ZBORNIK RADOVA
Prirodno-matematičkog fakulteta
Univerziteta u Novom Sadu
Serija za matematiku, 15,1 (1985)

REVIEW OF RESEARCH
Faculty of Science
University of Novi Sad
Mathematics Series, 15,1 (1985)

THE IMPLEMENTATION OF CF-GRAMMARS BY PROLOG LANGUAGE*

Dragan Kolar and Vojislav Stojković

Prirodno-matematički fakultet
Beograd, Studentski trg 16,
Jugoslavija

ABSTRACT

In this paper we have introduced the concept of partial derivation of the string. We have formulated and proved three theorems and their consequences, as well. According to these definition and theorems we can realize the implementation of grammars rules context-free grammars by PROLOG language, execute syntax analysis of strings and examine membership of strings to the language generated by grammars, which defines this grammar rules.

Let us suppose $G = (N,T,P,S)$ is a context-free grammar, $L = \{w | S \overset{*}{\Rightarrow} w,\ w \in T^{+}\}$ is a language generated by G and $A \to s_1 \dots s_n$ is a grammar rule, which is a member of a set P, where $s_i \in N \cup T$, $i = 1, \dots, n$.

DEFINITION. *Let us suppose* $w_0 \in T^{+}$ *and that there exists a derivation* $S \overset{*}{\Rightarrow} a_1 X a_2$, *where* $a_1, a_2 \in (N \cup T)^{*}$ *and* $X \in N \cup T$. *If there is a string* $w_1 \in T^{+}$, *so that* $X \overset{*}{\Rightarrow} w_1$ *and*

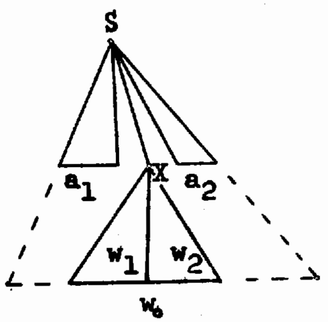
* This paper was reported at the ICM-83 in Warszawa, 1983.

AMS Mathematics Subject Classification(1980):68A05.
Key words and phrases: Logic programming, Prolog language, context-free grammar, partial derivation of string, implementation of grammar rules, syntax analysis of string.

$w_0 = w_1 w_2$, *where* $w_2 \in T^*$, *then such a derivation of the string* w_0 *from* X *is called a partial derivation of the string, in the denotation* $X(w_0, w_2)$.

COROLLARY 1. *If* $w_0, w_1 \in T^+$ *and* $w_0 = w_1 w_2$, *then* $w_2 \in T^*$.

COROLLARY 2. *If* $X \in T$, *then* $w_1 = X$, *that is* $w_0 = Xw_2$.



This definition can be represented graphically in the following way:

THEOREM 1. *The string* w_0 *is a member of set* L *if and only if* $S(w_0, \varepsilon)$ *is a partial derivation of the string, where* ε *is an empty string.*

PROOF. Let $w_0 \in L$. This means that $w_0 \in T^+$ and $S \overset{*}{\Rightarrow} w_0$. Let us prove that $S(w_0, \varepsilon)$ is a partial derivation of the string. Let $w_1 = w_0$. According to supposition $S \overset{*}{\Rightarrow} w_0 = w_1 \in T^+$ and $w_0 = w_0\varepsilon = w_1\varepsilon$, it follows that $S(w_0, \varepsilon)$ is a partial derivation of the string.

Let $S(w_0, \varepsilon)$ be a partial derivation of the string. This means that there is a string $w_1 \in T^+$, so that $S \overset{*}{\Rightarrow} w_1$ and $w_0 = w_1 w_2$. Since $w_0 = w_1\varepsilon$, it follows that $w_1 = w_0$, and there is a derivation $S \overset{*}{\Rightarrow} w_0 \in T^+$. That is $w_0 \in L$.

