

Microcomputer Databases for English Language Teaching Research¹⁾

Shigenobu TAKATSUKA

ABSTRACT

This paper is an introduction to the construction and the use of microcomputer databases for research into some aspects of English language teaching. It describes and analyzes how microcomputer databases are made using a database management system commercially available, and what roles they have to play in the research. It also identifies their advantages and limitations.

1. Introduction

In this paper an attempt will be made to show how microcomputer databases can be made and profitably utilized for English language teaching research. Although the database facility is only one of many uses for the microcomputer, which has been increasingly popular in schools, it seems to be the next step the teacher can take once he or she has become familiar with the word-processing facility.

The computer can carry out data processing which would be tedious and time-consuming if carried out manually. Although non-computer databases, which are usually stored on filing cards, have been used by many researchers/teachers, they are likely to be replaced by computer databases.

There are several databases available for the studies of English linguistics and applied linguistics, ranging from large-scale corpora generated by mainframe computers such as the Brown Corpus of American English and its British counterpart, the LOB (Lancaster-Oslo-Bergen) Corpus, to small-scale lists of words and sentences in the English textbooks published in Japan (Miura et al. eds. 1982 ; Aoki and Tanaka eds. 1984). Some are available on magnetic tape, some on microfiche, some on disc and others in print.

Computer-generated concordances based on large amounts of written English data such as the Brown Corpus are not easily accessible, though they would provide the researcher/teacher with better descriptive information on the English language. The databases such as that of Miura et al. (eds.)(1982) on the other hand have no problem of accessibility because they have been published in print, but they are subject to some limitations:

- (1) they are very voluminous,
- (2) they are likely to be limited in utility for the user because of the editor's specific purposes, e.g. categorizing all the sentences in the English textbooks used at junior high schools according to their communicative functions (Aoki and

Tanaka eds. 1984), and

- (3) the user cannot add new data to the database in order to be able to scan and manipulate the existing data from the viewpoint of the added data, e.g. the length and linguistic pattern of each sentence.

The microcomputer database seems to have advantages over the mainframe computer database in terms of both accessibility and flexibility. We can also identify the following advantages of microcomputer databases over non-computer, i.e. paper, databases:

- (1) data stored in a database file can readily be shared by other researchers/teachers simply by copying the original disc,
- (2) relevant information can easily and quickly be searched for by simple commands which specify conditions for information searches,
- (3) recorded data can be rearranged, e.g. in alphabetical order of a specific item(= field) of data, and
- (4) not only new records (=integral units of data item) but also new fields can be added to the original database by the user.

These seem to justify the compilation of a general-purpose microcomputer database, from which users can obtain whatever information they need and to which they can add new data which are relevant to their specific needs and interests enabling them to make information searches based on the newly-added data as well as the existing ones.

2. Examples of microcomputer database

In what follows seven different databases will be introduced with reference to what structures of data storage they have, what types of tasks the database software can handle with them, and what positive elements they have to offer the researcher/teacher. They are merely a few of the possible applications of the database program for English language teaching research and are not meant to represent an exhaustive list.

At present there are several database management systems commercially available, e.g. R:BASE 5000 and dBASEII, to name a few. All the databases in the present paper have been made with the 2.4 version of a database system called 'Japanese dBASEII'(Ashton-Tate Corp. 1984). The program enables us to store data as a collection of records which are divided into 'fields', each of which carries a different kind of information. The form and arrangement of information to be entered is called the 'structure' of the file. The structure should be so designed as to store information in a manner which makes it easy for the user to locate a specific piece of information. Some of the dBASEII specifications are given below.

- The maximum length permissible for one record is 1,000 characters,
- the maximum length permissible for one field is 254 characters, and
- the maximum number of fields permissible for one record is 32.

The microcomputers used in the study are the NEC VM2 and M2 models in the PC-9801 series. The names of the database files made are given below.²⁾

- | | | | |
|--------------|-----------------|---------------|--------------|
| (1) WORD.DBF | (2) GRAMMAR.DBF | (3) BIB.DBF | (4) HIGH.DBF |
| (5) TEXT.DBF | (6) VOCAB.DBF | (7) USAGE.DBF | |

(1) WORD.DBF

The first database file, WORD.DBF is designed so as to demonstrate what a database

file is like and what the database management system can do with the recorded data. The file contains all the words which are specified as those which should be taught in any of the grades 1, 2, and 3 by three versions of *The Course of Study for Lower Secondary Schools*, published in 1958, 1969, and 1977. Four character fields are made to contain four different items of information: word (Field WORD) and information from the three versions of the course (Fields A, B, and C). Each word marked with a combination of three pluses and minuses depending on which of the three courses of study include it as one that should be taught (+ + +, for example, indicates that the word is specified as obligatory by all three courses).

If you want to know which of those words that are obligatory in the 1969 version are omitted in the 1977 version, all you have to do is to type in the following command at a keyboard (Quotation marks are used to identify the beginning and the end of a character string which is stored in a field in the record). The printer can also produce a copy of whatever is displayed on the screen.

