A THEORY OF LEXICAL ACCESS IN SPEECH PRODUCTION

Willem J.M. Levelt
Max Planck Institute for Psycholinguistics
P.O. Box 310, 6500 AH Nijmegen, The Netherlands
pim@mpi.nl

ABSTRACT

The generation of words in speech involves a number of processing stages. There is, first, a stage of conceptual preparation; this is followed by stages of lexical selection, phonological encoding, phonetic encoding and articulation. In addition, the speaker monitors the output and, if necessary, self-corrects. Major parts of the theory have been computer modelled. The paper concentrates on experimental reaction time evidence in support of the theory.

Central to the skill of speaking is our ability to select words that appropriately express our intentions, to retrieve their syntactic and phonological properties and to compute the ultimate articulatory shape of these words in the context of the utterance as a whole (2).

In the multi-stage theory of word production (3) the first stage, *conceptual preparation*, involves activating a lexical concept, given the intention. In picture naming, for instance, there is no "hard-wired" link between the object depicted and the ultimate referential expression. The same object can be veridically referred to by a multitude of different terms. The mediating process here is called *perspective taking*. Its output is a *lexical concept*, i.e., a concept for which there is a word in the speaker's mental lexicon. In the computational model, lexical concepts figure in a semantic, spreading-activation network.

The lexical concept is input to a process called *lexical selection*. Lexical concepts spread their information to *lemmas* in the mental lexicon. Lemmas are syntactic words. The probability that a lemma is selected within a minimal time interval is its relative activation (following Luce's choice rule). From this hazard rate expected retrieval times can be computed for various experimental conditions. These predictions find solid experimental support (5). A selected lemma spreads its activation to the word's *phonological code*. The speed of accessing this code is word-frequency dependent (1).

During *phonological encoding* the segmental and metrical features of the word's phonological code are "spelled out". The metrical structures of adjacent words may get combined to compute larger-size metrical units, so-called *phonological words*. The spelled-out segments are incrementally ("from left to right") attached to the metrical frame, on the fly creating the phonological word's syllabification.

In *phonetic encoding* an articulatory gesture is computed for each phonological syllable as it comes available. This process probably involves a *syllabary*, a store of high-frequent syllabic gestures (4). Articulation can be initiated as soon as all of a word's syllabic gestures have been prepared.

A speaker self-monitors conceptual preparation, acoustic output, but also an intermediary level of representation, namely the syllabified phonological word (2, 6).

REFERENCES
1. Jescheniak, J. & Levelt, W.J.M. (1994). Word frequency effects in production. *Journal of Experimental Psychology LMC*, 824-843.
2. Levelt, W.J.M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.
3. Levelt, W.J.M. (1994). On the skill of speaking: How do we access words? ICSLP 94, 2253-2258.
4. Levelt, W.J.M. & Wheeldon. L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239-269.
5. Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition, 42*, 107-142.
6. Wheeldon, L. & Levelt, W.J.M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34*, 311-334.