THEOREM 2. *If we suppose that there are derivations* $s_i \overset{*}{\Rightarrow} w_i' \in T^+$, $i = 1, \ldots, n$ *and if the sequence of strings* (w) *is defined as follows:*

$$\begin{cases} w_n \in T^* & \text{is an arbitrary string} \\ w_{i-1} = w_i' w_i, & i = n, n-1, \ldots, 1, \end{cases}$$

then the following claims are valid:

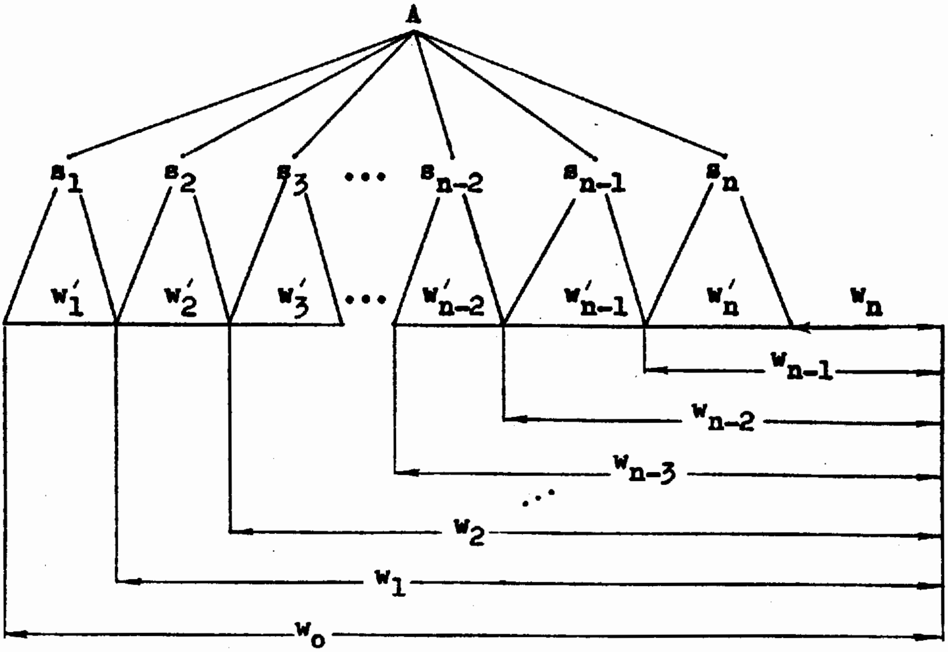
a) $A \overset{*}{\Rightarrow} w_1' w_2' \ \dots \ w_n'$

b) $w_0 = w_1' w_2' \ \dots \ w_n' \ w_n$

c) $|w_0| > |w_1| > \ \dots > |w_{n-1}| > |w_n| \geq 0,$

where |a| *is a denotation for the length of string* a.

This theorem can be represented graphically in the following way:



PROOF. a) Since grammar rule $A \to s_1 \ \dots \ s_n$ is a member of a set P and $s_i \overset{*}{\Rightarrow} w_i' \in T^+$, $i = 1, \dots, n$, it follows that there is a derivation $A \Rightarrow s_1 \ \dots \ s_n \overset{*}{\Rightarrow} w_1' \ \dots \ w_n'$.

b) According to the definition of the sequence of strings (w), it follows that: $w_0 = w_1' \ w_1 = w_1' \ w_2' \ w_2 = \dots$ $\dots = w_1' \ w_2' \ \dots \ w_{n-1}' \ w_{n-1} = w_1' \ w_2' \ \dots \ w_{n-1}' \ w_n' \ w_n$.

c) Since $w_{i-1} = w_i' w_i$, $i = n, n-1, \dots, 1$ and $w_i' \in T^+$, i.e. $|w_i'| \geq 1$, it follows that: $|w_{i-1}| = |w_i'| + |w_i| \geq 1 + |w_i|$

i.e.: $|w_{i-1}| > |w_i|$, $i = 1,\dots,n-1$.

If $w_n \in T^+$, then according to the supposition that $w_{n-1} = w'_n w_n$, it follows that: $|w_{n-1}| = |w'_n| + |w_n| \geqslant 1 + |w_n|$

i.e.: $|w_{n-1}| > |w_n| > 0$.

If $w_n = \varepsilon$, then $w_{n-1} = w'_n$ and $|w_{n-1}| = |w'_n| > 0 = |w_n|$.

