Building a CGI interface to WordNet verb data

Steven D. Tripp
Center for Language Research
University of Aizu

Several years ago I created a computer based language learning tool that allowed language learners to set parameters and see sentences created online. This CGI used only about 50 highly frequent verbs. WordNet is a hierarchical digital dictionary arranged as a lexical inheritance system containing about 10,000 verbs. The WordNet database also contains grammatical and lexical information about the verbs that can be extracted and re-purposed. The object of this research is to create a highly expanded Verb Machine and to test this interface with actual learners.

The Verb Machine is an interactive website which uses CGI to generate web pages according to user choices. The interface allows four binary choices that generate 16 verb forms in three time frames. Thus 48 basic patterns are potentially possible. These can be instantiated with 50 verbs and seven pronoun subjects. This generates a potential 16,800 sentences that can be assembled in real time. However, the verb data in WordNet (Fellbaum, 1999) could be added to Verb Machine to provide a more powerful tool. WordNet is a hierarchical lexical database that contains at least 10,000 English verbs. Verbs are linked vertically and horizontally according to meaning and implications. All senses are glossed, grammatical frames are provided, and in many cases example sentences are included. In short there is a very large amount of lexical information embedded in the WordNet database. The problem is how to make this available in a form useful to students.

The large amount of data embedded in WordNet renders the implementation nontrivial. For example, a simple expansion of the existing interface is not practical. While it is reasonable to ask a user to scroll through a list of 50 verbs, it is not plausible that a user will want to look through a list of 10,000 verbs. This raises questions about what kind of search mechanism is appropriate. As a first approximation, there are eight main categories of information to be displayed: Hypernyms, synsets, troponyms, senses, glosses, examples, syntactic frames, and generated sentences. How each category is displayed impinges on the display of the others. Usability literature (Eberts, 1994; Fleming, 1998, Galitz, 1997; Mandel, 1997; Neilson, 1993, 1997; Rosenfeld and Morville, 1998; Rudisill et al., 1996; Spool, 1999; Treu, 1994) suggests that navigation in database interfaces is a prime concern. Previous analysis by myself (Tripp, 2003) indicated that mature interfaces follow these principles: First, provide multiple views of the same data or what we might call analog, digital and hybrid views. A second principle is to provide direct and indirect access to locations. Direct access is provided through cursor movement. Indirect access is provided through menu driven searches. A third principle is to afford information based choices. A final principle is to suppress information under icons until the user decides that information is required. In addition, recent browser windows seem to follow a pattern of putting indexical information in a column on the left and categorical and navigational information horizontally across the top.

We have constructed several prototype interfaces and have been testing these with students. This presentation will demonstrate the use of prototyping as a design strategy (Isensee & Rudd, 1996; Tripp & Bichelmeyer, 1990) and report on the evaluation procedure (Treu, 1994).