CONCORDANCING AS A TOOL IN COURSE DESIGN*

JOHN FLOWERDEW

City Polytechnic, Hong Kong

Computerized text analysis programs (concordancers) are now available for use on personal computers. Drawing upon experimental work done at Sultan Qaboos University, Sultanate of Oman, this paper shows how such programs can be used as a tool in course design. The starting point is a corpus of written and/or spoken text from the target communicative situation. From this data-base computer text processing can provide criteria for: (a) the selection and grading of items for the syllabus, and (b) the authentic contextualization of these items in learning materials.

INTRODUCTION

Computerized text analysis, or concordancing, programs (Sinclair, 1986, 1991) are now becoming widely available for use on personal computers. A concordancer usually consists of two programs: a word frequency program, to provide data on the number of instances of all words in a corpus of text, and the concordancing program proper, to find all instances of a given word in a corpus and to present these instances in their immediate linguistic context (the Appendix shows what a concordance actually looks like). This paper will show how such programs can be applied to a specialized corpus of language and used in specific purpose course design.

The potential for computer-assisted corpus analysis, or concordancing, in language teaching and learning has been the focus of attention on the part of teachers and researchers for some time now. The best documented work in this area is that of Professor John Sinclair and his colleagues at the University of Birmingham, under the auspices of the Collins Cobuild English Project (Sinclair, 1991). The Cobuild project employed a corpus running into some 20 million words and its application has been in the area of dictionaries (Sinclair, 1987), grammars (Collins/University of Birmingham, 1990) and main course ELT syllabuses (Sinclair and Renouf, 1988; Willis and Willis, 1988).

In parallel to this Cobuild work, concordancing techniques have also been applied in the classroom, both for materials production (Johns, 1989; Tribble, 1990) and for use as a learning tool by students themselves (Tribble, 1990; Stevens, 1991). For these direct classroom applications much smaller corpora have been used. This paper presents a further application of small corpus concordancing, but this time in the field of ESP course design. The Cobuild lexical syllabus (Willis and Willis, 1988), the

*This is a revised version of a paper originally presented at the 24th Annual TESOL Convention. San Francisco, March 1990. I should like to acknowledge the helpful comments of Lynne Flowerdew and Martha Pennington on an earlier draft.
only publicly available application of concordancing to course design developed so far, was based on the full Cobuild corpus, as it was a "general" English syllabus and therefore needed to sample a wide cross-section of registers and genres. It is the contention here, however, that where a course is designed for a particular specific purpose a much smaller corpus of language, drawn from the given specific purpose area, is more appropriate. This paper will demonstrate the potential of this approach by means of a case study of a course developed in the Language Centre at Sultan Qaboos University (SQU), Sultanate of Oman.

BACKGROUND

In the foundation course for science students at the English-medium SQU, science and English are taught in parallel, the English course taking its material in large part from the science course. The function of the English course is two-fold: first, to help students cope on a day-to-day basis with the communicative demands which are put on them in their science classes, and, second, to foster the ongoing development of their communicative competence in English within their area of studies.

To fulfil these twin aims, a decision was made that the language presented in the English course should be developed around the language and communicative activities students are exposed to in science. By developing a corpus of language used in the science course, it was possible to provide a systematically selected and graded input to the communicative activities developed in the English course, which corresponded to the actual language students were exposed to in their science.

It may be felt that the corpus-derived language input approach just outlined suggests a "product-based" orientation, when a "process-based" orientation would be more in keeping with current thinking on syllabus/course design (Breen, 1987; White, 1988). However, it is emphasized that the procedure forms but one part of the course design procedure (it is a tool) and is not meant to produce the complete syllabus or course. The overall organizing principle of the English course at SQU, based around communicative activities associated with the study of science, was task-, and hence process-based. Tasks were used in a similar way to that proposed by Hutchinson and Waters (1987), with input from the concordancing introduced in what Hutchinson and Waters refer to as a "language focus" stage, leading up to the task. This procedure is illustrated in a later section of this paper, "Using the data from the analysis: an example".

