A Corpus for Spoken EFL Lexicon: learner-based definition vs. dictionary-based definition

TOMITA Kaoru
(English Phonetics)
NAKAYAMA Kazuo
(Psychology of Language)
Mark ANTHONY
(Applied Linguistics)

ABSTRACT

Reading a target language dictionary definition in a printed dictionary is one common way of learning a new word in foreign language. Digital dictionaries enable a learner not only to read the definition of a word but also to listen to its pronunciation, although digital dictionaries provide no definitions recorded in spoken sounds as yet. The definitions of words in dictionaries, however, may be too difficult for many learners to understand in speech. To modify the definitions so that the learners could understand them with ease in a listening context, the authors conducted word association research, obtaining associated words from EFL learners that could be cues for other EFL learners to accurately approximate the meaning of a new word. We will explore the effect of these cues on the learners' listening comprehension. Word-association cues vs. dictionary-based cues were presented to subjects in both listening and reading comprehension tests. The results showed that there was no significant difference between the mean scores of dictionary-based definitions and learner-based definitions in the reading test and that there was a significant difference between the mean scores of dictionary-based definitions and learner-based definitions in the listening test. It was concluded that word-association explanations described with words of cues obtained from student-based word association tasks resulted in better comprehension than dictionary-based definitions for EFL learners in listening tasks.

INTRODUCTION

Many digital dictionaries and multimedia language learning materials have been developed. Some of them provide learners with spoken sounds of words and spoken examples. With *Collins Cobuild Students' Dictionary Online* (Cobuild), learners can listen to speech sounds of a word by clicking lip-shape icons. *Merriam-Websters' Collegiate Dictionary* and *Longman Interactive English Dictionary* also contain spoken pronunciation of words. A handy electric dictionary, *Electric Book Player DD-170* (Sony Co. Ltd.) holds around 10,000 words with pronunciation by a native speaker. *Ex-word Ex-2500* (Casio Co.) has the pronunciation of around 6,000 synthesized words. The Internet service (*http://www.babylon.com*) provides freeware with which users can listen to the pronunciation of words. Many CD-ROM language-learning materials with tools to learn words have been developed. *World Friend* (Souiku Co.) provides a program called *Word Box* that helps learners exercise important words and phrases frequently used in conversation. *English Passport* (Nova Co.) contains a dictionary that users can use when they want to look for the meanings of words. *Virtual Homestay 1 to 3* (Mediakite Co.) let users look for the meanings of words and also refer to context by clicking words in sentences. The definitions of the words in this software help some learners as they are elaborate and carefully written. While these digital dictionaries and CD-ROM language-learning materials are useful in providing learners with the pronunciation of words and phrases and some sample sentences, there are no concise spoken definitions of the words and phrases recorded in spoken sounds.

This study investigates the necessity and possibility of constructing a spoken-form corpus of learner-based definitions of words for English as a foreign language (EFL) learners. Various corpora have been constructed in fields such as linguistics, second language acquisition, and speech processing. The construction of text corpus has a long history, to wit the *Brown Corpus*, started in 1960s, followed by the *LOB Corpus* and *London-Lund Corpus*. On the other hand, the construction of speech corpora started fairly recently. In 1990s, the compilation of speech corpora such as *ATR British English Speech Database, LDC* (*Linguistics Data Consortium*) *Database*, and *Voice Type Dictation* began. Owing to the rapid development of computer technology and the development of computer speech recognition and speech syntheses, projects of large-scale speech-corpus

constructions are developing rapidly (Sagisaka et al: 1996, Maekawa: 2000).

