A Distributed Representation of Lexical Semantic Information

Ping Li (PING@COGSCI.RICHMOND.EDU)
Beth Engiles (BENGILES@RICHMOND.EDU)
Department of Psychology
University of Richmond
Richmond, VA 23173, USA
Yang Gu (YANGGU@CUHK.EDU.HK)
Department of English
Chinese University of Hong Kong
Shatin, NT, Hong Kong

In this study we propose a new way of representing lexical semantic information in connectionist networks. Previous studies have attempted to derive semantic representations through the analysis of lexical cooccurrence constraints (e.g., Burgess & Lund, 1997; Lund & Burgess, 1996). Although representations derived this way can capture lexical information in significant ways, such representations are not by themselves semantic-explicit in that they do not contain explicit semantic properties or features of the lexical items per se. Li (1993) and Li & MacWhinney (1996) studied representations of verbs by using semantic judgments from human subjects, but their representations apply to only a restricted semantic field (meanings associated with the use of the English prefix *-un*).

To arrive at a representation that consists of realistic semantic properties, we analyzed an electronic dictionary, the Mini-LDOCE (Mini-Longman Dictionary of Contemporary English, Guo, 1995a; 1995b). We extracted out of this dictionary the most frequent words that appear as definitions of English nouns and verbs in the LDOCE. These words are referred to as the "defining words", which are used to represent the "defined words", that is, the target nouns and verbs in the dictionary, through a recursive mechanism described below. The degree of recursion indexes the relevance of the defining word to the defined word.

The recursive mechanism works as follows. We compiled a program to search the entire electronic dictionary for each defining word. If a defining word occurs in the actual definition of a particular noun or verb, the defining word is assigned the highest degree of relevance to that noun or verb. If it does not occur in the definition, the program searches through all other defining words in the definition, and determines if the defining word appears at the next level of relevance, which is part of the definition of a defining word. The search runs recursively for all defining words in all defined words, and it terminates when a given defining word has been assigned the lowest degree of relevance (after eleven sweeps of recursion). The different degrees of relevance is then graded on a eleven-point scale (from 0.0 to 1.0) to derive a matrix of relevance of defining words by defined words. This matrix forms the basis of a connectionist distributed representation of lexical semantic information in English.

Our goal is to construct a descriptively meaningful, semantically grounded input representation in order to circumvent the need of using only randomly generated, schematized input matrices as semantic representations in connectionist networks (Cottrell & Plunkett, 1994, MacWhinney, 1997). By feeding this representation into a connectionist network, along with phonological information, we can model various aspects of language processing and language acquisition with a realistic lexicon.

References

Burgess, C. & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 1-34.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers, 28*, 203-208.

Cottrell, G., & Plunkett, K. (1994). Acquiring the mapping from meanings to sounds. *Connection Science, 6*, 379-412.

Guo, C.-M. (1995a). *Mini LDOCE*. Tsinghua University, Beijing, China.

Guo, C.-M. (1995b). *Machine-tractable dictionaries: Design and construction*. Norwood, NJ: Ablex.

Li, P. (1993). Cryptotypes, form-meaning mappings, and overgeneralizations. In E. V. Clark (ed.), *The Proceedings of the 24th Child Language Research Forum*. Center for the Study of Language and Information, Stanford University, 162-178.

Li, P., & MacWhinney, B. (1996). Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science, 8*, 3-38.

MacWhinney, B. (1997). Lexical connectionism. In P. Broeder & J. Murre. (Eds.), *Cognitive approaches to language learning*. Cambridge, Mass.: MIT Press.