CORPUS AND ANALYSIS

For the first 10 weeks of their studies students receive approximately 25 hr of lectures in biology, supported by very short readings. (The ratio of lecture text to written text was about 10:1. This was because of the limited amount of reading students were given and because of the higher lexical density of written text compared to spoken.) A corpus was created consisting of verbatim transcriptions of these lectures (one set, but given by different lecturers) and readings. The corpus, which consisted of 104,483 words, was thus made up of the actual language students were exposed to in their study of science. The corpus
was analyzed by means of a frequency-concordancing program developed at SQU by David Poulton, chief technician in the Language Centre. However, similar programs are available commercially (e.g. Longman mini-concordancer, Oxford Micro-OCP, TACT, 1.2.) The analysis of the data is presented under two main headings: word frequency and concordancing.

Word frequency

Total word count. The total of 104,483 words in the corpus breaks down into 4232 separate items. In order to be able to make valid comparisons with other word counts, it is necessary to explain how this figure is arrived at.

Computerized word counts are less discriminatory than other traditional types, such as those based on dictionaries, for example, and therefore the total number of items tends to be larger. In particular, computerized counts include all derived forms of a given word as separate items. Thus, where the dictionary includes "convince", "convincing", "convincingly", and "convincible" under one single item, or head-word, the computer counts these as separate items. Derived forms have been estimated to make up over 50\% of word entries in dictionaries (Goulden et al., 1990). In addition, computerized counts include singular and plural of nouns and inflected verb forms as separate items, whereas these are included under single head-words in dictionaries. These items make up some 10\% of the present corpus. On the other hand, one area where computerized counts are more parsimonious than dictionary-based counts is in the treatment of homographs, words with the same spelling, but different meanings. As far as the computer is concerned, words with the same spelling represent one item. In dictionaries, however, where a single word has two unrelated meanings, these count as two separate entries, or head-words. In a recently published word count, Goulden et al. (1990) estimate that homographs account for some 5\% of total items.

Taking the above distorting factors into account—i.e. derived forms (+50\%), plurals (+10\%) and homographs (-50\%)-in order to provide a valid comparison with other published word counts, the total number of items for the specialized corpus should be reduced by some 55\%, to a figure of less than 2000 words.

Such a small number of words is a very realistic target for teaching. To make a comparison with the vocabulary knowledge of the average native speaker, Goulden et al. (1990), whose estimate was on the consecutive side, compared with other studies, estimated the average educated native speaker to have a passive vocabulary of around 17,000 base words, 8\% times our total of less than 2000. In terms of a potential objective for teaching, therefore, our specialized corpus word count represents a considerable refinement vis-d-vis a general word list. Instead of having to select for course design, as would a general English course, from among the 17,000-odd words known by the average educated native speaker (and, of course, these 17,000 words would vary from person to person), the ESP course can limit its focus to the much smaller 2,000-odd words in the specialist corpus. Of course, all language courses are based on a limited vocabulary objective insofar as they do not include the complete 17,000 words known by the average native speaker. Renouf, for example, in an analysis of nine major EFL courses [cited in Sinclair and Renouf (1988)].
showed that in the first book of each series the number of different word forms varies from 1156 to 3963 (a number not far off our total of 4232 forms). A decision must therefore have been made to include only a fraction of the words known by the average native speaker. However, the great power of the corpus-based word list lies in the fact that the course designer can be sure that the words selected are the most useful (i.e. the most frequently used). Of course, there are other well-established criteria for vocabulary selection in addition to frequency, such as "disponibilite" (or "coverage"), "teachability" and "classroom needs" (Halliday et al., 1964; Carter, 1987). The course designer may want to modify application of the frequency according to such criteria. In addition, the question of the need for receptive vs productive and spoken vs written usage of vocabulary (which latter, incidentally, the concordancer can identify) needs to be addressed.