Digital English dictionaries based on text corpora have been used in EFL classes. The purported benefits of digital dictionaries in EFL learning are that they require less time to access a definition than text dictionaries (Koga: 1995) and that it is also easy to find sample sentences in digital dictionaries. EFL learners are sometimes recommended to use English-English dictionaries when they encounter an unknown word, as reading English definitions of words helps them to "think in English" and referring sample sentences and guessing the meaning of the word in context is a natural way to understand English. The problem, however, is that reading the English definition of a word in an English-English dictionary may require time, for example, to determine which meanings is appropriate and, that contributes to exhaustion effect for EFL learners. These days, Japanese EFL learners are advised to use both English-English dictionaries and English-Japanese dictionaries (Matsumoto: 2000). It is true that English-English dictionaries have been revised for EFL learners to handle easily, as with the "learners' dictionaries." In listening comprehension classes, however, no digital dictionary that contains spoken-form definitions of words is now available. Due to the development of the multimedia technology, learners can find digital dictionaries which give them spoken sounds of words and phrases, but none which provide them with the spoken definitions, as yet. If the learners were free to access the spoken-form definitions besides definitions in written form, they could acquire English through the medium of spoken English, too. It was assumed, however, that dictionary-based spoken-form definition might prove less effective than related words suggested by other EFL learners. Dictionaries, for example, refer to less known words or to archaic usages, present multiple definitions, and have a new-learner unfriendly shorthand of references. If students could access a corpus of word definitions via speech, we thought that this might be easier for them to understand than the audio dictionary-definitions, and it would be an effective tool for them in improving their listening skills.

In constructing a speech database for EFL word comprehension, we have compiled spoken English learner-based definitions of English words. The learner-based definitions (LBDs) are key words collected in our word association experiment. Of course, to construct the spoken learner-based definition of words, one might record the definitions of the words in the dictionaries as they are. However, English-English dictionary definitions employed orally for the explanation of a new word for Japanese EFL learners

sometimes lack cues for referring to the concept of the word. Rephrased and spoken LBDs based on word association tasks were hypothesized as being easier for learners to understand aurally than spoken definitions cited from English-English dictionaries. Deficits of dictionary definitions seems to be in the attempt to be complete and precise: in a word's various usages past and present, technical and daily, and in collocations and grammatical construction, regionally, etc., a word could be defined ad infinitum. Our goal was the opposite: to define an English word in English as concisely and appropriately as possible. Our cue words were based on both modifications and selected borrowings of several different dictionary definitions and on EFL-student based word association tasks. It was thought that the understandings and explanations of EFL learners who had some knowledge of an English word would serve other native-language learners better than dictionary-based definitions, providing that they were accurate. Consider the subsequently arising definitions of "abacas," in which the dictionary's head noun is "frame" and the learner-based head noun is "calculator," and it seems clear that "dictionary-ese" is, although perhaps accurate, not the best key to the word.

Three experiments were made: the first to elicit LBDs of the target words. These were then modified to accord with selected portions of dictionary definitions. Dictionary-based definitions (DBDs) were abstracted from a variety of several English-English dictionaries, such as the Longman DCE and Chambers ULD. Both LBDs and DBDs were presented in written form to EFL learners in Experiment 2, a reading test and were recorded by a native-American English speaker and presented in spoken form in Experiment 3, a listening test to compare the effectiveness of LBDs in written and spoken form, respectively.

EXPERIMENTAL INVESTIGATION

Method

Subjects

Two hundred native Japanese speaking undergraduate students in a Japanese university participated in the research. They majored in technology. They were divided into five groups labeled G1～G5.

Materials

Ten key words, *abacus, babble, cactus, dabble in, ear, fabulous, gabble, habitation, ice berg* and *jack*, whose difficulty levels were higher than the 5,500 basic words as indicated by *Shogakukan Progressive English-Japanese Dictionary*, were chosen. These ten key words were used to gather LBDs in Experiment 1. Ten LBDs written based on both the associations occurring with the high frequency and the ones with low frequency but intuitive good sense in the researchers' opinion and ten DBDs quoted from dictionaries as are the examples below were used in experiment 2.

LDB of *abacus*:

An old style wooden calculator used in small shops by merchants and in schools by children for doing arithmetic, with small balls moved up and down by the fingers.

DBD of *abacus*:

A frame holding wires on which small balls can be moved, used for teaching children how to count, or especially in eastern countries, for calculating.

The ten LBDs and DBDs were read at moderate speed by a native-American English speaker and used in experiment 3.

Procedures

Three experiments were administered. Word association research collected the explanations that Japanese EFL learners gave for the target words. In Experiment 2, a reading test was administered to compare the reading comprehension of these LBDs and DBDs quoted from the Longman DCE and Chambers ULD. A listening test was administered to compare the listening comprehension of the spoken LBDs and DBDs in Experiment 3.

EXPERIMENT 1: WORD ASSOCIATION

Hypothesis

It was hypothesized that for each target word, EFL learners would associate certain words more frequently than others, and that we could find certain words as the dominant association for each target word.

