Grounding Abstract Word Definitions In Prior Concrete Experience

G. Chicoisne[1], A. Blondin-Massé[2], O. Picard[1], S. Harnad[1,3]

[1]Institut des sciences cognitives, UQAM, Montréal, Canada
[2]Lab. de combinatoire et d'informatique-mathématique, UQAM, Montréal, Canada
[3]Electronics & Computer Science, U. Southampton

We use dictionaries to look up unknown words, but meaning cannot be based on definitions all the way down. This is the "symbol grounding problem" [Harnad 2005]. Some word meanings must be grounded in sensorimotor category learning.

Longman's Dictionary of Contemporary English (LDOCE) and Cambridge International Dictionary of English (CIDE) have a *defining vocabulary* of 2000 words from which most but not all of the words are defined. We developed an algorithm, applicable to any dictionary, for recursively deleting words not used in any definition until the corpus is reduced to a subset – a *grounding kernel* – from which all the words of the complete dictionary are *reachable* by definition alone.

Just as dictionaries have a grounding kernel whose meanings cannot be learned through definition alone, so the mental lexicon with which each user comes to a dictionary must also have a kernel that is grounded by some means other than definition alone. Most words refer to *categories* (individuals, kinds, actions, events, traits and states) that are learned (or learnable) through *induction* from sensorimotor experience (though some categories may also have been shaped by evolution) [Harnad 2005]. Once learned, a category can be tagged with an arbitrary, shared name. These grounded word meanings then provide entry points for acquiring further categories and their names through *instruction,* based on boolean recombinations of the grounded names that define or describe them.

We compared LDOCE's and CIDE's defining vocabulary (DV) and grounding kernel (GK) against the rest of its words, on (1) *concreteness*, (2) *imagery* and (3) *age of acquisition* based on the MRC psycholinguistic database [Wilson

1988]. Both GK and DV proved significantly higher on all three scales (p<0.001) for all but 2 of the 12 comparisons (no difference in concreteness in LDOCE for either GK or DV). The difference was also consistently greater using our automatically computed GK than using the DV defined by the compilers of LDOCE and CIDE.

All categorization is abstraction, but we predicted that the grounding kernel would be more concrete, more imageable and acquired earlier than the rest of our vocabulary. The outcome supports the hypothesis that dictionary meanings are grounded in a mental lexicon that is in turn grounded in prior sensorimotor learning.

References

Harnad, S. (2005) To Cognize is to Categorize: Cognition is Categorization, in Lefebvre, C. and Cohen, H., Eds. *Handbook of Categorization*. Elsevier.

Wilson, M.D. (1988) The MRC Psycholinguistic Database: Machine Readable Dictionary. *Behav Res Methods, Instruments and Computers*, **20**(1), 6-11.