.LIST FOR B = "+" .AND. C = "-"

Table 1

(a)	WORD	A	B	C	(a)	WORD	A	B	C
00003	above	+	+	-	00248	hold	+	+	-
00010	almost	+	+	-	00252	horse	-	+	-
00016	America	+	+	-	00254	hotel	-	+	-
00017	American	+	+	-	00261	ill	+	+	-
00028	apple	-	+	-	00263	ink	+	+	-
00033	Asia	+	+	-	00279	knife	+	+	-
00038	autumn	+	+	-	00280	knock	-	+	-
00040	baby	-	+	-	00282	lady	+	+	-
00043	bag	-	+	-	00291	lesson	+	+	-
00044	ball	-	+	-	00295	lie	+	+	-
00045	baseball	-	+	-	00298	lily	-	+	-
00046	basket	-	+	-	00306	low	+	+	-
00047	bath	-	+	-	00308	mail	-	+	-
00051	bed	-	+	-	00312	map	-	+	-
00052	been	+	+	-	00317	meat	-	+	-
00055	behind	+	+	-	00319	men	+	+	-
00056	bench	-	+	-	00320	mile	+	+	-
00057	beside	+	+	-	00337	Mt.	+	+	-
00058	best	+	+	-	00342	myself	+	+	-
00059	better	+	+	-	00358	north	+	+	-
00061	bicycle	-	+	-	00377	orange	-	+	-
00065	blackboard	+	+	-	00381	ourselves	+	+	-
00075	bridge	-	+	-	00385	paint	-	+	-
00078	brown	-	+	-	00392	perhaps	+	+	-
00086	cake	-	+	-	00393	piano	-	+	-
00089	cap	-	+	-	00398	pocket	+	+	-
00092	cat	-	+	-	00399	pond	-	+	-
00094	center(-re)	-	+	-	00401	present	+	+	-
00096	chalk	+	+	-	00403	pupil	+	+	-
00098	children	+	+	-	00408	quite	+	+	-
00103	climb	-	+	-	00409	radio	-	+	-

Shigenobu Takatsuka

00104	clock	+	+	-	00411	rainy	-	+	-
00107	cloudy	-	+	-	00413	reach	+	+	-
00109	coat	-	+	-	00422	road	-	+	-
00110	coffee	-	+	-	00424	rose	-	+	-
00116	could	+	+	-	00425	round	+	+	-
00118	cow	-	+	-	00444	shine	+	+	-
00119	cross	+	+	-	00445	shoe	-	+	-
00123	dance	-	+	-	00451	shut	+	+	-
00130	diary	+	+	-	00461	skate	-	+	-
00132	did	+	+	-	00462	ski	-	+	-
00134	dish	-	+	-	00476	song	+	+	-
00136	doctor	+	+	-	00479	south	+	+	-
00138	dog	-	+	-	00490	stone	-	+	-
00139	doll	-	+	-	00499	sugar	-	+	-
00149	earth	+	+	-	00503	supper	+	+	-
00150	east	+	+	-	00509	tea	-	+	-
00153	egg	-	+	-	00512	telephone	-	+	-
00161	else	+	+	-	00513	television	-	+	-
00162	England	+	+	-	00516	tennis	-	+	-
00165	Europe	+	+	-	00525	themselves	+	+	-
00166	even	+	+	-	00530	thing	+	+	-
00182	feet	+	+	-	00537	though	+	+	-
00184	field	-	+	-	00551	train	-	+	-
00194	floor	-	+	-	00575	was	+	+	-
00211	gate	-	+	-	00584	were	+	+	-
00212	gentleman	+	+	-	00585	west	+	+	-
00221	grass	+	+	-	00596	wide	+	+	-
00226	had	+	+	-	00601	wish	+	+	-
00233	hat	-	+	-	00606	women	+	+	-
00242	herself	+	+	-	00607	wood	-	+	-
00244	hill	-	+	-	00623	yourself	+	+	-
00246	himself	+	+	-	00624	yourselves	+	+	-

(a)RECORD NUMBER

(a)RECORD NUMBER

Table 1 shows that some past and past participle forms of verbs (e.g. *been*, *did*, *was*, and *were*), some plural forms of nouns (e.g. *children*, *men*, and *women*) and directions (i.e. *north*, *south*, *west* and *east*) are dropped out of the 1977 version. On the other hand, what words are added to the 1977 version?

The command

.LIST FOR B = "-" .AND. C = "+"

supplies us with the following information shown in Table 2.

Table 2

(a)	WORD	A	B	C
00105	close	-	-	+
00224	ground	-	-	+
00449	shout	-	-	+
00496	student	-	-	+
00563	until	-	-	+
00613	yard	-	-	+

(a)RECORD NUMBER

These tables tell us that 'shut' and 'pupil' in the 1968 version are replaced by 'close' and 'student' respectively in the 1977 version.