Method

Forty subjects of G1 were given the ten target words and asked to write down any three

words that they associated with each of them and to write passages that explain each of the words (Appendix A).

Results

These results were collected (Appendix B), and the most frequently-occuring words were found as shown in the example below (figures in parentheses show the frequencies of occurrences. The maximum is 40):

Target word: abacus

Associated words: calculate (23), mathematics (18), school (5), wood (4), number (4), finger (4), computer (3), count (3), merchant (3), square (3), Asia (2), beads (2), China (2), difficult (2), elementary (2), Japan (2), old (2), rectangle (2), shop (2), tradition (2)

Discussion

A variety of words were associated with each of the ten target words, but it was generally clear which of the associated words were most typical, as the example shows. Aichison (1987) proposed four main types of associations found in word association tasks. They were, in order of frequency, co-ordination (e.g. *salt* and *pepper*), collocation (e.g. *salt* and *water*), super-ordination (e.g. *dog* and *animal*), and synonymy (e.g. *hungry* and *starved*). The most frequent associations with the ten target words in this research were, *calculate* with *abacus*, *baby* with *babble*, *desert* with *cactus, interest* with *dabble in*, *rice* with *ear*, *unbelievable* with *fabulous*, *noisy* with *gabble*, *house* with *habitation*, *Titanic* with *ice berg*, and *car* with *jack*. Seven of the associations were collocations, two were synonymy, and one was super-ordination. Aichison (ibid.) and Meara (1980) have both noted the tendencies for adults to give co-ordination associations and for children to give syntactic collocations to words in their native language. The Japanese EFL learners in this research gave syntactic collocations to the target words most often, aligning with the tendency of native-speaking children rather than adults. The same tendency could be found by analyzing all the associations of each key word. For example, with *abacus*, 47 kinds of words were elicited, and among them, two associations, *computer* and *machine* were co-ordinations and one, *tool* was a super-ordination. There was no synonymy, and the rest of the associations were syntactic collocations.

EXPERIMENT 2: READING TEST

Hypothesis

It was hypothesized that comprehensibility of written LBDs and DBDs (Appendix C) would not differ significantly in the reading test.

Method

Forty subjects of G2 and G3 were each provided with five written LBDs and five written DBDs for the ten target words, divided so that G2 had the LBDs for G3's DBDs and visa versa. G2 and G3 both received answer sheets (Appendix D) and were asked to connect the target word with the definition.

Results

One point was given for each correct response. The number of the subjects in each group was 40, and the number of stimuli was 10; the maximum score for LBDs and DBDs were both 400. Table 1 shows the score for each LBD and DBD.

Table 1 Score for Each LBD and DBD in Reading Test

Key word	LBD	DBD	Sum
abacus	39	34	73
babble	30	23	53
cactus	39	37	76
dabble in	39	32	71
ear	40	39	79
fabulous	37	32	69
gabble	23	36	59
habitation	40	40	80
ice berg	40	40	80
jack	39	39	78
Sum	366	352	718

Table 2 shows the mean score and standard deviation for the LBDs and DBDs.

Table 2　Mean and SD for the LBDs and DBDs in Reading Test

Type	n	M	SD
LBD	40	9.15	1.02
DBD	40	8.80	1.36

The mean score for the LBDs was 9.15, and it was a little higher than the mean score for the DBDs, 8.80, but the statistical analysis showed no significant difference, $F(1, 78) = 1.42$.

Discussion

The score of LBD of *gabble* was lower than its counter part, three LBDs of *habitation*, *ice berg* and *jack* showed the same score with their counter parts, and the scores of six LBDs were higher than their counter parts. In total, there was no significant difference between the mean scores of the LBDs and the DBDs.

EXPERIMENT 3: LISTENING TEST

Hypothesis

It was hypothesized that the comprehensibility of the LBDs would be higher than that of the DBDs in the listening test.

Method

The forty subjects of G4 and G5 each received an answer sheet (Appendix E) including the ten target words. Each group listened to five LBDs and five DBDs for the ten words, divided so that G4's LBDs applied to the words for which G5 heard DBDs, and visa versa.

