On the Translation of Languages from Left to Right

Donald E. Knuth

Mathematics Department, California Institute of Technology, Pasadena, California

There has been much recent interest in languages whose grammar is sufficiently simple that an efficient left-to-right parsing algorithm can be mechanically produced from the grammar. In this paper, we define $LR(k)$ grammars, which are perhaps the most general ones of this type, and they provide the basis for understanding all of the special tricks which have been used in the construction of parsing algorithms for languages with simple structure, e.g. algebraic languages. We give algorithms for deciding if a given grammar satisfies the $LR(k)$ condition, for given k, and also give methods for generating recognizers for $LR(k)$ grammars. It is shown that the problem of whether or not a grammar is $LR(k)$ for *some* k is undecidable, and the paper concludes by establishing various connections between $LR(k)$ grammars and deterministic languages. In particular, the $LR(k)$ condition is a natural analogue, for grammars, of the deterministic condition, for languages.

I. INTRODUCTION AND DEFINITIONS

The word "language" will be used here to denote a set of character strings which has been variously called a *context free language*, a *(simple) phrase structure language*, a *constituent-structure language*, a *definable set*, a *BNF language*, a *Chomsky type 2 (or type 4) language*, a *push-down automaton language*, etc. Such languages have aroused wide interest because they serve as approximate models for natural languages and computer programming languages, among others. In this paper we single out an important class of languages which will be called *translatable from left to right*; this means if we read the characters of a string from left to right, and look a given finite number of characters ahead, we are able to parse the given string without ever backing up to consider a previous decision. Such languages are particularly important in the case of computer programming, since this condition means a parsing algorithm can be mechanically constructed which requires an execution time at worst proportional to the length of the string being parsed. Special-purpose