Frequency data as criteria for syllabus selection and grading. The previous section has shown the power of frequency data in helping to provide an overall total for the number of vocabulary items to be incorporated into a course. Another important application of the word frequency data in course design is in establishing the relative importance of vocabulary items and thereby providing criteria for syllabus selection and grading.

Table 1 shows the relative frequency of items in the corpus. There is an inverse relation between frequency of occurrence and number of items occurring with a given frequency. Thus, while there are 1417 items occurring only once, there are only 549 items occurring twice, 356 items occurring 3 times, 256 items occurring 4 times, etc.; until by the time we reach 10 occurrences there are only 72 items occurring this number of times.

Table 1. Relative frequency of items*

<table>
<thead>
<tr>
<th>Number of occurrences</th>
<th>Number of items occurring this many times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1417</td>
</tr>
<tr>
<td>2</td>
<td>549</td>
</tr>
<tr>
<td>3</td>
<td>356</td>
</tr>
<tr>
<td>4</td>
<td>256</td>
</tr>
<tr>
<td>5</td>
<td>184</td>
</tr>
<tr>
<td>6</td>
<td>142</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
</tr>
<tr>
<td>8</td>
<td>99</td>
</tr>
<tr>
<td>9</td>
<td>83</td>
</tr>
<tr>
<td>10</td>
<td>72</td>
</tr>
</tbody>
</table>

*Total words =104,483, total items = 4232.

These data offer a very powerful tool in the hands of the course designer, each level of frequency offering a potential cut-off point for selection and grading of items (the number of words targeted by the syllabus designer, of course, depending on the entry level of learners and time available for teaching/learning). Table 2 shows how progressive exclusion of the lower-frequency items radically reduces the number of target lexical items. By excluding single occurrence items, for example, the number of target items is immediately reduced from 4232 to 2815 (by over a third). By excluding items occurring 2 or less times, the
Table 2. Total items, progressively excluding low-frequency items

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total items</td>
<td>4232</td>
</tr>
<tr>
<td>Less single-occurrence items</td>
<td>2815</td>
</tr>
<tr>
<td>Less two or less occurrences</td>
<td>2266</td>
</tr>
<tr>
<td>Less three or less occurrences</td>
<td>1912</td>
</tr>
<tr>
<td>Less four or less occurrences</td>
<td>1654</td>
</tr>
<tr>
<td>Less five or less occurrences</td>
<td>1470</td>
</tr>
<tr>
<td>Less 10 or less occurrences</td>
<td>975</td>
</tr>
<tr>
<td>Less 20 or less occurrences</td>
<td>602</td>
</tr>
</tbody>
</table>

target is further reduced to 2266 (by nearly a half). By excluding items occurring 10 or less times, the target shrinks to 975 (less than a quarter of the total). By the time items occurring 20 times or less are excluded we are left with only 602 words (less than a seventh). As mentioned above, criteria other than frequency, of course, may also enter into the selection procedure.

*Frequency of specialist corpus vs frequency of general corpora.* Purely as far as word frequency is concerned (and ignoring for the moment the role of word frequency as a necessary condition for concordancing), in order for there to be particular value in creating a specialist corpus, it must be demonstrated that the specialist corpus has a different make up from a general corpus; otherwise an already available general frequency list could be used to the same end. I will now consider this question.

A large percentage of the most frequently occurring items of the specialist corpus are grammatical, or form, words, as opposed to lexical, or content, words. The 10 most frequent items in the specialist corpus, for example, are all grammatical words. In this respect it is true that the corpus is very much like any other, be it an ESP corpus or a "general" one. The Cobuild general corpus, for example, also has grammatical words as its 10 most frequent items (see Table 3).