Results

One point was given for each correct response. The number of the subjects in each group was 40, and the number of stimuli was 10; the maximum score for LBDs and DBDs were both 400. Table 3 shows the score for each LBD and DBD.

Table 3 Score for Each LBD and DBD in Listening Test

Key word	LBD	DBD	Sum
abacus	36	30	66
babble	27	14	41
cactus	33	22	55
dabble in	28	13	41
ear	35	36	71
fabulous	32	14	46
gabble	18	22	40
habitation	39	38	77
ice berg	39	34	73
jack	33	31	64
Sum	320	254	574

Table 4 shows the mean score and SD for the LBDs and the DBDs.

Table 4 Mean and SD for the LBD and DBD in Listening Test

Type	n	M	SD
LBD	40	8.00	1.82
DBD	40	6.35	1.80

The mean score for the LBDs was 8.00, and it was higher than the mean score, 6.35 for the DBDs. Statistical analysis showed a significant difference, $F(1, 78) = 16.3^{**}$, $p < 0.01$.

Discussion

The scores of two LBDs, those of *ear* and *gabble* were lower than their counter parts, and the scores of eight LBDs were higher than their counter parts. In total, incidence of identification of LBDs was statistically higher than that of DBDs.

SUMMARY AND CONCLUSION

The mean score of the LBDs in the listening test was higher than that of the DBDs. LBDs written based on learner-based word association were easier for EFL learners to

understand aurally than the DBDs cited from English dictionaries, although among ten LBDs of this study, the raw scores of *ear* and *gabble* were lower than those of DBDs. As for the LBD of *gabble*, it showed a lower score than its DBD in the reading test, too. The mean score of LBDs was higher than that of DBDs in the listening test, but not in the reading test. It is supposed that the cue words in the LBDs are especially effective in the listening test, in which EFL learners have to process the language according to the flow of speech. In this situation, the key learner-based associations, which are supposed to lead them to understand the sentence, work effectively.

Although the LBDs written based on the word association task were easier for the learners to comprehend in the listening test, there remain several points to be clarified in future studies. The DBDs might have been supposed to be more accurate than LBDs when they are in written form. Additionally, some of the associations of the learners could not be used as definitions because they are not accurate enough to describe the word, although they may have worked in providing added cues. Some learner associations, as well, work negatively (Ishinou: 1998). Some of the associations that become effective cues for some learners may not be so for other learners. Furthermore, even when we achieve accurate explanations using learners' associations, they are not always easier for learners to comprehend aurally than spoken DBDs. There is also the consideration that words have various meanings and it seems insufficient to choose only one meaning and explain it. There was also a problem in the association test when learners were not familiar with a word and had to be provided with a Japanese equivalent, as this might result in inaccurate associations because of differences in the meanings of the English and Japanese dictionary equivalents. It might be better to have the learners study the meaning of the words before the experiment and present only English words in the experiment.

The question whether it is really necessary to construct a corpus of spoken definitions of words will have to be answered in future studies. These experiments did not make clear the need for spoken materials for TESL lexicon. It is still controversial whether learners are better off using English-English dictionaries as a reading reference. It is claimed that reading definitions helps the learners to think in English and that problems do not arise of a Japanese "equivalent" having a slightly different meaning. Although definitions in dictionaries are reliable, in EFL class activities in Japan, except for rare cases, English-English dictionaries are not used very often. The reason is that the learners do not think they can understand the definitions well, and technical words, such

as the names of the animals and plants are easier to reference in an English-Japanese dictionary, as Matsumoto (2000) points out. In TEFL context, however, English-English dictionaries serve not only for comprehension of word meaning, but also help learners to improve production skills. With the spoken LBDs, the learners may not only improve their listening skills but also come to know how to explain things in their own, sometimes basic English. This concept goes very well with that of Basic English(1) (Gotou: 1997). Shirakawa (2000) points out that many good CD-ROM learning materials are published and soon go out of fashion, even if they are very effective. Theoretical and data-based supports are necessary for the development of multimedia language learning materials. The points that were not clarified in the present study, how the collected associations help learners to comprehend the LBDs and whether using spoken LBDs really helps to improve production skills, need to be addressed in future studies.

NOTES

(1) Basic English is the abbreviation of British American Scientific International Commercial English.