(2) GRAMMAR.DBF

The GRAMMAR database file stores all the sentence types, sentence patterns, and grammatical points which are specified as those which should be covered at a certain grade by the three versions of *The Course of Study*. Four fields are allocated to contain a sentence type, pattern, or grammatical point (Field PATTERN) and information from the three versions (Fields A, B, and C). In the file, each sentence type, sentence pattern, or grammatical point is followed by three numbers 1, 2, and 3, for example, which means that the item is specified by the 1958 version as needing to be taught in the first grade, in the second grade by the 1969 version and in the third by the 1977 version. Numeric fields are chosen for the fields A, B and C so that arithmetic processing can be done later.

If you want to see what items are reserved for teaching at later grades, you can get relevant information as shown in Table 3 by typing in

```
.LIST FOR A < B .AND. A <> 0
```

In the command, the inequality $A < B$ means that for a given sentence type, sentence pattern, or grammatical point the year specified for teaching it in the 1969 version is larger (later) than in the 1958 version. Table 3 shows that the adjectival use of infinitive and exclamatory sentences are, among others, reserved for teaching in later grades.

Table 3

(a)	(b)	A	B	C
00024	G: TO (ADJECTIVAL)	2	3	2
00029	P: DON'T V COMMAND	1	2	0
00030	P: HERE+V (=BE)	1	2	2
00031	P: HOW+ADJECTIVE+S+V (=BE)	1	2	2
00038	P: MAY, MUST QUESTION	1	2	2
00039	P: S+ASK, TELL, WANT+O+TO (NOMINAL)	2	3	3
00041	P: S+V (<>BE)+C (=ADJECTIVE)	2	3	2
00061	P: S+V+O+C (=ADJECTIVE)	2	3	0
00065	P: S+V+O+C (=V-ING)	2	3	0
00068	P: V (=BE) COMMAND	1	2	2
00070	P: WHAT (+DET)+ADJECTIVE+NOUN+S+V (=BE)	1	2	2
00071	P: WHAT (+DET)+ADJECTIVE+NOUN+S+V (=HAVE)	1	2	2
00076	T: EXCLAMATORY SENTENCE	1	2	2

(a) RECORD NUMBER (b) PATTERN (G: GRAMMATICAL POINT, P: SENTENCE PATTERN, T: SENTENCE TYPE)

Similarly, the following command tells the computer to identify the items which are originally in the 1969 version but are dropped out of the 1977 version.

```
.LIST FOR B <> 0 .AND. C = 0
```

In the command, the symbol $<>$ is used to indicate 'not equal to'.

Table 4

(a)	(b)	A	B	C
00004	G: PASSIVE-FUTURE	2	2	0
00014	G: PRESENT-PERFECT-PROGRESSIVE	3	3	0
00018	G: RELATIVE ADVERB (RESTRICTIVE)	3	3	0
00019	G: RELATIVE ADVERB (WHEN, WHERE)	3	3	0
00029	P: DON'T V COMMAND	1	2	0
00037	P: LET US+V	1	1	0
00047	P: S+V+IO+DO (=HOW, WHAT, WHEN, WHERE+TO)	3	3	0
00049	P: S+V+IO+DO (=THAT CLAUSE)	3	3	0
00061	P: S+V+O+C (=ADJECTIVE)	2	3	0
00064	P: S+V+O+C (=V)	3	3	0
00065	P: S+V+O+C (=V-ING)	2	3	0
00069	P: WHAT (+DET)+ADJECTIVE+NOUN+S+V (<>BE, HAVE)	3	2	0
00081	T: TAG QUESTION	3	3	0

(a) RECORD NUMBER (b) PATTERN (G: GRAMMATICAL POINT, P: SENTENCE PATTERN, T: SENTENCE TYPE)

Table 4 tells us that passive-future and present perfect progressive sentences, the SVOC structure with non-noun objective complements and others are left out of the junior high school English syllabus.

If the next Course of Study also gives a list of required words and allocates sentence patterns, sentence types, or grammatical points to the three grades, information from the new version can be added to the existing files so that we can compare the four versions.

(3) BIB.DBF

The structure of the BIB database file is designed, as the name suggests, to register books and articles and find where they are and what books and articles are available on a specific topic. The BIB database file stores data on a considerable number of books and papers in the area of English language teaching research with reference to their author, publishing year, title, source (i.e. publisher or journal, volume, number and pages), category, and sub-category if necessary. The file is expected to help our students to find references on a specific topic for their studies.

The following represent typical ways in which information searches are made:

(1) information can be searched for simply by indicating a 'key' word in the title of a book or a paper. If you want access to all the books and papers with titles containing the key word, 'Communicative' or 'communicative' for example, all you have to do is to type in at the keyboard:

```
LIST FOR "ommunicative"$ Title1 .OR. "ommunicative"$ Title2
```

In this command, the dollar sign is a sort of shorthand for 'contained in' and " " is used to identify the boundary of a character string. The string "ommunicative" is used here to include both 'Communicative' and 'communicative'.

