5

STORAGE AND COMPUTATION IN THE MENTAL LEXICON

R. H. Baayen, Radboud University and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

1. INTRODUCTION

In the seventies of the previous century, the mathematical properties of formal languages have provided a key source of inspiration to morphological theory. Models such as developed by Lieber (1980) and Selkirk (1984) viewed the lexicon as a calculus, a formal system combining a repository of morphemes with rules for combining these morphological atomic units into complex words.

This approach to the lexicon was driven by two fundamental assumptions. First, the lexicon was assumed to be a compositional derivational system. Complex words were believed to be generated from simpler forms (Bloch, 1947; Chomsky and Halle, 1968). Second, following Bloomfield (1933), the set of atomic elements was assumed to comprise any word or formative that is not predictable by rule. Rule-governed combinations of these atomic units, the regular complex words, were assumed not to be available as units in the lexicon, as storage would introduce unnecessary redundancy in the model. Instead of being listed (i.e., stored without analysis or substructure), regular complex words were generated (produced or parsed) by rule. Unsurprisingly, the goal of morphological theory was seen as accounting for which words belong to the set of possible words in the languages of the world. The question of whether a regular complex word exists in a language was regarded as a question addressing performance rather than competence, and hence irrelevant for morphological theory.

Although many other formalisms have been developed to replace sequences of rules (Halle and Marantz, 1993; McCarthy and Prince, 1993), these formalisms did not challenge these fundamental assumptions of generative morphology. In optimality theory, for instance,
forms still enter into derivational relations, even though the algorithm that relates underlying forms to surface forms is not based on a sequence of rules but on constraint satisfaction. This type of theory of the lexicon is to a surprising extent equally adequate as a competence theory for how a pocket calculator works. A pocket calculator has a set of atomic elements, the symbols on its keys. Its chip is endowed with a small set of arithmetic rules that, when supplied with a legal string, compositionally evaluate this string. Whenever a pocket calculator is requested to evaluate a string such as “2 + 3”, it computes the outcome. It has no memory that holds the output of previous evaluations of the same string. It never learns from past experience. The balance of storage and computation is shifted totally to the maximization of computation and the minimization of storage.

The first goal of this chapter is to show that the pocket calculator provides a fundamentally flawed metaphor for understanding morphological structure and processing in the mental lexicon. To this end, we first survey evidence from experimental studies of lexical processing, and then consider another source of information, the fine phonetic detail that is present in the acoustic signal. We then address the second goal of this chapter, to provide an indication of the kind of formal mathematical model that may help us to better understand process and representation in the mental lexicon.

2. EXPERIMENTAL EVIDENCE

Over the last twenty-five years, the regular and irregular past tense forms in English and related languages have provided a rich testing ground for theories of morphological processing. Whereas English regular verbs have a past tense form in the dental suffix -ed (e.g., walked, claimed), irregular verbs have past tense forms that range from suppletion (go/went) to invariance (cut/cut) and from pure vocalic alternation (give/gave) to combinations of vocalic alternation and a variant of the dental suffix (sell/sold). Bybee and Slobin (1982) and Bybee and Moder (1983) called attention to the many kinds of subregularities that characterize the irregular verbs, which older structuralists had already characterized as semi-productive (e.g., Van Haeringen, 1940).

Most researchers understand regular past tense forms as being derived from their present tense stems (Bloch, 1947; Chomsky and Halle, 1968) by a simple rule adding the dental suffix. Although irregular past tense forms might also be analyzed as governed by various unproductive rules, such rule-based descriptions tend to be baroque, fairly arbitrary and uninsightful. For understanding the semi-regularities of the irregular past tense, connectionist models offered an alternative that obviated the need for a series of ad hoc unproductive rules (Rumelhart and McClelland, 1986; McClelland and Patterson, 2002b). The response of generative linguists (Pinker and Prince, 1988, 1994; Pinker, 1991) was to defend the Bloomfieldian model by claiming that regular and irregular morphology belong to two completely independent cognitive systems, the dual mechanisms of rule (for regulars)