REFERENCES

Aichison, J. (1978) *Words in the Mind*: *an introduction to the mental lexicon*. Oxford: Basil Blackwell Ltd.

Cambell, N. (1993) The ATR British English speech database. *ATR Technical Report*. Kyoto: ATR Interpreting Telephony Research Laboratories.

Gotou, H. (1997) *English Made Simple*: *Basic English*. Tokyo: Shouhakusya.

Ishiou, A. (1998) *Cognitive Psychological Study on Stroop Interference*. Tokyo: Kazama Press.

Kita, K. et al. (1996) *Speech Processing——corpus based approach*. Tokyo: Morikita Press.

Koga, T. (1995) Effect of digital dictionary on foreign language reading process. *Proceedings of 11th Japan Education Technology*.

Makino, Y. (1997) *English Vocabulary Acquisition Theory*. Tokyo: Kagensha.

Maekawa, K. et al. (2000) Preface in trends in database for phonetic research. *Journal of the Phonetic Society of Japan*, 4:2, 3-61.

Matsumoto, S. (2000) Handy electric dictionary. *The English Teachers' Magazine*, 48:15, 60-61.

Meara, P. (1980) Vocabulary acquisition: a neglected aspect of language learning. *Language Teaching and Linguistics*, 13, 221-246.

Sagisaka, Y. et al. (1996) *Computing Prosody*. New York: Springer.

Shirakawa, S. (2000a) Development of CALL series for advanced learners of university students. *The English Teachers' Magazine*, 49:5, 58-59.

Yanagi, Y. (2000a) CD-ROM World Friends, The English passport. *The English Teachers' Magazine*, 48: 14, 60-61.

Yanagi, Y. (2000b) CD-ROM Virtual Homestay, Sumaset Heights. *The English Teachers' Magazine*, 49:3, 60-61.

APPENDIX A

Answer sheet for word association task

Name Number

Ten key words are listed in the following pages. You can find Japanese meanings in the parentheses. Put the words that you associate with these key words in English. And try to explain the key words in English as clearly and plainly as possible without using dictionary.

abacus （そろばん）

Associated words
（ ）（ ）（ ）
Learner-based definition sentence

・
・
・

APPENDIX B

Associated words with ten key words

The number of the kinds of the associations is in the bracket. The more than two times associated words are listed with their frequency in the parenthesis.

abacus [47]

calculate (23)
finger (4)
count (3)
Asia (2)
difficult (2)
old (2)
tradition (2)
mathematics (18)
wood (4)
merchant (3)
beads (2)
elementary (2)
rectangle (2)
school (5)
computer (3)
square (3)
China (2)
Japan (2)
shop (2)

babble [60]

baby (25)
speak (5)
childish (3)
express (2)
mouth (2)
voice (2)
foreign (5)
sound (4)
communication (3)
infant (2)
pretty (2)
number (5)
word (4)
cute (3)
language (2)
understand (2)

cactus [45]

desert (16)
dry (7)
water (5)
needle (3)
spine (3)
rose (2)
thorn (10)
green (7)
strong (4)
pain (3)
America (2)
warm (2)
plant (9)
Mexico (6)
hurt (3)
sand (3)
flower (2)

dabble in [57]

interest (24)
hobby (5)
little (3)
challenge (2)
short (2)
try (10)
gamble (3)
study (3)
give up (2)
test (2)
curious (7)
learn (3)
amateur (2)
serious (2)

ear [44]

rice (24) golden (8) fall (7)
wheat (7) corn (6) farm (6)
autumn (5) bread (5) harvest (3)
hood (3) agriculture (2) field (2)
fruit (2) grain (2) wave of rice (2)
yellow (2)

fabulous [60]

unbelievable (17) surprise (8) huge (5)
money (5) unreasonable (5) expensive (4)
amount (3) excessive (3) great (3)
large (3) unexpected (3) very (3)
big (2) extraordinary (2) governor (2)
illegal (2) law (2) many (2)
price (2)

gabble [75]

noisy (11) chat (10) talk (8)
irritating (4) speak (4) foreigner (3)
woman (3) communication (2) crazy (2)
fast (2) ladies (2) laugh (2)
meaningless (2)

habitation [39]

house (22) live (12) home (9)
life (9) family (8) comfortable (5)
residence (5) apartment (4) address (3)
space (3) stay (3) building (2)
human (2) immigrant (2)

ice berg [43]