Table 3. The 10 most frequent items in the specialist corpus and in Cobuild

<table>
<thead>
<tr>
<th>Item</th>
<th>Specialist corpus</th>
<th>Cobuild</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>8315 (8.7%10)</td>
<td>the</td>
</tr>
<tr>
<td>and</td>
<td>3099 (3.2%10)</td>
<td>of</td>
</tr>
<tr>
<td>of</td>
<td>3014 (3.1%)</td>
<td>and</td>
</tr>
<tr>
<td>is</td>
<td>2886 (3.0%)</td>
<td>to</td>
</tr>
<tr>
<td>a</td>
<td>2429 (2.5%)</td>
<td>a</td>
</tr>
<tr>
<td>in</td>
<td>2225 (2.3%)</td>
<td>in</td>
</tr>
<tr>
<td>are</td>
<td>1785 (1.9%)</td>
<td>that</td>
</tr>
<tr>
<td>to</td>
<td>1709 (1.8%)</td>
<td>l</td>
</tr>
<tr>
<td>it</td>
<td>1366 (1.4%)</td>
<td>it</td>
</tr>
<tr>
<td>this</td>
<td>1356 (1.4%)</td>
<td>was</td>
</tr>
</tbody>
</table>

Given this overall similarity, however, differences in the ordering of certain grammatical items are nevertheless considerable. In Cobuild, for example, "was" is the tenth most frequent item, whereas it is only fiftieth in the specialist corpus, indicating perhaps the greater use of past narrative in general usage than in descriptive science. Certainly, the lower frequency of the past form in the specialist corpus has possible implications for work.
on tenses in course design. To take another example, in Cobuild the word "so" is not a high-frequency item [it does not appear in the top 200 items printed in Sinclair and Renouf (1988)], but in the specialist corpus it is very frequent, occurring in thirteenth position. This might be accounted for by the greater use of cause-and-effect markers in academic discourse, a hypothesis which, if corroborated by the concordancer, needs to be taken into account in course design.

Turning now to lexical items, there is considerable variation between Cobuild and the specialist biology corpus. For example, none of the top 20 nouns in Cobuild occurs among the top 20 nouns of the specialist corpus (Table 4).

Table 4. Top 20 nouns in Cobuild and in biology

| Top 20 nouns in Cobuild: time, people, way, man, years, work, world, thing, day, children, life, men, fact, house, kind, year, place, home, sort, end. |
| Top 20 nouns in biology: cell, water, membrane, food plant, root, molecules, plants, wall, energy, concentration, organisms, cytoplasm, animal, stem, structure, body, part, animals. |

As further support for the use of the specialist corpus, it should be pointed out that, even where items are common to both general and the specialist corpus, the items in the specialist corpus may have particular uses, a possibility which will be corroborated when these items are concordanced (see the section below on concordancing).

Technical vs subtechnical vocabulary. As a further counter to the use of the specialist corpus, it might have been predicted that many of the lexical items particular to the specialist corpus would be technical words and would not need to be incorporated into an ESP syllabus, as they would be explained by the content teacher. However, this is not the case. With the nouns cited in Table 4, for example, "cytoplasm" is the only word clearly falling into the "technical" category (although a case might also be made for cell, membrane, and molecule). If the majority of these items are not technical words, neither are they "general", or common-core; rather they might be classed as semi- or subtechnical; that is to say they are words in general usage, but which have a special meaning within the technical area (Inman, 1978). Wall, energy, concentration, structure, body, and animal clearly fit into this category. Because these words are not likely to be glossed by the content teacher their subtechnical meanings are clearly the domain of the ESP teacher.

Items conspicuous by their low frequency or absence. A significant point about the specific purpose corpus vis-a-vis a general corpus is that some items which might have been predicted to be high-frequency are conspicuous by their low frequency or total absence. Given that the corpus is an academic one, it might have been predicted, for example, that the full range of logical connectors would be well represented. However, as Table 5 shows, whilst a small group of connectors are very frequently used - "so" (1183), "then" (266), "first" (103), "next" (72) - others are less common - "however" (13), "therefore" (11), "thus"
Table 5. Connectors

Important connectors:

so (1183), then (266), first (103), next (72)

Less important connectors:

however (13), therefore (11), thus (8), finally (8), as a result (4)

Connectors not appearing at all:

what is more, furthermore, nonetheless, nevertheless, hence, consequently, in conclusion, in contrast, after that

"finally" (8), "as a result" (4)—and others do not appear at all—"what is more", "furthermore", "nonetheless", "nevertheless", "hence", "consequently", "in conclusion", "in contrast", "after that". As well as telling course designers, therefore, which items to include in a syllabus, the frequency list also tells them which items not to include.

