

Second Language Phrases Acquisition:

a Corpus-based Survey

ACQUISITION DES EXPRESSIONS DE LA SECONDE LANGUE: ENQUETE BASEE SUR LE CORPUS

Zhou Xiping¹

Huang Kaisheng²

Abstract: English phrases play an important role in language description and acquisition. The most frequently used English phrases that should take priority of learning can be sorted out mainly by quantitative analysis with the help of modern English corpus. Second language phrases acquisition deserves a considerable research topic in second language vocabulary acquisition.

Key Words: lexical phrases, second language phrases acquisition, corpus

Résumé: Les expressions anglaises jouent un rôle important dans la description et l'acquisition linguistiques. Les expressions anglaises utilisées le plus fréquemment qui donnent la priorité à l'apprentissage peuvent être dégagées principalement par des analyses quantitatives avec l'aide du corpus d'anglais moderne. L'acquisition des expressions anglaises de la seconde langue mérite d'être un sujet de recherches important dans l'acquisition du vocabulaire de la seconde langue.

Mots-Clés: expressions lexicales, acquisition des expressions de la seconde langue, corpus

1. INTRODUCTION

With the research of first language vocabulary acquisition, second language vocabulary acquisition research has emerged since the 1970s. From 1980s to 1990s, many scholars such as Linnaud (1986), McKeow & Curtis (1987), Nation (1990), Nattinger & DeCarrico (1992), Coady & Huckin (1997) published some books specializing in this area (Song, 2002). Recently, some Chinese scholars (Dai, 2000; Lian, 2001; Song, 2002; Liu, 2002) have also made similar research. Among them, Nattinger & DeCarrico (1992), Lian (2001) focused their research on phrases acquisition, that is, the function of lexical phrases and its application in language teaching and learning. This created a new research area to second language vocabulary acquisition.

However, among thousands and millions of English phrases, what should second or foreign language learners (say, English learners in China) priorly learn? There is a widely accepted principle in second or foreign language teaching and learning: the most frequently used language items are dealt with first. This

principle should also be concerned in terms of English phrases acquisition. The most frequently used English phrases can be sorted out mainly by quantitative analysis with the help of modern English corpus.

2. BACKGROUND AND AIM OF THE SURVEY

2.1 Role of phrases in second language acquisition

“Lexical phrase” is a term first put forward and categorized by Becker (1975). Nattinger & DeCarrico (1992: 36-47) further modified its concept. They said, lexical phrases are lexico-grammatical units that occupy a position somewhere between the traditional poles of lexicon and syntax; they are “prefabricated chunks” that consist of more than one word (hence called multiword). They are of 4 categories:

A. Polywords: Polywords are short phrases which functions very much like individual lexical items. They are continuous and allow no variability. They can be

¹ School of Foreign Languages, Zhongnan University of Economics and Law, China.

² School of Foreign Languages, Zhongnan University of Economics and Law, China.

*Received 8 December 2006 ; accepted 10 February 2007

either canonical like “hold your horses, at any rate, a piece of cake” or noncanonical like “as it were, by and large, once and for all”.

B. Institutionalized expressions: They are lexical phrases of sentence level, usually functioning as separate utterances. They are mostly continuous and invariable. They are mostly canonical, for example, “a watched pot never boils, how do you do” etc.

C. Phrasal constraints: They are short-to-medium-length phrases. They are continuous and allow variation of lexical and phrasal categories. They can be both canonical and noncanonical. “a...ago (a day ago, a long time ago, etc.) , ...as well as...(this one as well as that one)” are canonical. An example for noncanonical phrasal constraint is “the -er, the -er (the sooner the better, the busier the happier, etc.)”.

D. Sentence builders: Sentence builders are lexical phrases that provide the framework for whole sentence. They are both continuous and discontinuous. They allow considerable variation of phrasal and clausal elements. “Not only X, but also Y (not only was her mother injured in the accident, but also her father)” is a canonical sentence builder, while “the -er X, the -er Y (the sooner all this work is finished, the sooner we will all be able to go home)” is a noncanonical one. “The -er, the -er” is a phrasal constraint, but when expanded to encompass clausal categories (symbolized by X and Y), it becomes a sentence builder.

