Developing a TV drama corpus-informed general spoken formulas list for elementary-level EFL learners
Hiroya Tanaka (Hokkai-Gakuen University, Japan)

This study aims to develop a general formulaic sequences list for elementary-level EFL learners informed by an American TV drama corpus. The use of formulaic language is considered to play a significant role in spoken discourse as well as in written discourse (e.g., Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2008) and researchers have developed lists of formulaic language for various multiword units such as phrasal verbs (Garnier & Schmitt, 2015), written and spoken academic formulas (Simpson-Vlach & Ellis, 2010), and non-transparent multiword expressions (Martinez & Schmitt, 2011). There has been, however, little attention paid to date to pedagogical formulaic-sequence materials for informal conversations which could be used for elementary-level EFL learners, although oral communication is largely informal in nature even in workplaces (e.g., Crosling & Ward, 2002).

Davies (2012) claims that a TV drama corpus, the Corpus of American Soap Operas in his case, is at least spoken-like and contains more words dealing with everyday life and personal relationships than spoken proportion of BNC and COCA. Thus, it is fair to say that a TV drama corpus has a potential to inform the use of formulaic sequences in informal conversations. The author has developed a corpus of American TV sitcom scripts to reveal the use of formulaic sequences in informal conversations and to make a pedagogical general spoken formulas list for elementary-level EFL learners. The scripts were collected from various sitcom fansites, which resulted in a five-million word corpus of 18 sitcom titles from 2006 to 2016. Following some of the previous studies (Coulmas, 1979; Wood, 2010; Wray & Namba, 2003), the author created and used a checklist in order to identify spoken formulas in the corpus. The requirements in the checklist are; each formula has to 1) have two or more words (2- to 5-gram), 2) have the frequency of more than 10 per million words, 3) have a pragmatic integrity, 4) be transparent rather than opaque. The first two requirements were checked using a corpus concordance software, CasualConc (Imao, 2016), and the last two requirements were checked manually using learner dictionaries on the Internet. The poster will present a tentative list of general formulas informed by the five-million word American sitcom corpus and will discuss further possibilities of the list.

References


