

## COMPOUND: A PROGRAM THAT UNDERSTANDS NOUN COMPOUNDS

David R. McDonald\*

Department of Computer Science,  
Carnegie-Mellon University,  
Pittsburgh, Pennsylvania, 15213.

Compound is a program that accepts a noun compound of arbitrary length and attempts to construct a meaning for it. A noun compound is a sequence of two or more nouns used to describe a single referent. Compound was designed to function as a part of a natural language understanding system. Some of the knowledge (e.g., context) necessary to understand some compounds must come from sources that a natural language understanding system could provide. Compound produces meanings for compounds such as "water meter cover adjustment screw", "glass wine glass", "golf club", and "computer maintenance". The first example is fairly straightforward and provides no difficulty for the program. In the second example, the word "glass" has two meanings (i.e., one as a container from which people drink liquids, and the other as a clear material made from silicon dioxide). The context provided by the compound itself is sufficient to disambiguate which meaning of "glass" is meant in both cases. In the "golf club" example, the word "club" has two meanings (i.e., as an organization, and as a stick used in games). The compound "golf club" also has two meanings, one as an organization that provides golfing facilities, and the other as the stick used in the game of golf. The context in which the compound occurs can provide sufficient information to disambiguate it. For example, consider the two sentences:

1. The golf club hit the ball.
2. The golf club fired its manager.

The last compound, "computer maintenance", seems straightforward, as most people assume it describes a "maintenance" action in which a computer is maintained. However, the phrase "computer maintenance of a data base" brings out a second possible meaning.

Compound uses the semantic network system, NFTL, developed by Fahlman (1) as its underlying knowledge representation system. Knowledge about the meanings of words, as well as other real-world knowledge is stored in the NETL network. Compound accesses this network to find all the possible relationships between the nouns of a compound that could be the basis of an interpretation. If no interpretations are found, Compound reports this to the user. If one interpretation is found, it checks with the user before adding it to NETL's data base. If more than one interpretation is found, it invokes several heuristics to attempt to select the best

This research was sponsored in part by the Defense Advanced Projects Agency (DOD). ARPA Order No. 3597, monitored by the Air Force Avionics Laboratory under Contract F33615-78-C1551, and in part by the National Research Council of Canada. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies either expressed or implied, of the Defense Advanced Research Projects Agency, the US Government, or the National Research Council of Canada.

interpretations for the compound. The current heuristics include:

1. Context: the context in which a compound occurs often disambiguates it.
2. Cognates: a meaning for the compound may already exist in NETL's data base.
3. Instances of an Interpretation: instances of a proposed meaning of the compound may exist in NETL's data base.
4. Similar Compounds: there may exist in NETL's data base meanings of compounds that are similar to an interpretation for the compound under consideration. For example, the meanings of "car maintenance" and "truck maintenance" are similar to the meaning of "computer maintenance" where a "computer" is the thing maintained, since all are machines. This heuristic prefers this meaning, rather than the one where a "computer" does the maintenance, as in "computer maintenance of a data base".

5. Occurrences: the number of times an interpretation has occurred gives a weak indication of how good an interpretation is.

Once Compound selects and ranks the interpretations it considers best, it asks the user to make a final selection of the meaning to be added to NETL's data base. The interpretations stored in NETL's data base are used by the similar compounds heuristic to help disambiguate subsequent compounds. Through this mechanism, the program's performance improves over time, a simple but effective form of learning.

Compound is written in Maclisp and runs on a DKC POP 20 under TOPS-20. It takes a few seconds of CPU time to produce meanings for most compounds. It may take more time, if ambiguities exist

## REFERENCE

- [1] Fahlman, S. E., NETL: A System for Representing and Using Real-World Knowledge Cambridge: MIT Press, 1979.