Nattinger & Decarrico (1992) then emphasized the important role of lexical phrases in an overall description of language and the pedagogical applications in language acquisition. Ellis (1994) also pointed out that learning lexical phrases proved to be essential in second language acquisition especially at the earlier stage. Generally speaking, the role of lexical phrases in second language acquisition is of the following 5 aspects according to Lian (2001):

1st. A considerable command of lexical phrases helps to improve the speakers’ fluency in conversation;

2nd. It helps second or foreign language learners to have a kind of native-like selection of vocabulary, for first language learners use a large number of prefabricated language (that is, lexical phrases) in certain social contexts.

3rd. Lexical phrases learning may contribute to enhance one’s pragmatic awareness. Formulaic expressions frequently embody the societal knowledge that a given speech community shares and they enable the speaker to say the right thing at the right time in the right places (Ellis, 1994:85).

4th. Lexical phrases approach can balance the structural approach with the communicative one (Nattinger & Decarrico,1992), because “the structural approach accounts for one aspect of competence by concentrating on analysis but does so at the expense of access, whereas the communicative approach

concentrates on access to the relative neglect of analysis (Widdowson, 1989: 132).

5th. Since lexical phrases help to improve learners’ fluency, native-like selection of vocabulary and their pragmatic awareness, their confidence of learning and speaking the target language can be better built up.

6th. From the above definition and classification of lexical phrases, we find that lexical phrases are much wider in concept than the term “phrases” we usually refer to. Most of the time, when we use the term “phrases” we usually refer to noun phrases, verbal phrases, adjective phrases, adverbial phrases, etc. that falls into one or two categories of lexical phrases. This is the narrow sense of lexical phrases. The term used in the survey is of this kind. One reason is that this is the rather traditional understanding of phrases in most of the dictionaries and textbooks. Another is of technical reason: some lexical phrases, like “a...ago, the -er X, the -er Y”, are too difficult to calculate their frequencies in large English corpus. This is one of the limitations of the research.

2.2 Aim of the survey

Since phrases learning plays such a important role in second language acquisition, how to learn phrases effectively is of practical value. Pedagogically, if English learners, especially those who learning English as a foreign language in a non-English speaking country (China for instance) can focus on the most-frequently used phrases, the best return may be achieved especially in the early stage according to a general principle of vocabulary acquisition that will be further discussed later on. The most-frequently used phrases can be sorted out by using modern English corpus. This is exactly the main purpose of our survey.

As Nation and Newton (1997: 238) put, there has been a long tradition of research into what vocabulary will provide the best return for learning. The majority of these pieces of research have been frequency counts, which have provided lists of most frequent and widely used words of a language. They then conclude focusing learners attention on the high-frequency words gives a very good return for learning effort. The best example is found in West’s *General Service List of English Words* (1953): it contains 2000 most frequent head words, which accounts for at least 85% of the words on any page of any book no matter what the subject matter and give an even greater coverage of spoken English. Clearly the 2000 high-frequency words of English should receive attention first because without these it is not possible to use English in any normal way (Nation and Newton,1997: 239). Though the interrelation between vocabulary control and vocabulary acquisition has not theoretically proved yet, vocabulary control has been long applied in second language vocabulary teaching practice (Song, 2002).

Hazenber and Hulstun (1996) argue that, a minimal

receptive second language vocabulary should be defined for non-native university students. An empirical investigation is therefore been carried out. The same can be done to second language phrases acquisition as well: Chinese learners should begin their phrases learning with the most frequently and widely used ones, and a minimal receptive phrases list can be worked out through corpus-based survey. Corpus-based survey of word frequency has been practiced in China, however, quantitative analysis of large amount of English phrases is much more complicated and comparatively very difficult than that of single words (Huang and Zhang, 2001).

3. RESEARCH PROCEDURE

3.1 Phrases collecting

First, define headwords of which most useful phrases are made up. The choice of headwords is mainly based on Longman Dictionary (the 1995 edition) where the most frequently used 2000 English words selected by large corpus are labeled. After that, a considerably large amount of English phrases (altogether 4286 after a careful index) has been collected from the following sources either electronic or in paper form:

A. *COBUILD on CD-ROM*, an authoritative and comprehensive guide to the English language (an electronic dictionary developed by Attica Cybernetics from the COBUILD Database---*Collins Dictionary of the English language, Collins English Grammar, and Collins English Usage*);

B. the contemporary textbooks and syllabuses for junior English course, senior English course and college English course;

C. other 4 usually-referred dictionaries(*Longman Dictionary of Contemporary English, Collins Cobuild Essential English Dictionary, Random House Webster's Dictionary, Oxford Advanced Learners' Dictionary*).

