An Innovative English Teaching System Based on Computer Aided Technology and Corpus Management

Wencui Gong
Jilin Institute of Chemical Technology, Jilin, China
153949848@qq.com

Abstract—With the development of modern science and technology, more and more computer technologies have been successfully applied in English teaching. Based on computer aided technology and big data corpus management, this paper improves the traditional teaching method into an innovative teaching mode with a big data corpus as English learning resource. On this basis, a computer multimedia teaching system was set up to realize automatic matching of subtitles and vivid restoration of contexts. The teaching system achieved excellent results in application verification. The research results can promote computer technology in English teaching.

Keywords—Corpus management, English teaching method, innovative thinking, computer technology

1 Introduction

Corpus is a big electronic library with certain capacity established based on certain linguistic principles, with adoption of random sampling method and through collecting continuous language use texts or discourse segments appeared naturally. Applying corpus to English teaching can provide rich context for learners and is good for the language knowledge construction for the learners [1-3] The idea of making use of rich context of corpus to assist learners to accomplish active construction for language knowledge and the idea of teaching design based on constructivism stressing the important function of "context" for meaning construction is consistent. Researches about corpus in recent years have achieved great development, but there is certain technical difficulty in the application of powerful English corpus [4-7]. Therefore, in the mainly applications in abstract linguistic research field, dictionary compilation field as well as propositional work of important examinations, the majority of teachers lack necessary understanding of corpus [8]. This situation produces extremely low application rate of corpus in the field of computer assisted English teaching and its application potentiality has not been well explored [9-12]. There are also few researches on the important application value of corpus in assisting English teaching in educational technology circle.

With the continuous improvement of computer performance and continuous decrease of price as well as the rich English electronic publications and network English

text resources and the popularization of scanner in teaching field, it is possible for individual teachers or small teaching collective to construct corpus for teaching [13-15]. Applying corpus into English teaching can help to change the traditional teaching mode focusing on teacher and assist the application of data driven learning model based on constructivism [16-19].

2 Basic Function of Corpus Index Software

2.1 KWIC index

The basic analysis methods of corpus are full text retrieval and word index. Word index form in most of corpus index software is KWIC index, which is "keyword search with context" [20]. Different corpus index software has different naming for key words and the key word in Concordance 3.2 is called as Headword. The following Figure is the KWIC index (making use of Make Full Concordance function) attained from full text retrieval for 1-6 text books of college English intensive reading with adoption of Concordance 3.2, in which Figure 1 is the alphabetical order of key words and Figure 2 is frequency order of key words.

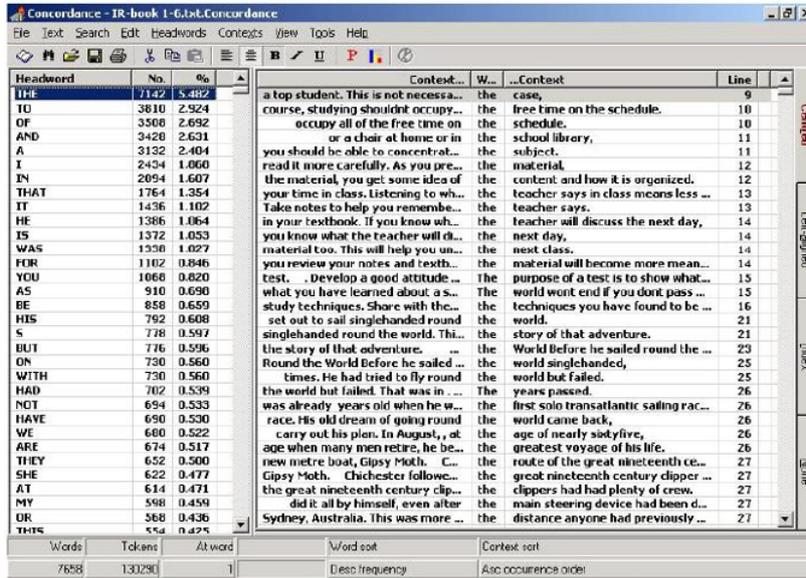**Fig. 1.** According to the keywords of the KWIC alphabetical index

