them focus on the salient patterns which emerge from a study of a concordance and which, according to Hunston (1995), “make explicit the intertextual information which the examination of a single text can call on only implicitly” (p. 17).
In response to a teacher’s question as to the difference between *day by day* and *day after day*, I first asked a number of colleagues, then checked these two phrases in Cobuild’s 20 million word corpus. In keeping with our intuitions, I found that *day after day* is normally used to convey some kind of negative feeling, possibly of frustration with the monotony, whereas *day by day* is typically used either for neutral situations, or to convey a sense of something worthwhile, developing incrementally in a steady, positive manner. Teachers were given the following citations (Figure 4 & Figure 5) to illustrate this tendency:

**Figure 4.** Concordance of “day after day” associated with some kind of negative feeling (*Cobuild Direct*)

| 1. feel his father’s death, spending day after day in silence, self- |
| 2. hungry and wretched. I have stood day after day watching the wago |
| 3. mitigation of soup. Thus we starved day after day and night after |
| 4. ng to windward in 60 knots of wind day after day after day.” <LTH> |
| 5. our people, in the poor women who day after day discover suffering |
| 6. that I couldn’t face meeting him day after day, and partly that I’d |
| 7. ould want them to do? Sit at home day after day contemplating suicide |
| 8. more than I can stand # the horror, day after day at the court and in |
| 9. anically repeats the words ‘Muslim’ day after day. <LTH> My friends, b |
| 10. r after year .. the same old thing day after day. And the impulse |

**Figure 5.** Concordance of “day by day” to convey a sense of something positive developing (*Cobuild Direct*)

| 1. competence is being increased day by day. <CQ1> Amleto |
| 2. having mastered the art of living day by day, we can always m |
| 3. smoking, your health is improving day by day. so why not take |
| 4. a result her memory is improving day by day. In particular, s |
| 5. They discovered, bit by bit, day by day, what algebra was |

Corpus data can also provide useful information about the relative frequency and distribution of words in different contexts; for example, one of the TELEC staff, Mr. Kam-yin Wu (Message in Language Corner 2, *TeleNex*, TELEC, April 23, 1996), writing in response to a teacher’s question about *moreover* and *furthermore* checked the use of each word in both a 2.1 million corpus of academic writing and in a 1.2 million corpus of spoken English before replying with the following observation and figures in support (Table 1): “Contrary to what you heard in the workshop, both *moreover* and *furthermore* are used more frequently in academic writing than in speech”. 
Table 1
Frequency of “moreover” and “furthermore” (*TeleNex Corpus*)

<table>
<thead>
<tr>
<th></th>
<th>Frequency of occurrence in the corpus of academic writing</th>
<th>Frequency of occurrence in the corpus of spoken English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moreover</td>
<td>141</td>
<td>8</td>
</tr>
<tr>
<td>furthermore</td>
<td>62</td>
<td>10</td>
</tr>
</tbody>
</table>

Part of the function of *TeleNex* is to clarify common misconceptions, and such information may help to make teachers more critical of the teaching materials they are using at present. A growing awareness of how different forms are distributed according to text types will hopefully help teachers to make informed decisions when deciding what to teach, and how much emphasis to give different forms, for example anyone familiar with Hong Kong students’ English compositions will be familiar with the preponderance of connectives such as *besides*, typically used to mean *in addition*. It seems reasonable to assume that part of the reason for this overuse is due to teachers’ introducing connectives to students in the form of a list along with other favorites such as the aforementioned *moreover* and *furthermore*. This sort of indiscriminate teaching stems largely from teachers’ misconceptions about usage, as hinted at in the following query from a teacher (Message in Language Corner 1, *TeleNex*, TELEC, October 30, 1995): “Is it true that *besides* is informal and thus should not be used in formal writing (e.g. a thesis)?”

Searches in a number of different corpora (formal and informal, written and spoken) reveal that *besides* does occur in formal writing as well as in everyday, spoken English, but that its overall use in English is really quite rare; and when it *is* used, its syntactic function is as likely to be intrasentential as intersentential.

In the TELEC student corpus, in only 313,000 words of student writing, *besides* appeared 146 times. Contrasting this with the native speaker corpus we find only 116 instances of *besides* in 5 million words of text; one does not need to be a statistician to conclude that perhaps our students are overusing this word.