Frequency and functional/notional selection and grading. An important application of the frequency list for course design is in deriving intuitions about functional and notional areas which might be important for the syllabus. For example, frequent occurrences of "called" (414), "call" (57), "means" (172), and "mean" (31) identify defining, or naming, as an important function, worthy of inclusion in a syllabus. Other examples of important functional/notional areas, as identified by various lexical items used to realize them, are listed in Table 6. That these items are indeed used with these particular functional/notional uses has to be corroborated by concordancing (see below).

Table 6. Indicators of functional/notional areas

Naming or defining

called (414), call (57), means (172), mean (31) referring back
remember (155) Checking/topic changing
okay (252), right (194), now (561), well (175) Boosting and down-toning
just (263), very (490), quite (108), really (77) Spatial location
in (2238), on (481), around (161), between (159), across (87), below (15), beneath (7)

Frequency in the evaluation and revision of currently used syllabuses and materials. As well as its important role in the design of new courses, frequency data can also be extremely useful in the evaluation and revision of existing courses.

With the availability now of optical scanners, a complete set of materials can very easily be fed onto a word-processor and a frequency list created of the vocabulary items contained therein. If this frequency list is then compared with that of the authentic corpus, discrepancies can serve as a basis for evaluation and revision.

Concordancing

Whilst frequency data, as has been seen, tell which items to select for a syllabus, concordances tell how these items are actually used. This information is useful in three
main areas of course design: (1) for the syllabus, (2) for providing instances of use for direct incorporation into instructional materials, and (3) for the evaluation of current syllabuses and materials.

**Concordancing for syllabus design.** Concordancing has a number of applications for syllabus design. First, given that any item may have a number of different uses, concordancing can identify which uses of items to teach (and, by extension, which uses not to teach). The concordance of "well", for example (partly reproduced in the Appendix), allows us to identify three important uses of this item in the corpus (dictionaries contain many more)—as an adverb (as in "show up well"), as a conjunct (as in "as well as causing some diseases"), and, importantly for ESP students needing lecture comprehension skills, as a discourse marker (as in "well, I think this might be interesting"). Concordancing of "or" identifies two main uses of this item in the corpus (dictionaries, again, have many more)—"exclusive" "or" (as in "the thorax is pushed to one side or the other") and to introduce a synonym or paraphrase (as in "side roots or lateral roots", "produce new cells or grow"). Concordancing thus reduces those uses of a given item to be presented for learning to those actually occur in the corpus, eliminating time-consuming attention to other uses pointed to by dictionaries and reference grammars.

A second application of concordancing for syllabus design is in showing the syntactic patterns in which words occur. In this respect, incidentally, much published instructional material is shown up to be inauthentic. For example, concordancing shows that the connector "then" rarely occurs as sentence initial [as it is often taught [e.g. Bates and Dudley-Evans (1976)]], but is more usually found between subject and verb:

"the viruses then do the same"
"these goblet cells then secrete mucus" "the cells then expand"

-or between auxiliary and main verb:

"the liquid is then discharged" "classes which are then divided" "it can then put down more layers"

Similarly, spatial prepositions are most often used to introduce post-modifying phrases (reduced relative clauses), as in:

"the lysosome in the cytoplasm of most cells" "the membranes around the sap vacuole"
"the microtubules just below the plasma membrane"

Compare this with course books which prefer to teach prepositions in their role as adjuncts, by means of statements such as "The circle is at the top", "The square is under the circle" [e.g. Bates and Dudley-Evans, 1976].