Fig. 2. frequency of keywords KWIC index

2.2 Corpus basic parameters statistics

Concordance 3.2 includes some statistical functions for the basic parameters of corpus. Figure 3 is the statistical results for the basic properties of 1-6 text books of college English intensive reading with adoption of Concordance 3.2:
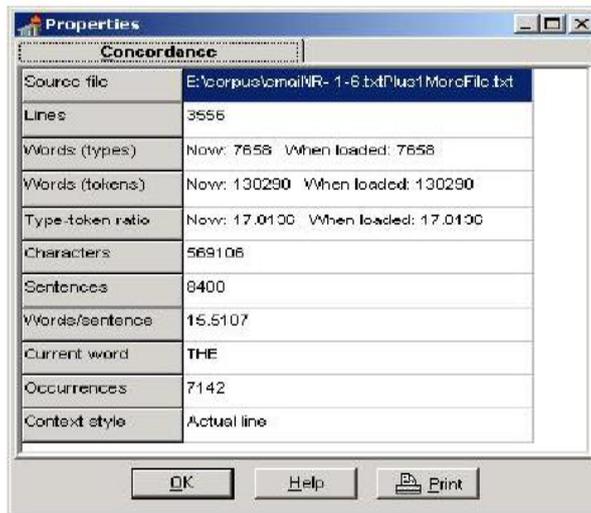

Fig. 3. Corpus properties

In which, it includes types, tokens, type-token ratio, words/sentences and other parameters mentioned in the former introduction of common statistical parameters in corpus.

Make statistics for the word length distribution situation in corpus with word length distribution function, which is with important reference value for judging the difficulty of corpus text and language style [21-24].

Figure 4 is the word length distribution map of 1-6 text books of college English intensive reading with adoption of word length distribution.
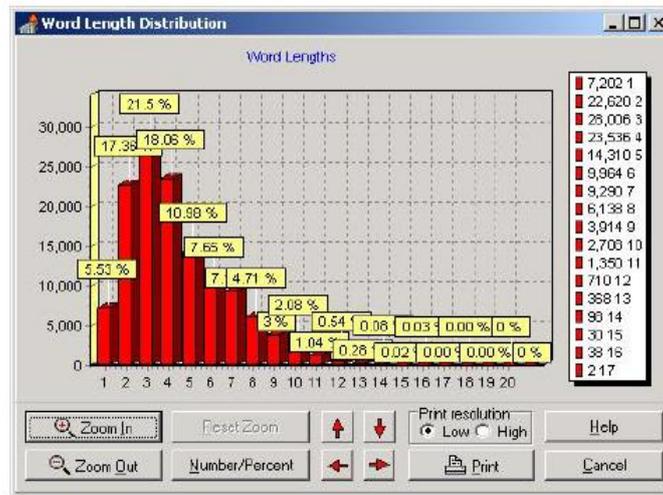

Fig. 4. Word length distribution

It can be seen from Figure 4 that there are 235364 tokens with word length at 4, accounting for 18.06% of the total tokens; there are 39149 tokens with word length at 9, accounting for 9% of the total tokens.

2.3 Collocation statistics of key words

Most of corpus retrieval software provides statistics for the frequency of collocation of key words and other parameters and some also provide the calculation function of collocation force between collocation words and key words [25-27]. Figure 5 is the statistical result for the collocation words of key word able in 1-6 text books of college English intensive reading with adoption of the collocation function of Concordance 3.2 [28-31].