Thus the theorem is proved.

COROLLARY 1. $w_0 = w'_1 w'_2 \dots w'_n$, *if* $w_n = \varepsilon$.

PROOF. According to claim b) of Theorem 2, $w_0 = w'_1 w'_2 \dots w'_n w_n$, and supposition $w_n = \varepsilon$ it follows that $w_0 = w'_1 \dots w'_n$.

COROLLARY 2. $|w_0| > |w_1| > \dots > |w_{n-1}| > |w_n| = 0$, *if* $w_n = \varepsilon$.

PROOF. According to claim c) of Theorem 2, $|w_0| > |w_1| > \dots > |w_{n-1}| > |w_n| \geqslant 0$ and the proof of claim c) of the Theorem 2, if $w_n = \varepsilon$ it follows that $|w_0| > |w_1| > \dots > |w_{n-1}| > |w_n| = 0$.

COROLLARY 3. $w'_1 w'_2 \dots w'_n \in L$, *if* $A = S$.

PROOF. According to claim a) of Theorem 2, it follows that $S \overset{*}{\Rightarrow} w'_1 w'_2 \dots w'_n$. Since $w'_i \in T^+$, $i = 1,\dots, n$ it follows that $w'_1 w'_2 \dots w'_n \in T^+$. According to the previous it follows that $w'_1 w'_2 \dots w'_n \in L$.

COROLLARY 4. $w_0 \in L$, *if* $A = S$ *and* $w_n = \varepsilon$.

PROOF. According to Corollary 1, it follows that $w_0 = w'_1 w'_2 \dots w'_n$ and according to Corollary 3, it follows that $w_1' w_2' \dots w_n' \in L$, it follows that $w_0 \in L$.

THEOREM 3. *If* $s_1(w_0,w_1)$, $s_2(w_1,w_2),\dots,s_n(w_{n-1},w_n)$ *are partial derivations of the strings, then* $A(w_0,w_n)$ *is a partial derivation of the string.*

PROOF. Let $s_i(w_{i-1},w_i)$, $i = 1,\dots,n$ are partial derivations of the strings. This means that $s_i \in N \cup T$, $w_{i-1} \in T^+$, $i = 1,\dots,n$, $w_n \in T^*$ and there are strings $w_i' \in T^+$, so that $s_i \overset{*}{\Rightarrow} w_i'$ and $w_{i-1} = w_i' w_i$, $i = 1,\dots,n$. Then according to claim a) of Theorem 2 it follows that $A \overset{*}{\Rightarrow} w_1' w_2' \dots w_n' \in T^+$ and according to claim b) of Theorem 2, it follows that $w_0 = w_1' w_2' \dots w_n' w_n$. This means that $A(w_0,w_n)$ is a partial derivation of the string.

Thus the theorem is proved.

COROLLARY 1. *Let* $A = S$. *If* $s_1(w_0,w_1)$, $s_2(w_1,w_2)$, $\dots$, $s_n(w_{n-1},w_n)$ *are partial derivations of the strings, then* $S(w_0,w_n)$ *is a partial derivation of the string.*

COROLLARY 2. *Let* $w_n = \varepsilon$. *If* $s_1(w_0,w_1)$, $s_2(w_1,w_2)$, $\dots$, $s_n(w_{n-1},\varepsilon)$ *are partial derivations of the strings, then* $A(w_0,\varepsilon)$ *is a partial derivation of the string.*

COROLLARY 3. *Let* $A = S$ *and* $w_n = \varepsilon$. *If* $s_1(w_0,w_1)$, $s_2(w_1,w_2)$, $\dots$, $s_n(w_{n-1},\varepsilon)$ *are partial derivations of the strings, then* $S(w_0,\varepsilon)$ *is a partial derivation of the string, i.e.* $w_0 \in L$.

The proof of the corollaries of Theorem 3 is trivial.

According to these theoretic results, we can realize the implementation of grammar rules of context-free grammars by PROLOG language, execute syntax analysis of strings and examine the membership of strings to the language generated by grammar, which defines these grammar rules.