In the native speaker corpus, *besides* appears at the beginning, in the middle, and even at the end of sentences; only half of all occurrences were sentence initial. Yet in the *TeleNex* student corpus we find that 90% of all instances were sentence initial; Figure 6 shows a sample from the student corpus:
1. Besides this, it makes us getting more lazy because the robot
2. besides the dog, the fat cat also see the snake, but the cat se
3. Besides, he is very rude to passengers. Last Friday, I took MT
4. Besides, I can serve them very well too. The main reason I prefe
5. Besides, it has a projector on its top. It has a radar on its head
6. Besides, I have the experience for looking after children for year
7. Besides, they have wrong concept. They think that they will becom
8. Besides, I carried my mother also. Unfortunately, she went down t
9. Besides, I put my mother on the table. She was only 20 cm tall.
10. Besides, all this, there is no more happiness for our wealth. It i
11. Besides, many student have mot a lot of money to spend on their le
12. Besides, Hong Kong is a democratic place. Everyone can enjoy righ

Figure 6. Concordance of “besides” (TeleNex Students Corpus)

Contrast this with the citation lines from our 5 million word native speaker collection in Figure 7. Notice that in the few cases where Besides is sentence initial, the letter <p> indicates that it is commonly used to begin a new paragraph.

1. olks’ home in San Francisco. Besides charity work Mui, who is prese
2. are not set by experience. <p> Besides being a theory about the basis
3. rty, proved burdensome. {para} Besides these practical political diff
4. h the total amount needed. <p> Besides manipulation of fertilizer app
5. and Mr Tony Galsworthy. {para} Besides briefly going over internation
6. </i>, produced in 414 BC. <p> Besides historiography and mathematic
7. the road, he explained. {para} Besides, he added, more than 17,000 ve
8. uld support the Conservatives. Besides, a National Government might w
9. add Coward to that mythology, besides which he was, judging by the b
10. ther ways of building strength besides lifting weights. Some sports m
11. dopt, the use of other tools” besides trade sanctions to promote Ame
12. o do other things with our day besides eat. <p> Fairly basic foods li
13. It is all quite innocent, and besides, it’s in the cause of her art.
14. ical authorities, do much else besides. They can declare that a certa
15. f the manifest and a good deal besides? I cannot really believe that

Figure 7. Concordance of “besides” (TeleNex Corpus)

In my response to this teacher’s query I concluded by highlighting three key points:

1. Besides is not very common in English. (Students should be encour-
egaged to use in addition, another thing, also, and, etc.)
2. When it does occur, half the time it will be used in the middle of the sentence, and not only at the beginning. (Corpus-derived data in the form of longer citations could be used to illustrate the other syntactic functions.)
3. When it *does* occur at the beginning of a sentence, it is only very rarely used with a comma after it; (the discourse function of this particular use is to draw special attention to an additional point the speaker or writer is trying to make.)

When preparing files for the grammar database, the native speaker corpus is a fertile source of illustrative examples reflecting the immediacy and freshness of real language, in contrast to the rather stilted and unnatural invented examples found in many ELT materials and reference books. In an earlier discussion of the grammar database, Wu and Tsui (in press) note that “grammar will be seen to be relevant to communicative language teaching only if it goes beyond the sentence level and is considered in relation to texts, spoken or written”. One of the key guiding principles of the database design is that extracts from naturally occurring text *beyond the sentence level* are used as examples wherever appropriate in order to illustrate the way language functions in specific texts. One of the advantages of presenting these examples in a hypertext format is that short examples can be used, with access to the wider context available through a small “back to context” icon ( ). See Wu and Tsui (in press) for a full discussion.

We also use concordance lines in KWIC format in grammar database files, either to illustrate a grammatical point or to focus on a particular aspect of usage. The lines can be accessed by clicking on a pop-up icon (), for example, in the following excerpt from a set of files on nominalization (Figure 8), the lines are provided to demonstrate the different forms of the verb *to go* with *-ing* nouns referring to sports, hobbies and so forth.

![Sample screen from TeleNex grammar file: Forms of Nominalization](image)

*Figure 8.* Sample screen from *TeleNex* grammar file: Forms of Nominalization
“In addition to such examples drawn from native speaker usage, we also select representative examples from the student writing corpus to illustrate specific students’ problems, as in this section dealing with students’ avoidance of certain nominalizations (Figure 9). Teachers have the option to click on the “further information” icon to see a possible rewording.

\[50\%\]_Sample screen from TeleNex grammar file: Students’ Problems: Avoiding Nominalization

**PRESENTATION OF CORPUS DATA: SOME OBSERVATIONS**

The way we present corpus data in files and messages has been refined with recourse to the available literature, experimentation and through the feedback we have received from TeleNex teachers. Regarding layout, teachers have indicated that due to the limited time they have available for reading files and messages, in general they prefer shorter, edited concordances of 10 to 15 citations at a time, with the lines numbered and main points explicated where appropriate. One of the advantages of presenting citations in a hypertext format is the possibility of building in pop-ups containing either wider context, or some relevant commentary. It is our policy at TELEC to limit the use of metalinguistic terminology as far as possible, and beyond the short list of new terms necessary for discussing concordances, (e.g. corpus, concordance, citation, collocation) we have managed to use only terms which teachers are already likely to be familiar with.
The transmission of information on a computer screen poses special problems of legibility, and early research into this aspect of computer ergonomics warns against inappropriate use of color, (see for example, Bouma, 1980; England, 1984, 1987). Therefore, in order to find the most eye-friendly format for presentation of concordance lines, we have experimented with a range of different colors—bearing in mind that due to differences in luminosity and saturation, some electronic colors are easier to read than others. For concordance lines we have settled on dark blue, and to highlight significant patterns, we use dark red in bold. Regarding our choice of font, we are constrained by the need to use a fixed-width font such as Courier to support the KWIC format; we find that 10 point is large enough to read easily, and still allow a reasonable amount of context on either side of the keyword.