It can be seen from the results that collocation function offers the statistical results of four words of one key word before and after the collocation position [32-34]. It can be seen from the example that the appearing frequency of be at 1-left position and to at 1-right position is obviously higher than other collocation words [35]. After making lemmatization for various forms of verb be, the situation becomes more obvious. It

explains that in 1-6 text books of college English intensive reading, the fixed collocation "be able to" is with the highest appearing frequency among various uses of word able and should be emphasized. This frequency information can't be attained when searching the usage of one word in English dictionary, while if this frequency advantage is very distinctive, it is very necessary for the construction of language knowledge and skills.
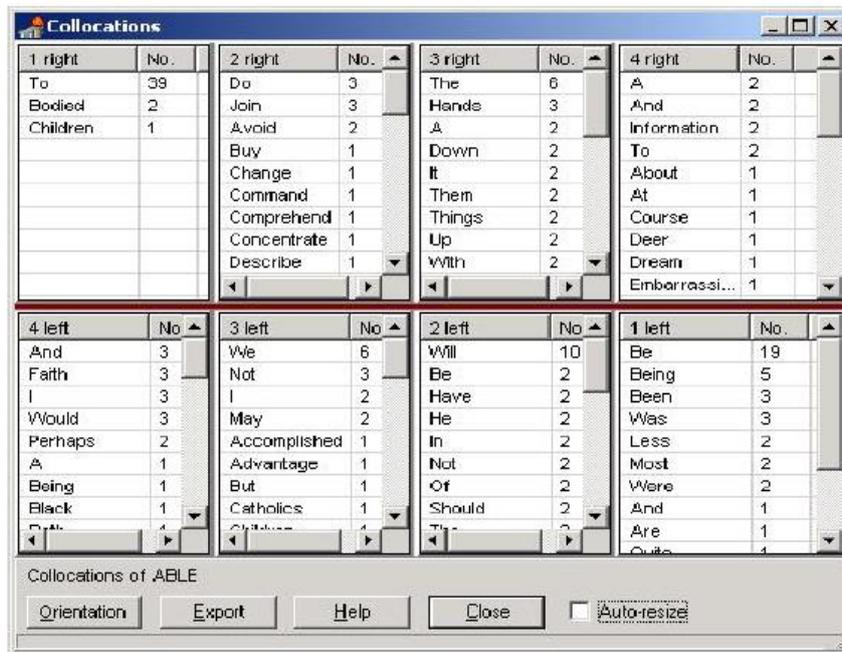
Fig. 5. The statistical results of key words collocation

3   Source of Corpus

3.1   Electronic publications

With the increasingly rich electronic publications, it becomes very convenient and efficient to make use of the text provided by electronic publications as the source of corpus in teaching, especially some English learning materials, English texts and encyclopedia etc. in the form of electronic text [36].

Mp3 voices and corresponding texts are usually provided in English learning materials, such as various kinds of published VOA, BBC news English learning CDs as well as electronic discs of English salons and English abstracts; all of them can extract the text as corpus.

Some special English corpuses provide literature in text form and academic article, which can be used as an important source of corpus. For example, English Classics

1000 published by Fudan University Press provide two searching methods based on works name and author name, covering famous literature and academic works of English history, which is a valuable source of corpus.

English news is in close relationship with current affairs and life, which can timely reflect the changes and developments of contemporary English and represent the characteristics of contemporary English; therefore, it is an important source of corpus for English teaching. English news text can be collected from some famous English news websites and English news learning websites both at home and abroad.

3.2 Film subtitles

Use film subtitles to construct an audio and visual corpus for English learning is a field that has been ignored for a long time. This kind of corpus can be combined with related film and television resources and can be used to improve the English reading ability, translation ability and listening ability [37]. There are usually two ways to attain subtitles [38]; one is to extract dialogue from script text and the other is the extract from the plug-in subtitle file of film and television material.

Subtitles of most film and television materials on current network adopt plug-in form. Plug-in subtitle files focus on srt files, which can be opened directly by notepad as text file and is good to be used as the source of audio and visual corpus. We take the plug-in subtitles of the second part of film the Sound of Music as example to explain the structure of srt file [39-40].