We shall represent the string as a list. The elements of the list are the symbols of the string. An empty string corresponds to an empty list. The partial derivation of the string, in the denotation $s(w_0,w_1)$, in PROLOG language, is presented with the goal s(LØ,L1), where LØ and L1 are lists, whose elements are the symbols of the strings w_0 and w_1. If $s \in T$, then a partial derivation of the string is presented by the goal lht(LØ,s,L1), where lht is a predicate defined by the fact:

```
lht([H|T],H,T).
```

This means that the list denoted by [H|T] has a head H and a tail T.

According to Theorem 3, the grammar rule $A \to a_1 a_2 \ldots \ldots a_n$, symbolically, could be written in PROLOG language in the following way, by the PROLOG rule:

$$a(L\emptyset,L) :- \begin{Bmatrix} a_1(L\emptyset,L1) \\ lht(L\emptyset,a_1,L1) \end{Bmatrix}, \begin{Bmatrix} a_2(L1,L2) \\ lht(L1,a_2,L2) \end{Bmatrix}, \ldots, \begin{Bmatrix} a_n(L_{n-1},L) \\ lht(L_{n-1},a_n,L) \end{Bmatrix}.$$

where L_i, $i = \emptyset,1,\ldots, n-1$ is a list, whose elements are symbols of string w_i, $\{{b_1 \atop b_2}\}$ is the denotation for choosing one of the goals b_1 and b_2, depending on $s_i \in N$ or $s_i \in T$, $i = 1,\ldots \ldots, n$, and the comma is the conjunction operator.

As we can see, each grammar rule takes an input string, analyses some initial portion, and produces the remaining portion as output for further analysis. The arguments required for the input and output strings are not written explicitly in a grammar rule, but the syntax implicitly defines them.

The implemented grammar rules form the base of knowledge. According to Corollary 3 of Theorem 3 the PROLOG system to the question:

```
? - s(LØ,[ ]).
```

answers the string w_0, represented by the list LØ, is a member of the language generated by the grammar, which defines these grammar rules.

EXAMPLE. Let us consider the following grammar:

$$G = \{N,T,P,E\}$$
$$N = \{E,T,F\}$$
$$T = \{+,\times,(,),a\}$$
$$P = \left\{\begin{array}{l} E \rightarrow T+E \mid T \\ T \rightarrow F\times T \mid F \\ F \rightarrow (E) \mid a \end{array}\right\}.$$

This grammar generates all the arithmetical expressions that consist of terminal symbols in accordance with the given grammar rules. The implementation of these grammar rules is as follows:

```
e(LØ,L):-t(LØ,L1), lht(L1,+,L2), e(L2,L).
e(LØ,L):-t(LØ,L).
t(LØ,L):-f(LØ,L1), lht(L1,×,L2), t(L2,L).
t(LØ,L):-f(LØ,L).
f(LØ,L):-lht(LØ,(,L1), e(L1,L2), lht(L2,),L).
f(LØ,L):-lht(LØ,a,L).
```

For example, if we ask the question:

```
? - e([a,+,a],[ ]).
```

we shall get the answer:

```
yes.
```

If we ask the question:

```
? - e([x,ȧ,+],[ ]).
```

we shall get the answer:

```
no.
```

The answers yes or no have the following meanings: the string is syntactically correct or the string is not syntactically correct.

The list representation of the string has no restrictions since by the corresponding program each string, written in the conventional way, may be converted into the list whose elements are the symbols of the string.

REFERENCE

[1] W.F. Clocksin, C.S. Mellish, *Programming in Prolog Springer-Verlag, Berlin, 1981.*

Received by the editors June 9, 1985.

REZIME

IMPLEMENTACIJA KONTEKSTNO-SLOBODNIH GRAMATIKA POMOĆU PROLOG JEZIKA

U radu je definisan pojam parcijalnog izvodjenja niske, formulisane su i dokazane tri teoreme i više njihovih posledica. Na osnovu uvedene definicije i teorema može se izvršiti implementacija gramatičkih pravila kontekstno-slobodnih gramatika pomoću PROLOG jezika, izvršiti sintaksna analiza niski i ispitivati pripadnost niski jeziku generisanom gramatikom, koja je definisana tim gramatičkim pravilima.