Abs-121

David Wible (National Central University, Taiwan) and Nai-Lung Tsao

The StringNet Lexico-Grammatical Knowledgebase and its LexChecker applications for lexicography and language teaching

This workshop introduces a lexico-grammatical knowledgebase of English called StringNet and a suite of its applications called LexChecker.

A StringNet Navigator web interface supports keyword searches and navigation of the knowledgebase. StringNet (Wible and Tsao 2010) itself consists of hybrid n-grams, which, unlike traditional n-grams, can include part-of-speech (POS) grams. Thus, not only the string consider yourself lucky but also the patterns consider [pron rfix] lucky, consider yourself [adj], [verb] yourself lucky, inter alia. With StringNet Navigator, a click on any POS slot provides a pop-up showing the exact words attested in BNC in that slot in that pattern and their frequency in that slot. Each hybrid n-gram links to all BNC examples of it.

StringNet exploits the POS slots to capture subordinate and super-ordinate relations among and between hybrid n-grams. These are navigated by following [parent] or [child] links beside each hybrid n-gram listed in search results. For example, the two distinct n-grams consider yourself lucky and count yourself lucky are related by a common [parent] [verb] yourself lucky. Conversely, the hybrid n-gram consider yourself [adj] is the common parent of the [child] consider yourself fortunate. This structure represents a dense relational dimension of new, navigable lexical knowledge.

LexChecker applications of StringNet include error detection and correction (Tsao and Wible 2009), also implemented as a query doctor, a proxy for the common, risky practice of using Google searches for English error checking. LexChecker installed as a web-browser toolbar also can determine, for any string of text that a user mouse-selects in a webpage, whether that string is a frozen expression or an instance of a more general pattern and what that pattern is. LexChecker can also actively detect and highlight lexico-grammatical patterns in a webpage, unprompted. All tools and functions to be demonstrated are freely accessible.

References