It can be seen from this example that the plug-in subtitle file structure of srt form consists of three parts; the first part is number, which is the subtitles appearance order, as 1, 2, 3, 4 in above example; the second part is the duration time of this subtitle on screen, such as 00:00:15,975 --> 00:00:18,808 in above example, indicating that the following subtitles will be displayed on screen from 00:00:15,975 to 00:00:18,808; the third part is subtitle content, such as "Dear Father, now I know why you sent me here" in above example. This kind of subtitle file in srt form can be found on internet and usually Chinese-English corresponding subtitle can be found. There are many good websites providing subtitle downloading service for free and one of the comprehensive and famous websites is shooter website (http://www.shooter.com.cn).

3.3 Scan written text

At present, scanner can be attained easily for most of teachers. The error rate of transferring English written text into electronic text is very low, which can be accepted completely. When the teachers have English books, newspapers and other written materials suitable to be added into corpus on hand, while corresponding electronic text can't be attained from electronic resources and network resources due to copyright issue, adopting scanner will be the first choice.

4   Application Corpus for Teaching

Collect constructed text corpus for teaching with adoption of corpus index system and some other computer software tools with combination of above methods and many specific functions can be realized. We take Concordance 3.2 as example to explore some specific application mode of corpus for teaching.

4.1   Review vocabulary through KWIC index of intensive reading textbooks

In the former, we have introduced the KWIC index of all types in teaching material through indexing intensive reading textbooks. This can be a good material and tool for learners to review words. Through KWIC index, the learners can realize the construction and strengthening of language knowledge in a better way with the assistance of the specific application environment of each word in teaching material. Produce KWIC index into webpage form to upload to related websites assisting learners, as the tool of assisting learners to construct vocabulary knowledge.

4.2   Vocabulary learning model of learning environment model

Vocabulary learning model of learning environment model based on the constructivism of Jonathan stresses the active construction of learners for vocabulary knowledge under the guidance of teachers. In the process of vocabulary knowledge construction, rich and real example sentence with context for one or some key words is an essential condition. The learners need to make full exploration and consideration for example sentence and context and attain own conclusions. There can be adequate cooperation and discussion among members within the team. Main conclusion of this group can be proposed by taking this group as a unit to mutual reference and discussion among each group. Teachers can offer necessary help and guidance when necessary. It can also be verified in English dictionary. But it should be realized that not all the vocabulary information in dictionary is comprehensive and adequate. The significance of this learning mode lies in that realize active vocabulary meaning construction through exploring a lot of real contexts where vocabulary appears, but not lies in whether the summarization of learner is comprehensive and perfect; most of discussions are open and not taking seeking for standard answer as purpose. The following example is used to explain vocabulary learning model of learning environment model based on the constructivism of Jonathan.

Figure 6 is to guide the students to make usage classification process for round with adoption of all contexts for round (32 in total) extracted from 1-6 text books of college English intensive reading. Figure 7 is deleting key word interest and used to guide students to make exercise of filling in blanks with adoption of all contexts for interest (10 in total) extracted from 1-6 text books of college English intensive reading with adoption of concordance 3.2. It has to point out that the main purpose of exercise of filling in blanks is reading and speculation for context and it is not important whether it can attain the deleted key word. On one hand, the context provided by the teacher is not adequate enough, which can attain the key words easily, under

this situation, it is difficult to get results. But the exploration and discussion of learners within the team for context reaches the purpose of study and communication; on the other hand, the answer of filling in blanks is not the sole and can be open (especially under the situation with less co-occurrence times of context), all the answers can be accepted as long as meeting grammar, logic and context condition. It can be seen that this kind of vocabulary learning model of learning environment model based on the constructivism of Jonathan is on the contrary to traditional indoctrinate vocabulary teaching. Real corpus, open problem and discussion and cooperation among team members are good for improving the interest and enthusiasm of learner. This method also provides a new idea for computer assisted English teaching.
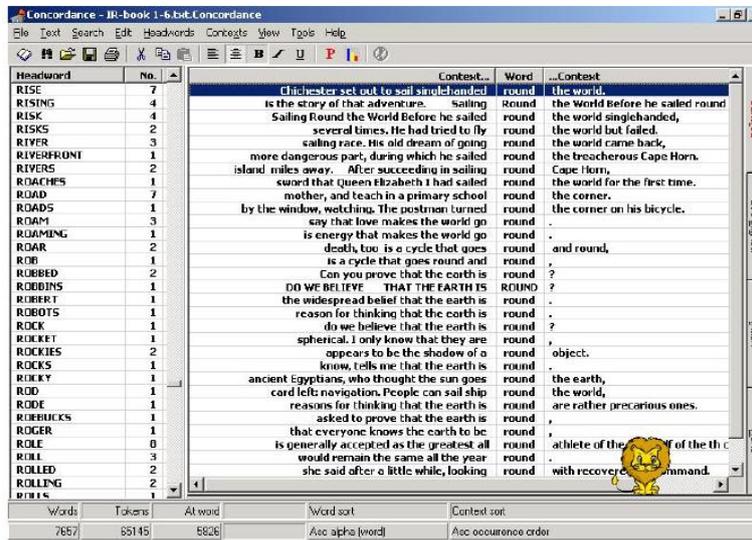