A third application of concordancing for syllabus design has already been referred to in the above section on frequency. This is in corroborating intuitions derived from the
frequency list on important functional and notional areas which might be included in a syllabus (Table 6). Some of these items are fairly transparent in their function - e.g. "in", "on", "across" etc., as indicators of spatial location-and therefore do not really need corroboration as to their function. Others, however, are less transparent and do need corroboration by the concordancer. For example, the discourse markers used to check or indicate topic change - "okay", "right", "now", "well" - all have other possible functions; for example, "okay" could be used to agree to a request, "right" could be used to evaluate a student comment. It is only by looking at authentic instances of use, as provided by the concordancer, that the analyst/course designer can be sure of the true functions of items such as these.

Concordancing for providing instances of use for direct incorporation into instructional materials. Any of the examples cited as means of identifying areas for inclusion in the syllabus could also be incorporated directly into the teaching materials developed to promote the syllabus. High face validity is given to an ESP course if the learning materials contain actual examples of use which are drawn from the content area and which the learner is likely to have come across, or will be likely to come across, in his specific area studies. In teaching post modification, therefore, learners are more likely to see the relevance of phrases drawn from the corpus such as the following:

"the lysosome in the cytoplasm of most cells" "the membranes around the vacuole"
"the microtubules just below the plasma membrane" "animals with a small surface/volume ratio"

than they are the type of phrases more usually employed to teach this grammatical construction, such as, "the girl in the corner", "the little boy down the lane", "the man with the beard", etc. Similarly, in teaching discourse markers, the authentic instances of their use provided by the concordancer can be presented directly to learners.

"... okay, any questions?" "... okay, don't worry"
"... now, a good example of a glycoprotein is..." "... now, any cell that is very active..."
"... well, I think this might be interesting..." "... well, maybe if I think out the question..."

In addition to high face validity, the use of authentic examples in materials, of course, ensures that an accurate representation of actual use is presented to the learner. With fabricated examples, however closely modeled they may be on authentic instances of use, there is always a danger that learners will be presented with a distorted picture of actual use.

Concordancing for the evaluation of current syllabuses and materials. As well as being a powerful tool in the development of a syllabus and materials, as with raw frequency data, concordancing can also fulfil an important role in the evaluation of a course, once it has been developed. To illustrate how this type of evaluation might apply I will take
two areas where concordancing has revealed a discrepancy between published materials and the specialist corpus.

The first of these examples concerns the teaching of definitions. Many commercially available course books teach learners to express definitions by means of a formula such as "X is/can be defined as . . ." [e.g. Allen and Widdowson (1974) and Master (1986)]. In the corpus, however, there is only one instance of the lexical item "define"; on the other hand, there are 417 instances of the lexical item "called", used in a defining function. Clearly, the commercially available materials, in presenting an inauthentic formula, are doing a disservice to learners and need to be replaced by materials presenting a more accurate model of how definitions are expressed.

The second area concerns the tendency of published materials to present idealized patterns in their teaching of syntax. I have already referred to the teaching of connectors as sentence initial, while the corpus reveals that they are more often found between subject and verb. As another example, published materials have a tendency to teach the passive as a simple subject + auxiliary + past participle construction ("The work was finished", "The cake was eaten", etc.) (Ewer and Latorre, 1969; Allen and Widdowson, 1976; Master, 1986). Many instances in the corpus, however, contain an adverbial between the auxiliary and past participle ("Water is actively passed.", "The nerve cells are also linked together.", "The viruses are then released."). Many other examples of the distorted picture of syntax presented by published teaching materials could be cited, based upon the corpus.

USING THE DATA FROM THE ANALYSIS: AN EXAMPLE

A number of examples of how the data are used in course design have been given already in the above sections. Although this paper is primarily aimed at showing the potential of concordancing as a tool in course design in general (albeit by means of a case study of SQU), in order to see what a syllabus/set of materials employing concordancing might look like, this section will show how the data has been used in the production of a unit of teaching material. It is emphasized that this is only one way of using the concordancing technique. Concordancing can play a role in the design of other types of courses/elements of courses.

