Journal of Automata, Languages and Combinatorics **14** (2009) 1, 33–62


# MARKERS AND DETERMINISTIC ACCEPTORS FOR NON-DETERMINISTIC LANGUAGES

HELMUT JÜRGENSEN
*Department of Computer Science, The University of Western Ontario*
*London, Ontario, Canada N6A 5B7*

ABSTRACT

We investigate the question as to how the intuitive structure of words in a formal language is reflected by the structure of the reduced deterministic acceptor for that language. Markers in words are known to facilitate the understanding of structure and the design of word recognition algorithms. Explicit markers, that is, special symbols, are rather artificial. To avoid such special symbols, but to benefit from the effect of markers we consider implicit markers – special words which provide segmentation information about input words. We propose a formal definition of markers based on the notion of relativized solid codes and indicate how this concept can assist in the understanding of the structure of words and the construction of acceptors via automaton extensions.

*Keywords:* Formal languages, automata, Nerode equivalence, determinism, non-determinism, marker, solid code, relativized code

## 1. Ideas

A language is a set of words over a finite, non-empty alphabet. Every language is accepted by a deterministic automaton. Such an acceptor can be assumed to be reduced in the following sense: the acceptor has no equivalent states; moreover, the reduced deterministic acceptor is unique up to isomorphism. We start with this statement and the following seemingly contradictory one: Every context-free language is accepted by a non-deterministic push-down automaton, but there are context-free languages for which there is no deterministic push-down acceptor. Of course, there is no real contradiction: the reduced deterministic acceptor for a given context-free language need not be a push-down acceptor.

The purpose of this paper is to present and question ideas in formal language theory from a point of view in information theory. In particular we should like to know how the structure of a language, possibly expressed in terms of informational dependency between parts of words, is reflected by the structure of its reduced deterministic acceptor. In natural languages we seem to use special markers to structure utterances. However, there does not seem to exist an informal understanding, let alone a formal definition, of what a marker is – beyond of course that of special symbols not occurring otherwise and referred to as explicit markers in the sequel. Explicit markers are quite common in programming languages; in the general theory of formal languages