Letter to the Editor

On the Search for Semantic Primitives

Dailey (1986) has transformed the classic problem of the "vicious cycle or circle" into the neoclassical paradigm of "NP-completeness". As presented, there can be no disagreement with his conclusion that the problem of extracting a minimum set of semantic primitives from a monolingual dictionary is NP-complete. However, I believe (1) his use of NP-completeness is inappropriate to the problem of finding semantic primitives in a monolingual dictionary and (2) his notion of how a primitive must emerge from an examination of definitions is not sufficiently perceptive.

Some suggestions for dealing with NP-complete problems include developing algorithms that are fast enough for small problems, that deal with special cases, or that exploit special features of a particular instance of the problem. With a large dictionary, such as *Webster's Third New International Dictionary*, it would seem that we are at a size where NP-completeness comes into play. Even here, it is possible to reduce the problem to a manageable level without invoking any special procedures, using only a correct formulation of the problem.

The definienda and definitions of a monolingual dictionary can be mapped into a graph theoretic structure with nodes corresponding to definienda (as Dailey does), but with links going from a word occurring in a definition to the word being defined (opposite to Dailey's direction). The problem of finding semantic primitives in this model is then equivalent to finding what is called in digraph theory the point basis of a digraph, i.e., those nodes from which all others in the dictionary are reachable. Using straightforward algorithms to do this, the size of the problem can be reduced considerably. (In Litkowski 1978 and 1980, I show how this approach reduced an initial set of 20,000 verbs to 4,000 "more primitive" verbs. Amsler (1980) similarly used many heuristics in analyzing verb definitions in a monolingual pocket dictionary.)

After applying such "gross" techniques for pruning, the problem is more tractable and it is possible to take advantage of special features of the problem at hand. In particular, it is possible to develop many heuristics which take into account many of the characteristics of a dictionary as well as many semantic considerations. These heuristics make it possible to reduce the problem even further, moving ever closer to semantic primitives.

With all that has been said above, I do not want to leave the impression that the problem is in any sense easy. If that were the case, the literature would be replete with claims based on analysis of dictionaries that such and such constitute primitives. Such analysis is filled with complexity and near overwhelming size, but it is *tractable*, even though we may not be able to develop an efficient algorithm to solve the problem once for all time.

In connection with the second point, about the nature of the primitives discoverable from a monolingual dictionary, I believe Dailey has fallen into a trap. As one reduces the digraph of a monolingual dictionary, even with all the heuristics, one is left with the conclusion that vicious cycles will inevitably remain, except for minor parts of speech, such as prepositions and conjunctions which appear to have functional definitions that act as primitives. The presence of a cycle as a primitive is disconcerting; I believe Dailey sees that as proof that the problem of finding semantic primitives is intractable.

I have argued (Litkowski 1978) that one will eventually reach the situation where all that remains (for the significant parts of speech) is a set of vicious cycles, agreeing in this respect with Dailey. However, I interpret these cycles, not as unknowable, but rather as individual concepts where each point in the cycle corresponds to the same concept and is merely labeled differently. I believe that such concepts should ultimately be describable in terms of some functions or procedures (as Schank (1972) does with his primitives). I ascribe the phenomenon present in monolingual dictionaries to the imperfections of lexicographers and to the as yet inscrutable nature of what functionality is captured by such primitive concepts.

Thus, I argue that the search for primitives in a monolingual dictionary is not an empty exercise. Rather, I believe that it can indeed uncover primitives and, in the process, help us to build better lexicons, to make them more consistent, and to make it possible to build a better set of frame constructs as elements for any computational system that relies on natural language.

Kenneth C. Litkowski
Gaithersburg, Md.

References

Amsler, R.E. 1980 The Structure of the Merriam-Webster Pocket Dictionary, TR-164. Department of Computer Sciences, University of Texas at Austin.

Dailey, D.P. 1986 The Extraction of a Minimum Set of Semantic Primitives from a Monolingual Dictionary is NP-Complete. *Computational Linguistics* 12(4): 306–307.

Litkowski, K.C. 1978 Models of the Semantic Structure of Dictionaries. *American Journal of Computational Linguistics* Mf.81: 25–74.

Litkowski, K.C. 1980 Requirements of Text Processing Lexicons. (Unpublished manuscript) Extended abstract in *Proceedings of the 18th Annual Conference of the Association for Computational Linguistics*, Philadelphia, Pennsylvania.

Schank, R.C. 1972 Conceptual Dependency: A Theory of Natural Language Understanding. *Cognitive Psychology* 3(4): 552–631.