Titanic (11) mountain (7) big (6)

penguin (6)	sea (6)	float (5)
cold (4)	North pole (4)	Antarctic (3)
Arctic (3)	freezing (3)	South pole (3)
bear (2)	cool (2)	dangerous (2)
huge (2)	ocean (2)	sharp (2)
solid (2)	terrible (2)	

jack [45]

car(25)	tool (11)	lift (8)
tire (7)	heavy (6)	up (5)
power (4)	iron (3)	fix (2)
flat tire (2)	trouble (2)	useful (2)

APPENDIX C

Learner-based definition (L) and Dictionary-based definition (D)

abacus

L: An old style wooden calculator used in small shops by merchants and in schools by children for doing arithmetic, with small balls moved up and down by the fingers.

D: A frame holding wires on which small balls can be moved, used for teaching children how to count, or especially in eastern countries, for calculating.

babble

L: To say something or make sounds that can not be understood very well, just like new born babies starting to baby talk.

D: To talk quickly and foolishly or in a way that is hard to understand or to make continuous sounds like a baby learning to speak.

cactus

L: Typical desert plants with many sharp protective thorns which grow in a dry areas, such as Mexico.

D: Any of a number of desert plants protected by sharp prickles, with thick fleshy stems and leaves.

dabble in

L: To pursue a study or hobby etc. superficially or from time to time, perhaps as a secondary interest.

D: To work at or study something with some interest but without serious intentions.

ear

L: The head of rice, corn or wheat whose color changes from green to gold in autumn.

D: The head of a grain-producing plant such as corn or wheat, used for food.

fabulous

L: Unbelievable, surprising or huge, for example, talking about a large amount of money.

D: Nearly unbelievable, or very good, pleasant and excellent.

gabble

L: To chat very fast, sometimes in a noisy and irritating way.

D: To say words so quickly that they cannot be heard clearly.

habitation

L: A house, home, apartment or place to live in with family.

D: A house or place to live in.

ice berg

L: A big piece of ice floating in the cold sea, into which the ship Titanic crashed.

D: A large piece of ice floating in the sea, most of which is below the surface.

jack

L: A tool to lift a car up when we put on the spare tire.

D: An apparatus for lifting off the ground anything of heavy weight, such as a car.

APPENDIX D

Answer sheet for reading test

Name　　Number

Match the words with the following sentences and put (a)～(j) in the blanks.

(a) abacus
(b) babble
(c) cactus
(d) dabble in
(e) ear
(f) fabulous
(g) gabble
(h) habitation
(i) ice berg

(　　) A house, home, apartment or place to live in comfortably, sometimes with family

(　　) A large piece of ice floating in the sea, most of which is below the surface

・
・
・

APPENDIX E

Answer sheet for listening test

Name Number

Listen to the sentence 1～10, which are the learner-based definition of the words and match them with the following choices (a)～(j).

(a) abacus
(b) babble
・
・
・

1. (　　)
2. (　　)
・
・
・

A Corpus for Spoken EFL Lexicon: learner-based definition vs. dictionary-based definition

TOMITA Kaoru
NAKAYAMA Kazuo
Mark ANTHONY

Reading a native language dictionary definition of a lexical item is a typical way of learning a new word in foreign language. Digital dictionaries enable a learner not only to read the definition of a word but also to listen to its pronunciation, although digital dictionaries provide no spoken definitions as yet. The definitions of words in dictionaries, however, may be too difficult for many learners to listen to spoken sounds. To modify the definitions so that the learners can understand them with ease in a listening context, the present authors conducted word association research. We obtained words that could be cues for EFL learners to accurately approximate the meaning of a new word. We compare the effect of word comprehension employing these cues to explanations of the words written originally on the learners' listening comprehension test. Word-association cues vs. dictionary-based cues were presented to subjects in both listening and reading comprehension tests. The results showed that there was no significant difference between the mean scores of dictionary-based definitions and learner-based definitions in the reading test and that there was a significant difference between the mean scores of dictionary-based definitions and learner-based definitions in the listening test. It was concluded that word-association explanations described with words of cues obtained from student-based word association tasks resulted in better comprehension than dictionary-based definitions for EFL learners, especially in listening tasks.