**PROCESSING CORPUS DATA FOR LANGUAGE AWARENESS PURPOSES: SOME CONSIDERATIONS**

The application of concordance lines as a tool in consciousness-raising is by no means uncontroversial, see for example McNeill (1993), and in exploiting concordances for language awareness purposes in teacher education, it is important to be aware of certain dangers inherent in the process of obtaining, editing, and presenting corpus data for use with teachers.

The first problem relates to time: searching on-line in a large corpus can throw up hundreds of citations which take time to analyze and edit, especially if unexpected patterns emerge. Secondly, the artificial nature of the KWIC format means that the analyst is at one remove from the writer’s original communicative purpose, and while most concordancers have a feature which enables the analyst to view the wider context, this is only possible on-line, and referring to the wider context of every line becomes incredibly time-consuming when dealing with concordances of any size.

In order to reduce concordances to a manageable size for presentation to teachers, editing is necessary, which, if done judiciously, can reflect proportionally the distribution of patterns of use within a corpus, (see for example, Figure 7); however, Berry (1994) warns against a danger that particular citations can be selected, and used to “cook the books” (p. 206).

There may also be a temptation to overgeneralize from insufficient data; a temptation which can be moderated by an awareness that a given concordance will only be a reflection of the corpus from which it was generated, and understanding that with any corpus, there may be a possibility of skewing due to overrepresentation of a particular genre or author.
It is worth noting too, that analysis of corpus data can lead to generalizations which sometimes contradict the traditional wisdom of teachers and coursebooks. Such contradictions are potentially destabilizing for teachers. For example, native speaker intuition, backed up by corpus evidence, suggests that beginning a sentence with a subordinating conjunction such as *Because* is common and perfectly acceptable, yet discussions in the conferencing corner reveal that Hong Kong students are penalized for beginning a sentence with *Because*, and are typically instructed to use the restrictive formula *It is because*. It is to be hoped that through continued exposure to corpus data, teachers will develop an open-minded, inquiring approach to help them deal with findings which run counter to the traditional wisdom of dictionaries, grammars and coursebooks. Sinclair (1994, p. 13) stresses the importance of being

open to the patterns observable in language in quantity as we now have it. The growing evidence that we have suggests that there is to be found a wealth of meaningful patterns that, with current perspectives, we are not led to expect. We must gratefully adjust to this new situation and rebuild a picture of language and meaning which is not only consistent with the evidence but exploits it to the full.

**FUTURE DEVELOPMENTS: A MOVE TOWARDS TEACHER-INITIATED CORPUS SEARCHES**

From unsolicited feedback in the grammar conferencing corner, teachers’ reactions to our use of concordance lines in grammar files and in messages have been very positive. We recently conducted a small-scale survey of teachers’ attitudes to language awareness through concordancing, and the 13 teachers who returned the questionnaire all indicated they found concordance lines useful in illustrating explanations. Even more encouragingly, since we have begun using concordances in response to queries on grammar and usage, a growing number of teachers have expressed interest in being able to use a concordancer themselves, and TELEC is currently looking into the possibility of developing a concordancing package to enable all TeleNex teachers to access both the native speaker corpus and the student writing corpus. Such a development would be of benefit to teachers for a number of reasons.

First, with respect to their own language ability, teachers may wish to focus on developing their understanding of English, which in turn would
hopefully enhance their ability to *use* the language, and as I hope I have indicated in the discussion above, access to a large corpus of native-speaker English through a concordancer has a unique potential to raise teachers’ awareness of patterning and usage within specific genres and texts, as well as across the language as a whole. Berry (1994, p. 107) notes that through the process of analyzing data in the form of concordance lines and frequency lists, teachers can develop confidence in their intuitions, as well as discover their limitations. Furthermore, the systematic analysis of concordance lines lends itself to an inductive approach and provides a focus for the development of analytical skills, with the *teacher* as researcher following Johns (1988), and in keeping with the current trend towards teacher-initiated action research. The data driven learning approach is, moreover, inherently motivating: teachers can conduct tailor-made searches relevant to their own interests and preoccupations. By using a discovery-learning approach, the teacher comes to ‘own’ the knowledge gained; and, as the findings from any given search will invariably suggest other avenues for exploration, the process of investigation and analysis will be never-ending.