Fig. 6. Round in intensive reading materials corpus in the context

Fig. 7. Interest in intensive reading materials corpus in the context

4.3 Assisted teaching with combination of English film and television information and film and television corpus

Assisted teaching with combination of English film and television information and film and television corpus can create two vivid contexts for learners. One is audio and

visual context composed of film and television information and the second is textual context which is easy to check and compare and composed of film and television subtitle text and corpus. It is in line with that teaching design of constructivism "stressing on the important role of context for meaning construction". Film and television material assisted English teaching needs to possess two conditions. One is English dialogue and the other is English subtitles (it is better to have Chinese subtitles and make it easy for reference). A lot of English films in DVD form provide two kinds of subtitles of Chinese and English. If play the film files in other forms (such as avi or rmvb), it can consider about using MPC (Multi-Player Classic) with plug-in subtitle to reach the purpose of selecting Chinese and English subtitles. It must pay attention that when use plug-in subtitles, plug-in subtitles must have the same name as its corresponding film file and also in the same file.

5 Conclusion

In a long-term run, corpus is mainly controlled by linguists as well as compliers of major examination proposition groups and dictionary as a powerful weapon for foreign language learning and research. Its potentiality in the field of computer assisted teaching has not been well explored and the popularity degree is very low. There are active researches on computer assisted teaching field, which is with wide application and weak disciplinary uniqueness in education technology circle, while neglect the researches that corpus assisted English teaching has great significance in certain discipline teaching field.

This paper introduces related knowledge and theory of corpus, discusses the theoretical basis of computer assisted English teaching based on corpus and makes preliminary discussion on the classification of teaching corpus as well as the application method of corpus with different contents in English teaching aroused by classification thought. It makes pioneering attempt in adopting film subtitle file to extract corpus for assisted teaching and constructs one film corpus library containing over 110 films with English subtitle and with over 1.1 million words. It is hoped to play the role of trigger and arouse the interest and emphasis of education technology circle for corpus assisted English teaching with great potentiality.

6 Acknowledgement

7 References

[1] Swann, P. (1992). Computer assisted language learning for English as a foreign language. Computers & Education, 19(3): 251-266. https://doi.org/10.1016/0360-1315(92)90119-P

[2] Lan, Y.J., Sung, Y.T., Chang, K.E. (2009). Let us read together: development and evaluation of a computer-assisted reciprocal early English reading system. Computers & Education, 53(4): 1188-1198. https://doi.org/10.1016/j.compedu.2009.06.002

[3] Khan, E.H. (1989). The use of computer-assisted learning in a primary school. Computers & Education, 13(4): 355-362. https://doi.org/10.1016/0360-1315(89)90051-1

[4] Solano-Flores, G., Raymond, B., Schneider, S.A., Timms, M. (1999). Management of scoring sessions in alternative assessment: the computer-assisted scoring approach. Computers & Education, 33(1): 47-63. https://doi.org/10.1016/s0360-1315(99)00018-4

[5] Liu, J., Liu, C. (2014). A rational reflection on the application of computer assisted instruction in teaching. International Journal of Technology Management, 250-253.

