FORMALIZED MATHEMATICS
Vol.2,No.5, November–December 1991
Université Catholique de Louvain

Context-Free Grammar - Part 1

Patricia L. Carlson
Teachers' College of English
Białystok

Grzegorz Bancerek
Warsaw University, Białystok
IM PAN, Warsaw

Summary. The concept of context-free grammar and of derivability in grammar are introduced. Moreover, the language (set of finite sequences of symbols) generated by grammar and some grammars are defined. The notion convenient to prove facts on language generated by grammar with exchange of symbols on grammar of union and concatenation of languages is included.

MML Identifier: LANG1.

The notation and terminology used here have been introduced in the following papers: [9], [7], [1], [8], [10], [11], [4], [2], [6], [5], and [3]. We consider context-free grammars which are systems

$\langle$symbols, a initial symbol, rules$\rangle$,

where the symbols constitute a non-empty set, the initial symbol is an element of the symbols, and the rules constitute a relation between the symbols and (the symbols)*.

We now define two new modes. Let G be a context-free grammar. A symbol of G is an element of the symbols of G.

A string of G is an element of (the symbols of G)*.

Let D be a non-empty set, and let p, q be elements of D^*. Then $p \frown q$ is an element of D^*.

Let D be a non-empty set. Then ε_D is an element of D^*. Let d be an element of D. Then $\langle d \rangle$ is an element of D^*. Let e be an element of D. Then $\langle d, e \rangle$ is an element of D^*.

In the sequel G will denote a context-free grammar, s will denote a symbol of G, and n, m will denote strings of G. Let us consider G, s, n. The predicate $s \Rightarrow n$ is defined as follows:

(Def.1) $\langle s, n \rangle \in$ the rules of G.

We now define two new functors. Let us consider G. The terminals of G yields a set and is defined as follows:

ISSN 0777–4028

(Def.2) the terminals of $G = \{s : \neg \bigvee_n s \Rightarrow n\}$.

The nonterminals of G yielding a set is defined as follows:

(Def.3) the nonterminals of $G = \{s : \bigvee_n s \Rightarrow n\}$.

Next we state the proposition

(1) (The terminals of G) $\cup$ (the nonterminals of G) = the symbols of G.

Let us consider G, n, m. The predicate $n \Rightarrow m$ is defined by:

(Def.4) there exist strings n_1, n_2, n_3 of G and there exists s such that $n = n_1 \frown \langle s \rangle \frown n_2$ and $m = n_1 \frown n_3 \frown n_2$ and $s \Rightarrow n_3$.

In the sequel n_1, n_2, n_3 denote strings of G. One can prove the following four propositions:

(2) If $s \Rightarrow n$, then $n_1 \frown \langle s \rangle \frown n_2 \Rightarrow n_1 \frown n \frown n_2$.

(3) If $s \Rightarrow n$, then $\langle s \rangle \Rightarrow n$.

(4) If $\langle s \rangle \Rightarrow n$, then $s \Rightarrow n$.

(5) If $n_1 \Rightarrow n_2$, then $n \frown n_1 \Rightarrow n \frown n_2$ and $n_1 \frown n \Rightarrow n_2 \frown n$.

Let us consider G, n, m. The predicate $n \Rightarrow_* m$ is defined by the condition (Def.5).

(Def.5) There exists a finite sequence p such that $\operatorname{len} p \geq 1$ and $p(1) = n$ and $p(\operatorname{len} p) = m$ and for every natural number i such that $i \geq 1$ and $i < \operatorname{len} p$ there exist strings a, b of G such that $p(i) = a$ and $p(i+1) = b$ and $a \Rightarrow b$.

The following three propositions are true:

(6) $n \Rightarrow_* n$.

(7) If $n \Rightarrow m$, then $n \Rightarrow_* m$.

(8) If $n_2 \Rightarrow_* n_1$ and $n_3 \Rightarrow_* n_2$, then $n_3 \Rightarrow_* n_1$.

Let us consider G. The language generated by G yielding a set is defined by:

(Def.6) the language generated by
$G = \{a : \operatorname{rng} a \subseteq$
the terminals of $G \wedge \langle$the initial symbol of $G\rangle \Rightarrow_* a\}$, where a ranges over elements of (the symbols of $G)^*$.

Next we state the proposition

(9) $n \in$ the language generated by G if and only if $\operatorname{rng} n \subseteq$ the terminals of G and $\langle$the initial symbol of $G\rangle \Rightarrow_* n$.

Let a be arbitrary. Then $\{a\}$ is a non-empty set. Let b be arbitrary. Then $\{a, b\}$ is a non-empty set.

Let D, E be non-empty sets, and let a be an element of $[\!: D, E :\!]$. Then $\{a\}$ is a relation between D and E. Let b be an element of $[\!: D, E :\!]$. Then $\{a, b\}$ is a relation between D and E.

We now define three new functors. Let a be arbitrary. The functor $\{a \Rightarrow \varepsilon\}$ yielding a strict context-free grammar is defined by:

(Def.7) the symbols of $\{a \Rightarrow \varepsilon\} = \{a\}$ and the rules of $\{a \Rightarrow \varepsilon\} = \{\langle a, \varepsilon \rangle\}$.

Let b be arbitrary. The functor $\{a \Rightarrow b\}$ yielding a strict context-free grammar is defined as follows:

(Def.8) the symbols of $\{a \Rightarrow b\} = \{a, b\}$ and the initial symbol of $\{a \Rightarrow b\} = a$ and the rules of $\{a \Rightarrow b\} = \{\langle a, \langle b \rangle\rangle\}$.