Table 5

(a)	(b)	(c)	(d)	(e)	(f)						
00023	Bachman, L. F. and Palmer, A. S. (1982)		The construct validation of some components of communicative proficiency. TESOL Q 16, 4, 449-465. testing communicative competence								
00058	Brown, J. W. (1984)		Communicative competence vs. communicative cognizance: Jakobson's model revisited. CMLR 40, 5, 600-615. communicative competence								
00059	Brumfit, C. J. (1979)		Communicative language teaching: an educational perspective. Brumfit, C. J. and Johnson, K. (eds.) (1979:183-191). communicative language teaching								
00060	Brumfit, C. J. (1984)		Communicative Methodology in Language Teaching. The Role of Fluency and Accuracy. Cambridge: C. U. P. communicative language teaching								
00061	Brumfit, C. J. and Johnson, K. (eds.) (1979)		The Communicative Approach to Language Teaching. Oxford: O. U. P. communicative language teaching								
00066	Canale, M. (1983)		From communicative competence to communicative language pedagogy. Richards, J. C. and Schmidt, R. W. (eds.) (1983: 2-27). communicative competence								
00092	Cziko, G. A. (1984)		Some problems with empirically-based models of communicative competence. Applied Linguistics 5, 1, 23-38. communicative competence								
(a)	RECORD NUMBER	(b)	AUTHOR	(c)	YEAR	(d)	TITLE	(e)	SOURCE	(f)	CATEGORY 1
(g)	CATEGORY 2										

Table 5 lists books and papers with titles containing the word 'communicative'.

(2) Alternatively, you can set information searches for all the books and articles written by a particular author, "Allwright" for example, by the following command (See Table 6).

.LIST FOR Author = "Allwright"

Table 6

(a)	(b)	(c)	(d)	(e)	(f)	(g)
00008	Allwright, R. L. (1976)		Language learning through communication practice. Brumfit, C. J. and Johnson, K. (eds.) (1979:167-182). communicative language teaching individualization			
00009	Allwright, R. L. (1978)		Abdication and responsibility in language teaching. Studies in Second Language Acquisition 2, 1, 105-121. individualization			
00010	Allwright, R. L. (1981)		What do we want teaching materials for? ELTJ 36, 1, 5-18. individualization			
00011	Allwright, R. L. (1982)		Perceiving and pursuing learner's needs. Geddes, M. and Sturtridge, G. (eds.) (1982: 24-31). individualization			
00012	Allwright, R. L. (1983)		Classroom-centered research on language teaching and learning: a brief historical overview. TESOL Q 17, 2, 191-204. classroom language teaching			
00013	Allwright, R. L. (1984a)		Differentiation. Unpublished lecture handout, M. A. E. L. T. Lancaster Univ. individualization			
00014	Allwright, R. L. (1984b)		The importance of interaction in classroom language learning. Applied Linguistics 5, 2, 156-171. classroom language teaching			

(a) RECORD NUMBER (b) AUTHOR (c) YEAR (d) TITLE (e) SOURCE (f) CATEGORY 1
(g) CATEGORY 2

(3) Yet another way of making an information search is to specify one of the categories under which each record of data (=each book or article) is put. The command typed in below can scan all the records of data which are categorized under the heading of 'CAI' (See Table 7).

.LIST FOR Category = "CAI"

Table 7

(a)	(b)	(d)
	(c)	(e)
		(f)
		(g)
00021	Asahi (1985)	朝日新聞 「朝日C A I シンポジウム」から、コンピューターの可能性 8月15日 CAI
00121	Fox, J. (1984)	Computer-assisted vocabulary learning. ELTJ 38, 1, 27-33. CAI vocabulary learning
00151	Higgins, J. T. (1983)	Computer assisted language learning. Language Teaching 16, 2, 102-114. CAI survey article
00152	Higgins, J. T. (1984a)	Reading and risk-taking: a role for the computer. ELTJ 38, 3, 192-250. CAI reading
00153	Higgins, J. T. (1984b)	Computers in Language Learning. Collins ELT. CAI
00154	Higgins, J. T. (1985)	Grammarland: a non-directive use of the computer in language learning. ELTJ 39, 3, 167-173. CAI grammar
00159	Holmes, G. and Kidd, M. E. (1982)	Second-language learning and computers. CMLR 38, 3, 503-516. CAI

(a) RECORD NUMBER (b) AUTHOR (c) YEAR (d) TITLE (e) SOURCE (f) CATEGORY 1
(g) CATEGORY 2

Additional records of data can be appended to the existing file. All the records then can be rearranged again and stored in a new file in alphabetical order of data in the 'Author' field by the following command.

.SORT ON Author TO B: (New File Name)

The following command tells the computer to order the records in alphabetical order of

the data in the 'Author' field and in ascending order of the data in the 'Year' field within the records with the same data in the 'Author' field.