[6] Murray, O. (2010). Teaching decision making to medical undergraduates by computer-assisted learning. Medical Education, 11(4): 262-264. https://doi.org/10.1111/j.1365-2923.1977.tb00606.x

[7] Burnett, J.D., Miller, L. (1984). Computer-assisted learning and reading: developing the product or fostering the process? Computers & Education, 8(1): 145-150. https://doi.org/10.1016/0360-1315(84)90064-2

[8] Ballou, W.E., Weitze, L.C. (1984). Special all-computer issue computer assisted bio review. The American Biology Teacher, 46(3): 183-187.

[9] Croft, A.C., Danson, M., Dawson, B.R., Ward, J.P. (2001). Experiences of using computer assisted assessment in engineering mathematics. Computers & Education, 37(1): 53-66. https://doi.org/10.1016/s0360-1315(01)00034-3

[10] Clayden, G.S., Wilson, B. (2010). Computer-assisted learning in medical education. Medical Journal of Australia, 22(5): 455-467. https://doi.org/10.1111/j.1365-2923.1988.tb00783.x

[11] Thoennessen, M., Harrison, M.J. (1996). Computer-assisted assignments in a large physics class. Computers & Education, 27(2): 141-147. https://doi.org/10.1016/0360-1315(96)00003-6

[12] Kish, G., Cook, S.A., Kis, G. (2013). Computer-assisted learning in anatomy at the International Medical School in Debrecen, Hungary: A preliminary report. Anatomical Sciences Education, 6(1): 42-47. https://doi.org/10.1002/ase.1303

[13] Al-Jasmi, F., Moldovan, L., Clarke, J.T.R. (2010). Hunter disease eClinic: interactive, computer-assisted, problem-based approach to independent learning about a rare genetic disease. BMC Medical Education, 10(1): 72-82. https://doi.org/10.1186/1472-6920-10-72

[14] Tsai, S.C. (2001). A computer-assisted tutorial on protein structure. Journal of Chemical Education, 78(6): 837-840. https://doi.org/10.1021/ed078p837

[15] Wilson, E. (1994). Using corpora as a resource in language teaching. Computers & Education, 23(1-2): 41-51. https://doi.org/10.1016/0360-1315(94)90031-0

[16] Shannon, J.D. (2010). Small-group interactive computer-assisted teaching. Medical Education, 24(2): 148-150. https://doi.org/10.1111/j.1365-2923.1990.tb02513.x

[17] Kashy, E., Gaff, S.J., Pawley, N.H., Stretch, W.L., Tsai, Y. (1995). Conceptual questions in computer-assisted assignments. American Journal of Physics, 63(11): 1000-1005. https://doi.org/10.1119/1.18044

[18] Hulsman, R.L., Ros, W.J.G., Winnubst, J.A.M., Bensing, J.M. (2002). The effectiveness of a computer-assisted instruction programmed on communication skills of medical specialists in oncology. Medical Education, 36(2): 125-134. https://doi.org/10.1046/j.1365-2923.2002.01074.x

[19] Jacobson, A.R., Militello, R., Baveye, P.C. (2009). Development of computer-assisted virtual field trips to support multidisciplinary learning. Computers & Education, 52(3): 571-580. https://doi.org/10.1016/j.compedu.2008.11.007

[20] Vogel, M., Wood, D.F. (2010). Love it or hate it? medical students' attitudes to computer-assisted learning. Medical Education, 36(3): 214-215. https://doi.org/10.1046/j.1365-2923.2002.01181.x

[21] Boyd, G., Keller, A., Kenner, R. (1982). Remedial and second language English teaching using computer assisted learning. Computers & Education, 6(1): 105-112. https://doi.org/10.1016/0360-1315(82)90019-7

[22] Schreck, R. (1978). An examination of selected procedures for assessing the efficacy of computer-assisted instruction in the teaching of English as a second language to adult foreign students. Genetics, 132(1): 179-191. https://doi.org/10.1101/gad.6.9.1810

