

## SpeechIndexer in the Foreign Language Curriculum

Jozsef SZAKOS

Department of Chinese and Bilingual Studies

The Hong Kong Polytechnic University

Hung Hom, Kowloon, Hong Kong

jozsef.szakos@inet.polyu.edu.hk

Ulrike GLAVITSCH

EMPA, Swiss Federal Laboratories for Materials Science and Technology

Media Technology Laboratory

Dübendorf, Switzerland

ulrike.glavitsch@empa.ch

The software SpeechIndexer allows for the synchronization of text with recorded speech, it is used for the documentation of endangered languages and serves as a language learning and teaching tool (Szakos & Glavitsch, 2004; Glavitsch, Simon & Szakos, 2011). Corpus-based instruction became popular in EFL classrooms (Frankenberg-Garcia, 2005). However, it takes a longer time to implement it in Chinese as a Foreign Language (CFL) teaching. Despite all the optimism of the previous years, teachers of Chinese rarely base their teaching and textbooks on corpus, even less on speech corpora. In our paper, we intend to provide our considerations on the integration of speech corpora in long-term foreign language instruction by using SpeechIndexer. We will introduce a step-by-step approach of distributing the tasks of corpus-based teaching/learning over one or several years of curriculum that leads teachers onto the corpus track and contributes to the autonomy of language learners.

The rising popularity of Chinese, combined with its unique difficulties for western learners (pronunciation, reading-writing, segmentation problems) provides a suitable testing ground for our proposals. Taking *New Practical Chinese Reader (NPCR)*, a six-volume teaching work as an example, we demonstrate how corpus-based instruction is woven into the intermediate and advanced levels of teaching (Liu, 2006). The standard work is published in English, German, Spanish, French and Russian versions, used in most of the classes (except for Taiwan). The first four volumes target beginners, while the last two are geared for intermediate learners.

The intermediate stage of learning a new language, which is critical for learner perseverance, can be most helped by corpus, especially by speech corpus methods. The curriculum consists of the original teaching materials expanded through supplementary recordings and transcriptions to broaden the learner's experience with the authentic language. Completing this stage may take at least a year for the learner. Teachers can select speech recordings from the vast amount of materials available (audio books, radio and TV podcasts, etc. ) that specifically match the knowledge level and interest of individual students or the whole class. The individual features of SpeechIndexer are to be gradually introduced in class and applied to the supplementary materials. The suggested order of introducing the features: (1) the player function trains the listening comprehension on marked speech recordings where the learner can follow the currently spoken text, (2) with the role play function students can train their speech production in the context of dialogues, (3) the built-in speech concordancer is used for contextual language learning, and (4) the transcribe function serves as a feedback on learning success (Glavitsch, Simon & Szakos, 2011).

In the combined process the learner gradually gains control of the methods and materials. This will be illustrated by various examples such as BBC radio podcasts, portions of audio books and dialogues from Chinese (NPCR) and English courses.

The novelty of our present paper is that SpeechIndexer increasingly becomes part of the curriculum, and the involvement takes place over a whole year, so the learner is not hampered and put off by suddenly surfacing software complexities (Farr, 2008). By substituting diverse languages in this teaching/learning method we hope to remove some obstacles in learning. Previously, a sudden and full implementation of SpeechIndexer in teaching proved impossible, in spite of clear understanding of all its benefits by teachers and students.

We hope to demonstrate that a gradual and graded integration of speech corpus methods will have a better chance in the classroom, achieving learner autonomy and independence.

### **References**

- Farr, F. (2008). 'Evaluating the Use of Corpus-based Instruction in a Language Teacher Education Context: Perspectives from the Users.' *Language Awareness* 17: 25-43.
- Frankenberg-Garcia, A. (2005). 'Pedagogical uses of monolingual and parallel concordances.' *ELT Journal* 59: 189-198.
- Glavitsch, U., Simon, K. & Szakos, J. (2011). 'SpeechIndexer: A Flexible Software for Audio-Visual Language Learning.' *Proceedings of the International Conference on Education, Informatics, and Cybernetics*. Orlando, Florida. 79-82.
- Liu, X. (2006). *New Practical Chinese Reader*. Beijing: Beijing Language & Culture University Press.
- Szakos, J. & Glavitsch, U. (2004). 'Seamless Speech Indexing and Retrieval: Developing a New Technology for the Documentation and Teaching of Endangered Formosan Aboriginal Languages.' *Proceedings of the International Conference on Education and Information Systems: Technologies and Applications (EISTA'04)*. Orlando, Florida. 88-93.