.INDEX ON Author + Year TO B:(New File Name)

(4) HIGH.DBF

The HIGH database file contains all the sentences which are to be found in the lessons of all the junior high school first-year English textbooks³⁾ which are ready for use from the 1987-1988 school year onward. This file might be called a computerized version of Miura et al. (eds.) (1982). In the file, fields are made to accommodate information indicating in which textbook, grade, lesson, or section, and on what page each sentence appears and whether or not a given sentence is a target sentence.

Once the text has been entered into the computer, it is easy to derive a concordance from it automatically. When you want to see how the word '*what*' is used, for example, the command to obtain relevant information is:

.LIST FOR "What"\$ S1 .AND. "what"\$ S1.

The LIST FOR command is used here to locate all the occurrences of the word '*what*' in the stored file (See Table 8).

Table 8

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
00037	What is this?						
	SE	1	3	3		14	*
00065	What do you have in your hand?						
	SE	1	5	3		24	*
00134	What's that?						
	SE	1	9	2		46	
00139	What's its name?						
	SE	1	9	3		47	
00180	What time is it?						
	SE	1	12	1		62	
00269	What's this?						
	NH	1	3	1		8	*
00347	What do you have in that bag?						
	NH	1	6	3		27	
00405	What time is it, Ken?						
	NH	1	9	1		42	*
00422	What time do you leave for school, Ken?						
	NH	1	9	3		44	
00464	What are doing now?						
	NH	1	11	2		53	*
00501	What am I?						
	NH	1	12	1		62	
00527	What are you doing, Lucy?						
	NH	1	13	1		67	
00573	What's in it, Ken?						
	NH	1	14	3		77	*
00584	What time is it in Japan now?						
	NH	1	16	0		80	

Microcomputer Databases for English Language Teaching Research

00666	What is this?	NC	1	5	1	27	
00668	What is that?	NC	1	5	1	27	
00761	What do you have in your hands?	NC	1	8	3	48	*
00764	What do you see on them?	NC	1	8	3	53	
00824	What time is it?	NC	1	11	1	63	*
00831	What time do you go to bed?	NC	1	11	2	64	

(a) RECORD NUMBER (b) SENTENCE
(c) TEXTBOOK (d) GRADE (e) LESSON (f) SECTION
(g) PAGE (h) TARGET SENTENCE

In this way the teacher can easily check in what contexts a particular word occurs in textbooks other than the one he is expected to use. This seems to be a valuable and time-saving aid enabling the teacher to have a general view of the treatment of a given word in the textbooks.

Alternatively, when you want to see all the examples of interrogative sentences, the command is:

.LIST FOR "?"\$ S1

Some of the examples are shown in Table 9.

Table 9

(a)	(b)						
		(c)	(d)	(e)	(f)	(g)	(h)
00011	Is this your bag?	SE	1	2	1	6	*
00016	Is that a school?	SE	1	2	2	8	
00021	Is that a store?	SE	1	2	3	9	
00027	Are you a teacher?	SE	1	3	1	12	*
00033	Is it a river or a lake?	SE	1	3	2	13	*
00037	What is this?	SE	1	3	3	14	*
00040	Hello, how are you?	SE	1	4	1	18	
00050	Do you like rock music?	SE	1	4	3	20	*
00052	Do you play the guitar?	SE	1	4	3	20	
00062	Do you go to school by bus in New York?	SE	1	5	2	23	
00065	What do you have in your hand?	SE	1	5	3	24	*

Shigenobu Takatsuka

00068	Do you like science?						
	SE	1	5	3		24	
00086	Is Hokusai famous in your country?						
	SE	1	6	3		33	
00088	Is he still alive?						
	SE	1	6	3		33	*
00090	Who is the tall boy?						
	SE	1	7	1		36	*
00093	And who is that boy?						
	SE	1	7	1		36	
00099	Do you play basketball?						
	SE	1	7	2		37	
00104	Who are those girls?						
	SE	1	7	3		38	
00110	Do you hear the sound?						
	SE	1	8	3		42	*
00120	Do you have any falls in Japan?						
	SE	1	8	3		42	*
00124	Are these your books?						
	SE	1	9	1		45	
00128	How about your sister?						
	SE	1	9	1		45	

(a) RECORD NUMBER (b) SENTENCE
(c) TEXTBOOK (d) GRADE (e) LESSON (f) SECTION
(g) PAGE (h) TARGET SENTENCE

Similarly, you can set information searches for all the examples of negative sentences just by typing in at the keyboard:

.LIST FOR "not"\$ S1 .OR. "n't"\$ S1

Some of the examples are shown in Table 10.