The unit in question is taken from the writing component of the SQU course. The syllabus for this unit is built upon three objectives: a task objective, a vocabulary objective and a grammar objective, as follows:

Task objective

To write a cohesive paragraph from diagrams, tables and prior knowledge describing structure and function (biology topic)

Vocabulary objective

Productive knowledge of 15 verbs used in describing structure and the interrelationship of parts (e.g. enclose, suspend, surround, contain, separate)
Grammar objective

To be aware of the relationship between transitivity and the passive voice.

To be able to judge when and when not to use the passive voice, with and without an explicit agent.

The task objective (writing a paragraph) is derived primarily from needs analysis, although concordancing does play a role in helping to identify the importance of the language functions of structure and function. The vocabulary and grammar leading up to the task, on the other hand, are derived primarily through concordancing. The verbs listed under the vocabulary objective are drawn from the frequency list, their use in describing structure being corroborated by means of concordancing. The grammar objective derives from an analysis of the concordances of the structure verbs, which demonstrate that such verbs vary in transitivity and therefore in their susceptibility to passivization.

The unit begins with an "input section", which presents a text on the structure of the eye. Information transfer and text labelling exercises help to familiarize learners with the text and draw attention to the language associated with structure and function. A "vocabulary" section presents instances of use of the structure verbs from the input text or the corpus (derived by means of the concordancer), which are compared with more "everyday", madeup examples, supported by visuals. A grammar section follows, again making use of examples from the input text or the corpus, and presenting and practising the relationship between transitivity and the passive. The unit culminates in a task which requires students to write a paragraph about the biological species "hydra" (something studied in the biology course), based upon a table and diagram. At this stage students make use of the language from the input and vocabulary and grammar sections (and which have been derived from the corpus) in negotiating the production of an authentic piece of English.

The degree to which the students make use of the language presented leading up to the task varies. More proficient students tend to be more creative and are able to deviate more from what is presented. Less proficient students, on the other hand, tend to rely more on what they have been taught. Experience has shown, however, that students at all levels make some use of the language presented and hence benefit from the use of the concordancing from which this language is derived.

CONCLUSION

This paper has shown how computerized text analysis can assist in the design and evaluation of an ESP course. The only necessary conditions for such a procedure are possession of the relatively simple frequency and concordancing programs, and a corpus of language drawn from the specific area of the targeted students. Armed with these, course designers can ensure that the language they present in their courses corresponds as closely as possible to the language that is actually required by learners in their specific purpose area. Although the technique itself is product-focussed, this paper has shown how concordancing can contribute equally well to both process- and product-oriented approaches to course design.
One anonymous reviewer of a previous version of this paper expressed the concern that teachers and course designers might simply use the frequency data from concordancing and "use the lists in class to walk students through". This certainly is not the approach envisaged here, as will be made clear in the course of this paper. Some might argue that this overt focus on language in a task-based approach is inappropriate [cf. Breen and Candlin (1980) and Prabhu (1987)], because, as Nunan (1989) [cited in Loschky and Bley-Vroman (1990: p. 163)] has put it, learners' attention, in the performance of a task, "is principally focused on meaning rather than form". However, there are signs that this view is losing ground (Nunan, 1989), with a growing awareness of the role instruction can play in the acquisition of grammar [see Loschky and Bley-Vroman (1990) for review]. Further, even in a meaning-focused approach such as that promoted by Breen and Candlin and Prabhu, learners will require units of meaning (lexical and grammatical items) with which to negotiate. In addition, there are a number of practical reasons for such an approach in the present context. First, the students come from an educational background based upon a traditional lock-step approach to learning, with much emphasis on rote; they thus appreciate something concrete to hold on to. Second, in post-course questionnaires the students repeatedly asked for more "grammar". Third, the large body of teachers involved in the course also requested a focus on language, as well as tasks. Fourth, language drawn from the science course had high face-validity with students. And fifth, the language presented in the lead-up to the tasks, experience demonstrated, was useful in the performance of the tasks.

Bearing in mind with all frequency data quoted that to make a valid comparison with published frequency lists, as outlined in the previous section, account needs to be taken of the different treatment of derivatives, plurals, inflected verb forms and homographs, and that numbers consequently need to be approximately halved. Frequency data relating to Cobuild are taken from Sinclair and Renouf (1988).

