# WN-LEXICAL: An ACT-R module built from the WordNet lexical database

**Bruno Emond(bruno.emond@{uqo.ca,nrc-cnrc.gc.ca})**
Département des Sciences de l'Éducation,
Université du Québec en Outaouais, 283 Alexandre-Taché,
Gatineau, QC, Canada. J8X 3X7

Institute for Information Technology,
National Research Council Canada,
1200 Montreal road, M-50
Ottawa, ON, Canada K1A 0R6

In spite of the application of the ACT-R cognitive architecture to many different domain areas (http://act-r.psy.cmu.edu/), few large-scale cognitive models have been built yet. One reason for this situation is certainly the high cost of building large-scale models, and the focus on using cognitive models to account for experimental data. However, there is a growing interest in building large-scale models, in particular in the area of natural language processing (Ball, Rodgers, & Gluck, 2004).

This paper presents a module for the ACT-R cognitive architecture, which supports building models of over half a million chunks. This module, WN-LEXICAL, is an implementation of the WordNet lexical database (http://wordnet.princeton.edu/). WN-LEXICAL can simply be added to the basic ACT-R architecture to offer some basic functionality as well as all of the WordNet lexical entries and relations encoded as chunks. The module can be used as a basis to build large-scale conceptual or natural language processing models.

At the current stage of development, the WN-LEXICAL module commits to few sub-symbolic computations, and aims to be relatively neutral in terms of architectural assumptions. The purpose of the implementation is to allow exploration and experimentation to establish what should be the module core functionality. In this respect, the current implementation offers some flexibility to the modeler by allowing to load WN-LEXICAL as a separate module, or to load all of its chunks in declarative memory. This feature is in agreement with neuroscience studies (Ullman, & al., 1997; Tyler, Marslen-Wilson, & Stamatakis, 2005).

The paper is divided in two sections. The first section describes briefly the content of the WordNet lexical database, and the second section explains some specific elements of the WN-LEXICAL implementation. A conclusion outlines some future development of the WN-LEXICAL module.

## WordNet

WordNet was developed by the Cognitive Science Laboratory at Princeton University under the direction of George A. Miller (Fellbaum, 1998). One motivation that guided the development of WordNet is its cognitively plausibility. This plausibility is based on three hypotheses. The separability hypothesis states that lexical knowledge is independent from other language related knowledge. The patterning hypothesis states that relations and patterns between lexical entities are central to natural language processing. And the comprehensiveness hypothesis states that any computation model of human language processing should have a store of lexical knowledge as extensive that people do (Miller, 1998). The origin of WordNet was to build a lexical-conceptual model and database, consisting of both lexical units and the relations between such units, structured into a relational semantic network.

In spite of the original focus on cognitive plausibility, most of the work that originated from the WordNet community has focused mainly on linguistic and artificial intelligence applications (Morato, Marzal, Llorens, & Moreiro, 2004). Very few efforts have been placed on integrating the resource into cognitive architectures with the exception of NL-SOAR (Lehman, Lewis, & Newell, 1991; Lewis, 1993). No such implementation had been done for ACT-R yet.

WN-LEXICAL fully implements the hypotheses or assumptions behind WordNet, first by supporting the independence of lexical knowledge, and second by implementing as closely as possible the knowledge representations used in WordNet. In WordNet nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets. Some of these relations are hypernym hyponym relations, synonyms antonyms relations, meronym holonym relations (member, part, and substance), reflexive lexical morpho-semantic relations, class memberships, causal relation between verbs, and attribute relations between nouns and adjectives. All of these relations are between synonym set pairs.

## WN-LEXICAL

The act-r WordNet implementation was developed from the prolog version of WordNet 2.0. The prolog files were parsed and complied in a suitable lisp format to be used by the ACT-R cognitive architecture. The WordNet lexicon is an interface module to the act-r declarative memory. WN-LEXICAL holds a total of 553,001 chunks.

The module is meant to be platform independent and only uses common lisp, no external database. It requires about 46MB of ram. Loading the module takes about 2 and a half

minutes on a 1GHz PowerBook G4 with 512MB of memory. It was developed with LispWorks.