Table 10

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
00019	That is not a park.						
	SE	1	2	3		9	*
00022	No, it is not.						
	SE	1	2	3		9	
00060	I don't like rush hour.						
	SE	1	5	2		23	*
00061	I don't, either.						
	SE	1	5	2		23	
00063	No, I don't.						
	SE	1	5	2		23	
00089	No, he isn't.						
	SE	1	6	3		33	*
00100	No, I don't.						
	SE	1	7	2		37	
00113	Don't fall in, Kenji.						
	SE	1	8	2		41	*

Microcomputer Databases for English Language Teaching Research

00116	Don't walk too fast, Kenji.					
	SE	1	8	2	41	
00122	They aren't so big.					
	SE	1	8	3	42	
00152	No, she doesn't.					
	SE	1	10	2	51	*
00168	No, I can't.					
	SE	1	11	2	55	*
00210	No, he isn't.					
	SE	1	13	2	68	*
00231	No, there aren't.					
	SE	1	14	2	73	*
00264	No, it isn't.					
	NH	1	2	2	6	*
00286	No, she isn't.					
	NH	1	3	3	11	
00297	No, I don't.					
	NH	1	4	2	14	*
00320	No, I'm not.					
	NH	1	5	2	19	
00367	No, he doesn't.					
	NH	1	7	2	30	*
00377	But she doesn't like sushi.					
	NH	1	7	3	31	*
00392	No, she doesn't have any sisters.					
	NH	1	8	2	38	*
00394	No, I don't.					
	NH	1	8	2	38	

(a) RECORD NUMBER (b) SENTENCE
(c) TEXTBOOK (d) GRADE (e) LESSON (f) SECTION
(g) PAGE (h) TARGET SENTENCE

The database file might attract a larger number of users if we added new fields such as sentence patterns, the number of words in each sentence, and a range of communicative functions for each sentence.

(5) TEXT.DBF

The TEXT database file contains all the sentences in the listening sections of *Break-through* Books 1, 2, and 3 (Richards and Long 1984, O.U.P.). The file is designed so that we can compare the language content of the English language textbooks published in Japan with the content of a British counterpart and find what words and linguistic structures are missing in the former which are frequent in the latter. The database uses the same structure of fields as the HIGH database file.

By juxtaposing the sentences which contain a particular word or a particular combination of words from the HIGH database file and those from the TEXT database file, the researcher/teacher may be able to identify differences between the two corpora in terms of the frequency and usage of that linguistic form. This information may serve as a basis on which materials developers in Japan can improve their textbooks. It may also be a resource which would help the practising teacher to gain new insights into

present day English since he is likely to lose sight of the reality of the language he has been teaching.

A sample survey, though small and preliminary, was made of the adverbial use of 'please' in the materials. The information obtained from the TEXT. DBF was compared with the treatment of 'please' in the junior high school textbooks. The information on the junior high school textbooks was taken from Miura et al. (eds.) (1982) since the HIGH database file has not contained data from second and third year textbooks. Each occurrence of adverbial 'please' was noted, providing a total of 86 and 47 examples from the two sources respectively. All the examples were classified according to what syntactic structures they cooccurred with.

Table 11

(a)	(b)
00014	B: Good evening. A table for two, please.
00015	A: Certainly, Sir. Please come with me.
00035	A: May I speak to Mrs Williams, please?
00036	B: Yes. Who's speaking, please?
00038	B: One moment please, Mr Hayes.
00039	A: Hello, is Mr Woods there, please?
00046	A: Could I speak to Miss Jones, please?
00047	A: Could you help me with this form, please? My English is not so good.
00054	B: How do you spell it, please?
00062	A: I see. And one more question, please. Your occupation. What do you do?
00108	A: Can I borrow it, please?
00110	A: And could I use your ruler for a few minutes, please?
00133	A: Would you be able to return this book to the library for me, please? It's overdue.
00151	A: I want some toothpaste, please.
00157	A: I'll have a large one, please.
00158	A: I want to send this letter to America, please.
00160	A: Yes, three air letter sheets, please. What do they cost?
00163	B: A dollar, please.
00165	A: How much is that ring, please?
00216	A: What time is my flight, please?
00222	B: Yes. What hours are you open today, please?
00276	A: Excuse me, how do I get to the station, please?
00283	A: Thank you. And where's the National Bank, please?
00289	A: Where's that, please?
00292	A: Excuse me, please. I'm looking for Wain Street.
00294	A: I'm trying to find Hill Street, please.
00297	A: Can you tell me how to get to Fox Street, please?
00299	A: Do you know where the Public Library is, please?

(a) RECORD NUMBER (b) SENTENCE

Table 11 gives some examples of 'please' from the TEXT. DBF and Table 12 gives a comparison of the two corpora in terms of the distribution of 'please' to those syntactic structures. It shows that the ESL textbook *Breakthrough* includes 'please' in a wide range of syntactic structures, while the Japanese junior high school textbooks include it in a very limited range of structures, predominantly in the imperative (76.6%).

Table 12

	<i>Breakthrough</i>	High School Textbooks
Imperative + please	6	36
Can you/Will you + please?	10	1
Can I/May I + please?	10	2
Questions asking for information + please	22	0
Statements + please	15	0
I want/I'd like/I'll have NP	6	0
I want to/I'd like to VP	7	0
Others (e. g. That'll be NP)	2	0
NP/Adverbial Phrase + please	18	1
Yes + please in response to offers	5	3
Please	0	4
Total	86	47

Similarly, the following command enables us to scan for word combinations, '*what is/are X like*' for example.

