

Abs-57

Johannes Widmann (University of Tübingen)

Lexical clusters in small pedagogic corpora

Pedagogic corpora as they have been conceived in the SACODEYL and BACKBONE ([www.uni-tuebingen.de/backbone](http://www.uni-tuebingen.de/backbone)) projects are an important step in bringing corpora closer to use in the communicative language classroom. Each corpus consists of interviews where the interviewees talk about pre-defined topics. With their focus on pedagogic purposes, the corpora provide teachers with a didactic infrastructure that harmonizes better with their teaching practice than it seems to be the case with typical large representative corpora (see Braun 2005 for an overview of the pedagogic concept behind the corpora).

In this talk I will look at the lexical potential of these corpora. Because of their carefully selected contents they allow for lexical tasks that go beyond classic data-driven learning. What is needed, however, are semi-automatic procedures for making the lexical potential of small corpora explicit to users and for integrating this lexical potential in a topic-based language learning approach. In order to find an answer to these two issues, I will present an analysis of the lexical potential of the English SACODEYL and BACKBONE corpora. The focus will be on various types of clusters, such as polywords, fixed expressions and collocations which are the key elements in a lexical approach as defined by Lewis (1993, 1997). I will extract such clusters using available corpus tools and methodologies. I will then look at the overall results across all interviews in order to analyze and describe the corpora with regard to their lexical cluster characteristics and I will discuss how much these automatic results are usable in language teaching with respect to a topic-based teaching approach.

The results will then be sorted per interview topic and I will compare the topics according to their lexical potential. Based on this lexical potential I will attempt to establish a pedagogical classification and sequencing of the topics. This classification will move from a description of the clusters to an interpretation of what the clusters are good for with respect to topic-based language learning. My interpretation will take into account a complementary inspection by hand which the automatic analysis has not revealed.

As a conclusion, I will argue that only once we achieve a systematic identification and pedagogical classification of topic-based multi-word units, we can contribute to making teachers aware of the advantages of using corpus techniques. It will be much easier for them to get a multi-faceted overview of the lexis of a text. Many researchers have claimed that such an awareness of corpus techniques is the key element for getting teachers more interested in using corpora (Frankenberg-Garcia 2006). My talk will thus demonstrate how even small corpora can fruitfully be exploited for a topic-based lexical approach.

References:

Braun, S. (2005). "From pedagogically relevant corpora to authentic language learning contents", *ReCALL* 17:1, 47-64.

Frankenberg-Garcia, A. (2006). "Raising teachers' awareness to corpora", keynote speech at TaLC 7, Bibliotheque Nationale de France, 1-4 July 2006:

Lewis, M. (1993). *The Lexical Approach*. Hove: Language Teaching Publications.

Lewis, M. (1997). *Implementing the Lexical Approach*. Hove: Language Teaching Publications.