WN-LEXICAL can be used in the cognitive architecture of ACT-R in two modes: as an independent lexical module, or as a mean to load all of its chunks into declarative memory.

As an independent module, WN-LEXICAL has no parameters or subsymbolic features that would contribute to the modeling of lexical access latency. The basic lexical access processes implemented in the module consists of a retrieval of WordNet chunks followed by a selection process. All request to WN-LEXICAL are done with a chunk type wnl-request which can take a word string, a synset-id, a wordnet operator, and a context-criterion. The synset-id and wordnet operator are wordnet specific. A synset-id is an identifier for a sense set, and the WordNet operator specifies which meaning relation is encoded in a chunk such as word sense (S), or definition (G), or synonym (SIM). The context criterion specifies how the selection should proceed. WN-LEXICAL makes a selection by using a context, which consists of the set of selected wn-chunks in the past. The current version has 3 possibilities: A) no criterion, then a random selection is made from the retrieved set, B) set-difference, then a random selection is made form the set difference of the retrieved set and the context, C) set-intersection, and then a random selection is made form the set intersection of the retrieved set and the context.

When WN-LEXICAL chunks are loaded in declarative memory, these chucks are accessed through the declarative memory buffer and the WN-LEXICAL buffer is not necessary. This modeling approach could be used to implement the lexical decision model of Van Rijn and Anderson (Van Rijn, & Anderson, 2003).

## Conclusion

WN-LEXICAL, is an implementation of the WordNet lexical database, which can be added as a module to the basic ACT-R architecture for building large-scale conceptual or natural language processing models. WN-LEXICAL can be used in two modes: as an independent lexical module, or as a mean to load all of the 553,001 chunks into declarative memory.

Future work on WN-LEXICAL will require to associate word frequencies to the WordNet lexical entries. The Brown corpus is one possibility (Francis, & Kucera, 1982). Other developments are to develop models of various semantic priming phenomenon such as automatic versus strategic priming, associative versus pure semantic priming, mediated versus direct priming, effects of lag, forward versus backward priming, conscious versus unconscious priming, list context effects, as well as word frequency, stimulus quality, and stimulus repetition (McNamara, 2005).

## References


Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J.H., Koroshetz, W.J., & Pinker, S. (1997). A Neural Dissociation within Language: Evidence that the Mental Dictionary Is Part of Declarative Memory, and that Grammatical Rules Are Processed by the Procedural System, *The Journal of Cognitive Neuroscience, 9*, 266-276.

Ball, J., Rodgers, S., & Gluck, K. (2004). Integrating ACT-R and Cyc in a Large-Scale Model of Language Comprehension for Use in Intelligent Agents. In R. M. Jones (ed.), *Intelligent Agent Architectures: Combining the Strengths of Software Engineering and Cognitive Systems: Papers from the 2004 AAAI Workshop*, pp. 24-31, Menlo park, CA : AAAI.

Tyler, L.K., Marslen-Wilson, W.D., & Stamatakis, E.A. (2005). Differentiating lexical form, meaning, and structure in the neural language system, *PNAS, 102(23)*: 8375 - 8380.

Lehman, J., Lewis, R., & Newell, A. (1991). Integrating knowledge sources in language comprehension. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Pittsburgh, PA: Cognitive Science Society Incorporated.

Lewis R. (1993). *An Architecturally-based Theory of Human Sentence Comprehension*. PhD thesis, Carnegie Mellon University.

Fellbaum, C. (ed) (1998). *WordNet: An electronic lexical database. Cambridge*. MA : MIT Press.

Miller, G.A. (1998). Foreword, In C. Fellbaum, (ed), *WordNet: An electronic lexical database.* Cambridge, MA : MIT Press.

Van Rijn, H., Anderson, J.R. (2003). Modeling lexical decision as ordinary retrieval, In *Proceedings of the International Conference on Cognitive Modeling*, pp. 207-212.

McNamara, T.P. (2005). *Semantic priming: perspectives from memory and word recognition*. New York, NY : Psychology Press.

Morato, J., M. Marzal, M.A., Llorens, J., Moreiro, J. (2004). WordNet Applications, In *Proceedings of the Second Global WordNet Conference*, pp. 270-278.

Francis, W.N., Kucera. H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.