The functor $\left\{ \begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array} \right\}$ yields a strict context-free grammar and is defined by:

(Def.9) the symbols of $\left\{ \begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array} \right\} = \{a, b\}$ and the initial symbol of $\left\{ \begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array} \right\} =$ a and the rules of $\left\{ \begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array} \right\} = \{\langle a, \langle b, a \rangle\rangle, \langle a, \varepsilon \rangle\}$.

Let D be a non-empty set. The total grammar over D yields a strict context-free grammar and is defined as follows:

(Def.10) the symbols of the total grammar over $D = D \cup \{D\}$ and the initial symbol of the total grammar over $D = D$ and the rules of the total grammar over $D = \{\langle D, \langle d, D \rangle\rangle : d = d\} \cup \{\langle D, \varepsilon \rangle\}$, where d ranges over elements of D.

In the sequel a, b are arbitrary and D denotes a non-empty set. Next we state several propositions:

(10) The terminals of $\{a \Rightarrow \varepsilon\} = \emptyset$.

(11) The language generated by $\{a \Rightarrow \varepsilon\} = \{\varepsilon\}$.

(12) If $a \neq b$, then the terminals of $\{a \Rightarrow b\} = \{b\}$.

(13) If $a \neq b$, then the language generated by $\{a \Rightarrow b\} = \{\langle b \rangle\}$.

(14) If $a \neq b$, then the terminals of $\left\{ \begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array} \right\} = \{b\}$.

(15) If $a \neq b$, then the language generated by $\left\{ \begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array} \right\} = \{b\}^*$.

(16) The terminals of the total grammar over $D = D$.

(17) The language generated by the total grammar over $D = D^*$.

We now define two new attributes. A context-free grammar is efective if:

(Def.11) the language generated by it is non-empty and the initial symbol of it $\in$ the nonterminals of it and for every symbol s of it such that $s \in$ the terminals of it there exists a string p of it such that $p \in$ the language generated by it and $s \in \operatorname{rng} p$.

A context-free grammar is finite if:

(Def.12) the rules of it is finite.

Let G be an efective context-free grammar. Then the nonterminals of G is a non-empty subset of the symbols of G.

Let X be a set, and let Y be a non-empty set, and let f be a function from X into Y. Then graph f is a relation between X and Y.

Let X, Y be non-empty sets, and let p be a finite sequence of elements of X, and let f be a function from X into Y. Then $f \cdot p$ is an element of Y^*.

Let X, Y be non-empty sets, and let f be a function from X into Y. The functor f^* yielding a function from X^* into Y^* is defined as follows:

(Def.13) for every element p of X^* holds $f^*(p) = f \cdot p$.

Let R be a binary relation. The functor R^* yielding a binary relation is defined by the condition (Def.14).

(Def.14) Let x, y be arbitrary. Then $\langle x, y \rangle \in R^*$ if and only if the following conditions are satisfied:

(i) $x \in \operatorname{field} R$,

(ii) $y \in \operatorname{field} R$,

(iii) there exists a finite sequence p such that $\operatorname{len} p \geq 1$ and $p(1) = x$ and $p(\operatorname{len} p) = y$ and for every natural number i such that $i \geq 1$ and $i < \operatorname{len} p$ holds $\langle p(i), p(i+1) \rangle \in R$.

In the sequel R is a binary relation. We now state the proposition

(18) $R \subseteq R^*$.

Let X be a non-empty set, and let R be a binary relation on X. Then R^* is a binary relation on X.

Let G be a context-free grammar, and let X be a non-empty set, and let f be a function from the symbols of G into X. The functor $G(f)$ yielding a strict context-free grammar is defined by:

(Def.15) $G(f) = \langle X, f(\text{the initial symbol of } G), (\operatorname{graph} f)^{\smile} \cdot \text{the rules of } G \cdot \operatorname{graph}(f^*) \rangle$.

The following proposition is true

(19) For all non-empty sets D_1, D_2 such that $D_1 \subseteq D_2$ holds ${D_1}^* \subseteq {D_2}^*$.

References

[1] Grzegorz Bancerek. The fundamental properties of natural numbers. *Formalized Mathematics*, 1(**1**):41–46, 1990.

[2] Grzegorz Bancerek and Krzysztof Hryniewiecki. Segments of natural numbers and finite sequences. *Formalized Mathematics*, 1(**1**):107–114, 1990.

[3] Czesław Byliński. Finite sequences and tuples of elements of a non-empty sets. *Formalized Mathematics*, 1(**3**):529–536, 1990.

[4] Czesław Byliński. Functions and their basic properties. *Formalized Mathematics*, 1(**1**):55–65, 1990.

[5] Czesław Byliński. Functions from a set to a set. *Formalized Mathematics*, 1(**1**):153–164, 1990.

[6] Agata Darmochwał. Finite sets. *Formalized Mathematics*, 1(**1**):165–167, 1990.

[7] Krzysztof Hryniewiecki. Basic properties of real numbers. *Formalized Mathematics*, 1(**1**):35–40, 1990.

[8] Andrzej Trybulec. Domains and their Cartesian products. *Formalized Mathematics*, 1(**1**):115–122, 1990.

[9] Andrzej Trybulec. Tarski Grothendieck set theory. *Formalized Mathematics*, 1(**1**):9–11, 1990.

[10] Edmund Woronowicz. Relations and their basic properties. *Formalized Mathematics*, 1(**1**):73–83, 1990.

[11] Edmund Woronowicz. Relations defined on sets. *Formalized Mathematics*, 1(**1**):181–186, 1990.

Received February 21, 1992