.LIST FOR ("What's"\$ S1 .AND. "like"\$ S1) .OR. ("what's"\$ S1. AND. "like"\$ S1)
 .OR. ("What are"\$ S1 .AND. "like"\$ S1) .OR. ("what are"\$ S1 .OR. "like"\$ S1)

Table 13

(a)	(b)
00174	A: That must be an interesting place. What's it like?
00180	A: What's the weather like in Hawaii?
00190	A: And what are the restaurants like?
00271	A: And what are your working hours (like)?
00491	A: What's the living-room like?
00794	A: What's your home town like, Akita?
00802	A: And what's Hong Kong like, Jane?
00821	A: What's the weather like then?
00825	A: And what's the weather like in the summer?
00827	A: What's the weather like in Sydney, Lois?
01320	A: What's accommodation like in London?

(a) RECORD NUMBER (b) SENTENCE

We can also compare this information with the data from junior high school English textbooks (Miura et al. eds. 1982). The comparison tells us that the common and useful question '*what is/are X like?*' is not dealt with at all by any of the junior high school textbooks. This might be a reason why Japanese learners of English are often said to use '*what is/are X?*' when they should use '*what is/are X like?*' instead.

(6) VOCAB. DBF

The VOCAB database file stores data from the following word count lists and word lists:

- A word frequency list for junior high school English textbooks (1984)
- West, M. (1953) *A General Service List of English Words* (Longman).

—The first 1,000 words from Thorndike, E.L. and I. Lorge(1944)

The Teacher's Word Book of 30,000 Words (Columbia Univ.)

—*Longman Dictionary of Contemporary English* (1978)

'List of words used in the dictionary'

—A list of required words in *The Course of Study for Lower Secondary Schools* (1977)

The file is designed to clarify some similarities and differences among the lists. Information obtained from this file may be used as a basis for the establishment of core vocabularies.

If you want to see which of the words that are used by *LDCE* as its defining vocabulary do not appear at all in the junior high school textbooks, all you need to do is to type in:

LIST FOR LONGMAN = " + " .AND. HIGH SCHOOL = 0

Table 14 shows a small portion of those words which satisfy this condition.

Table 14

(a)	(b)	(c)	(d)	(e)	(f)	(g)
00002	ability	0	340	-	+	-
00006	abroad	0	268	-	+	-
00007	absence	0	213	-	+	-
00008	absent	0	91	-	+	-
00012	accordance	0	0	-	+	-
00013	according	0	1446	+	+	-
00014	account	0	968	+	+	-
00015	accustom	0	162	-	+	-
00016	ache	0	58	-	+	-
00020	action	0	1079	+	+	-
00021	active	0	338	-	+	-
00022	activity	0	477	-	+	-
00023	actor	0	0	-	+	-
00024	actress	0	0	-	+	-
00025	actual	0	380	-	+	-
00028	add	0	869	+	+	-
00029	addition	0	0	-	+	-
00031	adjective	0	0	-	+	-
00032	admiration	0	155	-	+	-
00034	admit	0	574	+	+	-
00036	advance	0	817	+	+	-
00037	advantage	0	556	-	+	-
00038	adventure	0	413	-	+	-
00039	adverb	0	0	-	+	-
00040	advertize	0	174	-	+	-
00042	advise	0	188	-	+	-
00045	affair	0	758	+	+	-
00046	afford	0	385	-	+	-
00052	afterwards	0	0	-	+	-
00067	alcohol (ic)	0	0	-	+	-
00068	alike	0	176	-	+	-
00069	alive	0	213	-	+	-

Microcomputer Databases for English Language Teaching Research

00071	allow	0	751	+	+	-
00076	aloud	0	93	-	+	-
00080	although	0	0	+	+	-
00081	altogether	0	239	-	+	-
00087	amongst	0	284	-	+	-
00088	amount	0	996	+	+	-
00089	amuse	0	123	-	+	-
00090	amusing	0	0	-	+	-
00093	ancient	0	671	-	+	-
00095	anger	0	119	-	+	-
00096	angle	0	199	-	+	-
00099	ankle	0	0	-	+	-
00102	annoy	0	55	-	+	-
00105	ant	0	0	-	+	-
00106	anxiety	0	112	-	+	-
00107	anxious	0	203	-	+	-
00109	anybody	0	204	-	+	-
00100	anyhow	0	70	-	+	-
00114	anywhere	0	111	-	+	-
00115	apart	0	270	-	+	-
00177	apparatus	0	0	-	+	-
00118	appear	0	1740	+	+	-
00119	appearance	0	495	-	+	-
00121	appoint	0	480	-	+	-
00122	approve	0	205	-	+	-

(a) RECORD NUMBER (b) WORD (c) HIGHSCHOOL (d) WEST (e) THORNDIKE
(f) LONGMAN (g) COURSE

(7) USAGE.DBF

The USAGE database file will make a good example of collecting examples of the usage of a particular word or structure used in a context and storing them not on filing cards but on a disc instead. This has two major advantages. One is that data can easily be shared by other researchers/teachers, and the other is that information can easily be searched for and rearranged from a particular viewpoint.

