JOURNAL OF FORMALIZED MATHEMATICS
Volume 4, Released 1992, Published 2003
Inst. of Computer Science, Univ. of Białystok

Context-Free Grammar — Part I

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`.

WWW: `http://mizar.org/JFM/Vol4/lang1.html`

The articles [8], [5], [9], [1], [10], [11], [3], [2], [7], [4], and [6] provide the notation and terminology for this paper.

We consider tree construction structures as extensions of 1-sorted structure as systems
⟨ a carrier, rules ⟩,
where the carrier is a set and the rules constitute a relation between the carrier and the carrier*.

Let us note that there exists a tree construction structure which is non empty and strict.

We consider context-free grammars as extensions of tree construction structure as systems
⟨ a carrier, an initial symbol, rules ⟩,
where the carrier is a set, the initial symbol is an element of the carrier, and the rules constitute a relation between the carrier and the carrier*.

Let us observe that there exists a context-free grammar which is non empty.

Let G be a tree construction structure. A symbol of G is an element of G. A string of G is an element of (the carrier of G)*.

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

Let D be a set. One can check that there exists an element of D^* which is empty.

Let D be a set. Then ε_D is an empty element of D^*.

Let D be a non empty set and 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 denotes a non empty tree construction structure, s denotes a symbol of G, and n, m denote strings of G.

Let us consider G, s and let n be a finite sequence. The predicate $s \Rightarrow n$ is defined as follows:

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

Let us consider G. The terminals of G yielding a set is defined by:

(Def. 2) The terminals of $G = \{s : \neg \bigvee_{n : \text{finite sequence}} s \Rightarrow n\}$.

The nonterminals of G yields a set and is defined by:

(Def. 3) The nonterminals of $G = \{s : \bigvee_{n : \text{finite sequence}} s \Rightarrow n\}$.

One can prove the following proposition

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

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

(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 are 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
- (i) $\operatorname{len} p \geq 1$,
- (ii) $p(1) = n$,
- (iii) $p(\operatorname{len} p) = m$, and
- (iv) 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 $a \Rightarrow b$.

We now state three propositions:

(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 G be a non empty context-free grammar. The language generated by G yielding a set is defined by the condition (Def. 6).

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

One can prove the following proposition

(9) Let G be a non empty context-free grammar and n be a string of G. Then $n \in$ 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 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.

Let a be a set. The functor $\{a \Rightarrow \varepsilon\}$ yields a strict context-free grammar and is defined as follows:

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

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

(Def. 8) The carrier 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 carrier 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 $\left\{\begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array}\right\} = \{\langle a, \langle b, a \rangle\rangle, \langle a, \emptyset\rangle\}$.

Let a be a set. One can check that $\{a \Rightarrow \varepsilon\}$ is non empty. Let b be a set. One can verify that $\{a \Rightarrow b\}$ is non empty and $\left\{\begin{array}{c} a \Rightarrow ba \\ a \Rightarrow \varepsilon \end{array}\right\}$ is non empty.

Let D be a non empty set. The total grammar over D yielding a strict context-free grammar is defined by the conditions (Def. 10).

(Def. 10)(i) The carrier of the total grammar over $D = D \cup \{D\}$,

(ii) the initial symbol of the total grammar over $D = D$, and

(iii) the rules of the total grammar over $D = \{\langle D, \langle d, D \rangle\rangle; d$ ranges over elements of D: $d = d\} \cup \{\langle D, \emptyset\rangle\}$.

Let D be a non empty set. One can verify that the total grammar over D is non empty.

In the sequel a, b are sets and D is a non empty set.

Next we state several propositions:

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

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

(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 $D = D^*$.

Let I_1 be a non empty context-free grammar. We say that I_1 is effective if and only if the conditions (Def. 11) are satisfied.

(Def. 11)(i) The language generated by I_1 is non empty,

(ii) the initial symbol of $I_1 \in$ the nonterminals of I_1, and

(iii) for every symbol s of I_1 such that $s \in$ the terminals of I_1 there exists a string p of I_1 such that $p \in$ the language generated by I_1 and $s \in \operatorname{rng} p$.

Let I_1 be a context-free grammar. We say that I_1 is finite if and only if:

(Def. 12) The rules of I_1 are finite.

One can check that there exists a non empty context-free grammar which is effective and finite.

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

Let X be a set and let Y be a non empty set. Note that there exists a relation between X and Y which is function-like.

Let X, Y be non empty sets, 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 R be a binary relation. The functor R^* yielding a binary relation is defined by the condition (Def. 13).

(Def. 13) Let x, y be sets. Then $\langle x, y \rangle \in R^*$ if and only if the following conditions are satisfied:

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

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

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

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

(Def. 14) For every element p of X^* holds $f^*(p) = f \cdot p$.

In the sequel R denotes a binary relation.

One can prove the following 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 non empty context-free grammar, let X be a non empty set, and let f be a function from the carrier of G into X. The functor $G(f)$ yielding a strict context-free grammar is defined as follows:

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

We now state the proposition

(19) For all 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. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/nat_1.html.

[2] Grzegorz Bancerek and Krzysztof Hryniewiecki. Segments of natural numbers and finite sequences. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/finseq_1.html.

[3] Czesław Byliński. Functions and their basic properties. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/funct_1.html.

[4] Czesław Byliński. Functions from a set to a set. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/funct_2.html.

[5] Czesław Byliński. Some basic properties of sets. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/zfmisc_1.html.

[6] Czesław Byliński. Finite sequences and tuples of elements of a non-empty sets. *Journal of Formalized Mathematics*, 2, 1990. http://mizar.org/JFM/Vol2/finseq_2.html.

[7] Agata Darmochwał. Finite sets. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/finset_1.html.

[8] Andrzej Trybulec. Tarski Grothendieck set theory. *Journal of Formalized Mathematics*, Axiomatics, 1989. http://mizar.org/JFM/Axiomatics/tarski.html.

[9] Zinaida Trybulec. Properties of subsets. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/subset_1.html.

[10] Edmund Woronowicz. Relations and their basic properties. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/relat_1.html.

[11] Edmund Woronowicz. Relations defined on sets. *Journal of Formalized Mathematics*, 1, 1989. http://mizar.org/JFM/Vol1/relset_1.html.

Received February 21, 1992

Published January 2, 2004