For teaching purposes, corpus-based investigations can guide teachers in their decision as to how much time should be devoted to a given lexical item or grammatical structure. Some teachers, especially those with responsibility for more advanced students, may wish to use the concordancer for the preparation of exercises which would be given to students as handouts in accordance with Johns’ data-driven learning approach, referred to previously. Another major use of the concordancer would be to conduct research into students’ writing, to investigate patterns of usage, particularly recurrent errors, and avoidance strategies—as well as identifying what students *do* get right. And for language awareness activities in the classroom, students could compare concordances from the student corpus with concordances of the same item taken from a suitably matched native speaker corpus (Pickard, 1994).

**SELF-ACCESS CONCORDANCING: SOME CONSIDERATIONS**

In evaluating a self-access concordancing package for undergraduates in Hong Kong University’s language center, Pickard (1994) points to the need for “more exercises which guide students in the techniques of language analysis or metalinguistic awareness” (p. 221). Analysis is a skill which takes time to develop and my own experiences in introducing concordance lines to undergraduates in an ESP context suggest the need for patience, and systematic training (Allan, 1995). However, it is not only
in analyzing concordance lines that teachers will need guidance: referring to Ma’s (1993) ethnographic study into students’ use of a concordancer as an on-line resource on a writing project, Flowerdew (1993) observes that “learners in the study were severely limited in the use of concordancing...because they were unable to make use of many of the possible search techniques available to users of a concordancer” (p. 99). He follows this up by concluding that “if concordancing is to achieve optimal results, *adequate time must be devoted to familiarizing learners* [my italics] with the sort of things they can find out by using the concordancer.”

Therefore, in making available a concordancing package for *TeleNex* teachers, we intend to give a high priority to training. To this end, the package will incorporate a tutorial function which, through a series of help-files in hypertext format will provide a structured induction into the possibilities of concordancing—and guidance through the potential pitfalls. It will also be necessary to complement this tutorial-cum-help facility with familiarization workshops as new teachers join the network.

Such training will be vital, as it is important to bear in mind that concordances in KWIC format appear quite alien for many people; for most *TeleNex* teachers, seeing concordance lines in the grammar files, and conferencing corners, has been a new experience. Not the least of the problems reported by some of these newcomers is the initial shock of arbitrarily truncated lines, coupled with the limited context which can be a source of frustration for users who expect to but do not necessarily understand every line.

Another problem area is patterning: we cannot assume that patterns will be seen easily by all teachers. Hunston (1995) claims that “the visual patterning of the concordance lines facilitates ‘noticing’ of lexical/grammatical patterns” (p. 18), and while this may be true for an experienced lexicographer, linguist, or language teacher, our preliminary research indicates that such patterning is not always apparent to teachers who are not subject-trained, partly due to lack of familiarity with clause structure, and partly to the inappropriate reading strategies adopted in dealing with concordances. For example, some teachers report reading lines word by word, from left to right as with running text rather than scanning down the page for patterns.

It is to be hoped though, that when the TELEC concordancing package is available, *TeleNex* teachers will have benefited from the experience of having been exposed to an extensive range of concordances, the majority accompanied by more or less detailed explication, with significant patterns and collocations highlighted. Through familiarization with the sorts of searches which they have seen conducted by TELEC staff, it is envisaged that they will have a head start when it comes to running searches of their own, and analyzing the resulting concordances.
References


Wright, T., & Bolitho, R. (1993). Language awareness: a missing link in
Wu, K.Y., & Tsui, A.B.M. Teachers’ grammar on the electronic highway:

**Notes**

Quentin G. Allan is a teaching consultant at TELEC (Teachers of English
Language Education Center), The University of Hong Kong.

This article is a revised version of a paper originally presented at a confer-
ence jointly hosted by the Language Center, Hong Kong University of Sci-
ence and Technology and the Department of English, Lingnan College,
Hong Kong in June, 1996 on the theme: *Language Analysis and Descrip-
tion: Applications in Language Teaching*.

I am extremely grateful to all my colleagues at TELEC for providing such a
congenial work atmosphere, and would like to thank in particular Anne
Gordon for her assistance with data collection, and her cheerful support at
all stages of the enterprise; and Kam-yin Wu for his invaluable comments
on an earlier draft.

Correspondence regarding this article should be addressed to Quentin G.
Allan, TELEC, Department of Curriculum Studies, The University of Hong
Kong. (Fax: (852)-2517 2100; email: qgallan@hkucc.hku.hk).

In March, 1997 *TeleNex* was formally launched on the internet. Parts of the
databases and sample messages from the conferencing corners are avail-
able and these will be added to progressively. The projected completion
date for the database files is August, 1998.

*TeleNex* web site: http://www.telenex.hku.hk or http://www.netvigator.com
(under Study Cell)