3.2 Frequency sequencing

The most crucial and difficult thing in this survey lies in the occurrence searching of the large amount of phrases. The software we use is *MICROCONCORD* (VERSION1.0, Mike Scott & Tim Johns, Oxford University Press, 1991), which can search for the total occurrence of any given words, either single or string (more than two words). And the corpus we use is *BCET* (*Birmingham Collection of English Texts, 7.3 million words covering both general English and English for science and technology*). Calculating the occurrence of a phrase in a word bank is much more complicated, a lot of handwork have to be done (see Huang & Zhang, 2001, *A report of the research on basic phrases in academy English* for detailed information). For example, the total

occurrence of the phrase "speak of" is sorted out in this way:

speak of	26
speaks of	16
spoke of	23
spoken of	4
speaking of	18

(After lemmatization, the total occurrence of the phrase "speak of" in the corpus of 7.3 million words is 87.)

4. DISCUSSION ON THE RESULT

The occurrence of the 4286 phrases in the corpus varies from 1875 times to zero. That means the high-frequency phrases are widely used while some low-frequency ones seldom appears in real English texts. This fact should be noticed and considered. For instance, the following are 19 English phrases with their occurrence over 1000 times in the 7.3 million corpus, they must be given first priority to English phrases learning and using:

more than	1875
such as	1875
out of	1845
for example	1799
up to	1730
per cent	1651
both ...and...	1628
that is	1591
as well	1552
a few	1508
at least	1468
be about to	1382
rather than	1250
from...to	1237
either...or	1227
so that	1156
according to	1145
as well as	1135
of course	1083

On the other hand, phrases like "try one's best, learn...by heart, lose one's temper, in surprise, etc." are often used by English learners in China only occur once each in the corpus. It reveals that native speakers seldom use them in written English, though we don't know what's the case in spoken English.

5. CONCLUSION

Despite the limitation of the research mentioned in Part 2 and the fact that quantitative analysis of phrases still opens to criticism, two things are for sure: English phrases serve as an essential and important issue in the

description and application of the language; corpus-based analysis of English phrases can provide valuable guidelines in teaching and learning process especially to English acquisition as a second or foreign

language. Second language phrases acquisition deserves a considerable research topic in second language vocabulary acquisition.

REFERENCES

- Coady, J. and Huckin, T. *Second Language Vocabulary Acquisition* [M]. Cambridge University Press. 1997.
- Dai, Manchun. On the research of second language vocabulary acquisition [J]. *Foreign Language Teaching and Research* (2). 2000.
- Ellis, R. *The Study of Second Language Acquisition* [M]. Oxford University Press. 1994.
- Hazenbergs, S and Hulstijn, J.H. 'Defining a minimal receptive second language vocabulary for non-native university students: an empirical investigation'. *Applied Linguistics*, 17(2):145-163. 1996.
- Huang, Kaisheng and Zhang, Shubiao. 'A report of the research on basic phrases in academy English [J]'. *Journal of Huanggang Teachers College* (4). 2001.
- Lian, Jie. 'Role of lexical phrases in second language acquisition' [J]. *Foreign Language World* (4). 2001.
- Liu, Shiming. 'Some issues in second language vocabulary acquisition' [J]. *Journal of Jishou University* (1). 2002.
- Nattinger, J.P. and DeCarrico, J.S. *Lexical Phrases and Language Teaching* [M]. Oxford University Press. 1992.
- Nation, P and Newton, J. 'Teaching vocabulary' [J]. In J.Coady and T.Huckin (Eds.). *Second Language Vocabulary Acquisition* [M]. Cambridge University Press. 1997.
- Song, Gang. 'An review of second language vocabulary acquisition research overseas' [J]. *Language Teaching and Research* (1). 2002.
- Widdowson, H.G. 'Knowledge of language and ability for use' [J]. *Applied Linguistics* (1). 1989.

THE AUTHORS

Zhou Xinping, lecturer, School of Foreign Languages, Zhongnan University of Economics and Law. Wuhan, Hubei, 430073, P. R. China.

E-Mail: joanzhou2@163.com

Huang Kaisheng, associate professor, School of Foreign Languages, Zhongnan University of Economics and Law. Wuhan, Hubei, 430074, P. R. China.

E-Mail: huangdick@hotmail.com