At present the file contains only a small number of examples: 'as + Adjective/ Adverb + as + Number/Time' as in 'as many as 1,000 people', 'out-' verbs as in 'outnumber' and the structure 'have + O + V-ing' as in 'Tom had us laughing'. Table 15 shows some examples of an 'as...as' construction.

Table 15

(a)	(b)
	(c)
	(d) (e) (f)
00001	as...as (early) As early as 1795, a would-be colonist named Issac Weld went to live in the New World and then returned to England complaining of Americans that "civility cannot be purchased from them on any terms..." AmE W TIME 11/5/84:58
00002	as...as (early)

...Peking as early as July 1979 issued a circular to make smokers more aware of this risks.
AmE W TIME 6/18/84:15
00003 as...as (far back)
And that fringe, I know as a fact, as far back as 1976 tried to solicit some of us and wanted to get a third party started.
AmE W TIME 11/19/84:25
0004 as...as (high)
Heart replacements could run considerably higher than that (some guesses go as high as \$40 billion).
AmE W TIME 12/10/84:8
00005 as...as (high)
Many (sex therapists) report a success rate as high as 75% to 90% for orgasm and premature-ejaculation problems but less than 50% for desire problems.
AmE W TIME 4/4/83:50
00006 as...as (high)
Some lawyers' bills are as high as \$5,000.
AmE W TIME 1/16/84:41

(a) RECORD NUMBER (b) FORM (c) EXAMPLE (d) VARIETY (e) REGISTER
(f) SOURCE

3. Conclusions

In the preceding sections some examples of small databases have been used to demonstrate the simple and versatile features of a database management system. Although it has been proved to be able to undertake a certain range of text retrieval and storing tasks, the following limitations can be identified of the use of the database system commercially available.

Unlike a program written by the user, the commercial database program, though versatile to a certain extent, cannot satisfy the user's specific needs, e.g. counting the number of words in each sentence, displaying a fixed number of words preceding and following each word-token as a KWIC concordance, and displaying it in a different color. The ready-made program necessarily restricts the form of what we can produce. Here lies the advantage of the PC-KWIC developed by Kinoshita and Matsuo (1986).

For most general purposes, however, the advantages of the ready-made system seem to outweigh its disadvantages. The following steps must be taken to make the most of the database software:

- (1) The researcher/teacher must be prepared to explore other possible ways of using the database software.
- (2) Since most of the database files introduced in this study are so small, they are valuable only as an illustration. They should continuously be supplied with further data so that they can be more useful for a greater number of users, since the TEXT. DBF, for example, has obvious limitations determined by its size and the kind of text it contains.
- (3) Researchers/teachers must work together to put their databases to practical use, sharing the responsibility of entering data into the databases.

Notes

- 1) This paper is a revised version of a paper read at the 17th Annual Convention of the Chugoku Academic Society of English Language Education held at the School of Education, Shimane University on October 4, 1986.
- 2) You can have all the database files introduced in this paper if you send me floppy discs. Address: Shigenobu Takatsuka, School of Education, Okayama University, 3-1-1 Tsushima-naka, Okayama 700.
- 3) The HIGH database file will be completed with data entered from second and third year textbooks by the end of 1986 so that junior high school English teachers can use it.

Acknowledgments

I would like to express my thanks to Mr Masahiro Yamada of Fuzoku Junior High School (attached to the Faculty of Education, Okayama University) for his word frequency list for junior high school textbooks.

I would also like to thank Mrs Yoko Morimoto and Mr Shoji Itami of Okayama University for typing in several vocabulary lists and part of the junior high school first-year textbooks, respectively.

Special thanks are due to Mr Kevin Mark, my colleague at Okayama University, for his comments and revision of this paper.

References

- Aoki, S. and M. Tanaka (eds.) (1984) *A Complete List of Sentence Functions in Junior Highschool English Textbooks*. Hyogo University of Education.
- Byers, R.A.(1982) *Everyman's Database Primer: Featuring dBASEII*. Culver City, Calif.: Aston-Tate.
- Kinoshita, T. and M. Matsuo (1986) 'PC-KWIC: Utilization of Personal Computers for Material Development' *CASELE RESEARCH BULLETIN* 16, 35-44.
- Leech, G. and A. Beale (1984) 'Computers in English Language Research' *Language Teaching* 17, 3, 216-229.
- Miura, S., K. Tomoeda, and T.Kawata (eds.) (1982) *A Complete List of Words and Sentences in English Textbooks Published in Japan — Junior High School — 1982*. Hiroshima: Keisuisha Publishing Company, Ltd.

(昭和61年10月31日受理)