Numbers in parentheses refer to the number of instances of a word in the corpus.

This may be related to the fact that most of the corpus consists of spoken (albeit academic) language, where the range of connectors is narrower than in the written mode.

The unit in question was written by James Scott. For reasons of space and to highlight the role of concordancing, the description which follows is simplified. The unit is in fact more complex than as described and other pedagogical factors enter into its format besides those mentioned.

REFERENCES


APPENDIX: PART OF A CONCORDANCE FOR "WELL"

Actually to be precise / leave out

/ akaryotic / prokaryotic / and euk alright / we've heard lots of useful / and for this reason
hydro has got / and so you could add them here on / and the bacteria has a nice place / and
within the lumbricidae there and water can move out through the water / so that it's rather like
carbohydrate / it's called a pep being / as well as causing some d causing some disease / now
there a very similar substance is found the length yes yes / yes / in i / I don't know if you know
that e the animal cell / now you can see another / and these are the are conducting / is to give
strength its main function which is to co school / yeh / the smell of bad the moment there's one
and a hal

you know what these are / well a these words can be adjectives as well own food / how do
do they do it / yes well sensitive / it has to reproduce as well anules under normal conditions as
well rom the bacteria / so it can grow well here are various other families as well phloem / these
need some water as well vacuole might contain solutes as well as all because it contains
protein as well as very useful kind of organisms as well f organisms as well as being / as
well as found in animals and in plants as well as s another difference / not so / as well as hey
have a cellulose cell wall as well as e things are in the plant cell as well as hates / sulphates
/potassium / as well as of the functions of the xylem / as well as ignin / so the xylem supports
/as well as before in any chemistry classes / well at so that here we have / somewhere / well at
r has a concentration gradient as well but it cannot pass through the memb so this is a the
sieve element as well but this one here is maybe cut ther membrane is in the plant cell as
well / but on the outside of the membrane ma which doesn't shown up quite so well but we can
see one or two other thi what else / the stem needs it as well / but why else do the leaves
need wa the root / okay / any questions well call them the root cap cells / okay u got the
idea secretorybut where well can somebody tell me better what is
/can you think of any other root wh / collenchyma parenchyma and sclere contractile vacuoles /
these are us enough to know / if there are races finger-like projections of the cell glycoprotein
on the surface of the / have a look / we'll have a look a / heredity it contains what does it
/hold your horses / there are a hug ena has a flagellum /i don't think well hydra does have
cilia but it's not a / in fact it's an interesting / well i think this might be interesting to
cells nerve cells that's five and well i will write these down and add the 's the answer to
question thirteen well i think the you the safest thing to well i am very pleased to see you all
aga well i think you seem to have remembered well / i think we'll just leave it at th well i
think we can finish off there for well / if the pressure in the cells well / if there are any
questions later p well in the cell wall / and some of these well in the root as well the water
can well in your readings there is another wo well it depends on the actual substance well"it
won't explode or blow up but the well it protects its eyes because it has well it depends on what
sort of cell / i well / it comes from the air into the le well / it just means a place okay / now
"t well it doesn't form chromosomes in the well / it has photosynthesis chloroplast
 carrot which / other roots do as well tem / the roots and the flowers as well discharged through
the cytoproct / well t think people have studied hydra well n is / a finger-like projection /
well cant / it lubricates the surface / well is outside again from the phloem / well is the
function of the nucleus / well no one knows how many there are / well
eif halak / alhamdulillah / zain / e looking at plant anatomy / right do you have them with you
now / oh ne cell isn't it and .... / right t get the light from the sun very t to ask a special
question now / parts other types of molecules as s through these pits very easily / any questions
about this / okay /
then is secreted from the cell / eventually what will happen to / t its eyes from outside
stimulus) ma membranes / of adjacent cells / the c-o-2 comes from the leaf as ineering and
probably medicine as ng organisms / forms chromosomes / n because it can make some food as