[23] Stolow, R.D., Joncas, L.J. (1980). Computer-assisted teaching of organic synthesis. Journal of Chemical Education, 57(12): 868-873. https://doi.org/10.1021/ed057p868

[24] Jim, L.K. (1984). A computer-assisted instructional approach to teaching applied therapeutics. American Journal of Pharmaceutical Education, 48(1): 20-25. https://doi.org/10.1119/1.2341917

[25] Demaiziere, F. (1982). An experiment in computer assisted learning of English grammar at the university of Paris vii. Computers & Education, 6(1): 121-125. https://doi.org/10.1016/0360-1315(82)90022-7

[26] Diab, H.B., Demashkieh, I. (1991). A computer-aided teaching package for microprocessor systems education. IEEE Transactions on Education, 34(2): 179-183. https://doi.org/10.1109/13.81598

[27] Walsh, R.J., Bohn, R.C. (2010). Computer-assisted instructions: a role in teaching human gross anatomy. Medical Education, 24(6): 499-506. https://doi.org/10.1111/j.1365-2923.1990.tb02665.x

[28] Murray, T.S., Barber, J.H., Cupples, R.W., Hannay, D.R., Scott, D.B. (1976). Computer-assisted learning in undergraduate medical teaching. The Lancet, 307(7957): 474-476. https://doi.org/10.1016/S0140-6736(76)91489-6

[29] Harkin, P.J.R., Dixon, M.F., Reid, W.A., Bird, C.C. (1986). Computer assisted learning systems in pathology teaching. Medical Teacher, 8(1): 27-34. https://doi.org/10.3109/01421598609036842

[30] Kenny, G.N., Schmulian, C. (1979). Computer-assisted learning in the teaching of anaesthesia. Anaesthesia, 34(2): 159-162. https://doi.org/10.1111/j.1365-2044.1979.tb06271.x

[31] Brennan, K.F. (1992). A computer-aided approach to teaching band structure formation. IEEE Transactions on Education, 35(1): 60-68. https://doi.org/10.1109/13.123419

[32] Balman, T. (1981). Computer assisted teaching of fortran. Computers & Education, 5(2): 111-123. https://doi.org/10.1016/0360-1315(81)90020-8

[33] Khan, E.H. (1989). The use of computer-assisted learning in a primary school. Computers & Education, 13(4): 355-362. https://doi.org/10.1016/0360-1315(89)90051-1

[34] Smith, S.G. (1971). Computer-aided teaching of organic synthesis. Journal of Chemical Education, 48(11): 727-729. https://doi.org/10.1021/ed048p727

[35] Wharton, D.A. (1990). The use of computer-aided learning for teaching parasite life cycles. International Journal for Parasitology, 20(5): 709-712. https://doi.org/10.1016/0020-7519(90)90136-B

[36] Collart, M.E. (1973). Computer assisted instruction and the teaching-learning process. Nursing Outlook, 21(8): 527-32.

[37] Miller, K., Willett, J.E. (1987). Audio-enhanced computer-assisted learning in science teaching. Biochemical Society Transactions, 15(3): 351-354. https://doi.org/10.1042/bst0150351

[38] Castro, I.P. (1990). Computer aided teaching of digital signal processing. Computers & Education, 14(5): 433-443. https://doi.org/10.1016/0360-1315(90)90037-8

[39] Muñoz-Abella, B., Álvarez-Caldas, C., Rubio, L. (2011). Computer-aided tool for teaching mechanical clutch systems design. Computer Applications in Engineering Education, 19(3): 493-500. https://doi.org/10.1002/cae.20329

[40] Blandford, A.E., Smith, P.R. (1986). The evolution of computer-based learning software design: computer assisted teaching unit experience. Computers & Education, 10(2): 251-257. https://doi.org/10.1016/0360-1315(86)90025-4

8 Author

Wencui Gong received the B.A. Degree in English teaching from Beihua University, Jilin, China, in 2006; the M.A. degree in applied linguistics from Beihua University, Jilin, China in 2011. She has been on the faculty of Jilin Institute of Chemical Technology. Her research interests include English language teaching and linguistics.