A Note on Scattered Context Grammars with Non-Context-Free Components

Jiří Techet   Tomáš Masopust   (Alexander Meduna)

Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic

Modern Formal Language Theory, 2007

Scattered Context Grammar

Scattered Context Grammar (SC Grammar)

G = (V, T, P, S), where

V is a finite alphabet

T is a set of terminals, $T \subset V$

S is the start symbol, $S \in V - T$

P is a finite set of productions of the form

$$(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n),$$

where $A_1, \dots, A_n \in V - T$, $x_1, \dots, x_n \in V^*$

Propagating Scattered Context Grammar (PSC Grammar)

- each $(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)$ satisfies $x_1, \dots, x_n \in V^+$

SC Grammar—Derivation Step

Derivation Step

For $(A_1, \ldots, A_n) \rightarrow (x_1, \ldots, x_n) \in P$ and

$$u = u_1 A_1 \ldots u_n A_n u_{n+1}$$
$$v = u_1 x_1 \ldots u_n x_n u_{n+1}$$

we write $u \Rightarrow v \; [(A_1, \ldots, A_n) \rightarrow (x_1, \ldots, x_n)]$

Generated Language

$L(G) = \{x \in T^* : S \Rightarrow^* x\}$

Generative Power

- $\mathscr{L}(SC) = \mathscr{L}(RE)$
- $\mathscr{L}(CF) \subset \mathscr{L}(PSC) \subseteq \mathscr{L}(CS)$

SC Grammar—Example

Production Length

- $\text{len}((A_1, \ldots, A_n) \rightarrow (x_1, \ldots, x_n)) = |A_1 \ldots A_n| = n$

Example

SC grammar $G = (\{A, B, C, S, a, b, c\}, \{a, b, c\}, P, S)$ with

$$\begin{aligned} P = \{&(S) \rightarrow (ABC), \\ &(A, B, C) \rightarrow (aA, bB, cC), \\ &(A, B, C) \rightarrow (\varepsilon, \varepsilon, \varepsilon)\} \end{aligned}$$

$S \Rightarrow ABC \Rightarrow aAbBcC \Rightarrow aaAbbBccC \Rightarrow aabbcc$
$L(G) = \{a^n b^n c^n \,:\, n \geq 0\}$

Linear Scattered Context Grammars

Linear Scattered Context Grammar

- scattered context grammar $G = (V, T, P, S)$
- P is a finite set of productions of the following two forms:
  1. $(S) \rightarrow (x_1 A_1 \dots x_k A_k x_{k+1})$, where $A_i \in (V - T) - \{S\}$, $x_i \in T^*$ for all $1 \leq i \leq k$, for some $k \geq 1$,
  2. $(A_1, \dots, A_k) \rightarrow (z_1, \dots, z_k)$, where $A_i \in (V - T) - \{S\}$, and either
     - $z_i = x_i B_i y_i$, where $x_i, y_i \in T^*$, $B_i \in (V - T) - \{S\}$, or
     - $z_i \in T^*$

     for all $1 \leq i \leq k$, for some $k \geq 1$

Degree n of Linear Scattered Context Grammar

- $(S) \rightarrow (x_1 A_1 \dots x_n A_n x_{n+1}) \in P$ satisfies $n \geq m$ for all $(S) \rightarrow (y_1 A_1 \dots y_m A_m y_{m+1}) \in P$
- for each $p \in P$, $\text{len}(p)$ is constant for every grammar ($\text{len}(p)$ does not depend on the degree)

Right-Linear Scattered Context Grammars

Right-Linear Scattered Context Grammar

- linear scattered context grammar $G = (V, T, P, S)$
- P is a finite set of productions of the following two forms:
  1. $(S) \rightarrow (x_1 A_1 \dots x_k A_k)$, where $A_i \in (V - T) - \{S\}$, $x_i \in T^*$ for all $1 \leq i \leq k$, for some $k \geq 1$,
  2. $(A_1, \dots, A_k) \rightarrow (z_1, \dots, z_k)$, where $A_i \in (V - T) - \{S\}$, and either
     - $z_i = x_i B_i$, where $x_i \in T^*$, $B_i \in (V - T) - \{S\}$, or
     - $z_i \in T^*$

     for all $1 \leq i \leq k$, for some $k \geq 1$

Language Families

- $\mathscr{L}(SC, LIN, n)$ – linear scattered context grammars of degree n
- $\mathscr{L}(SC, RLIN, n)$ – right-linear scattered context grammars of degree n

Main Results I

Theorem

For each $n \geq 1$,

$$
\begin{aligned}
\mathscr{L}(SC, LIN, n) &\subset \mathscr{L}(SC, LIN, n+1),\\
\mathscr{L}(SC, RLIN, n) &\subset \mathscr{L}(SC, RLIN, n+1),\\
\mathscr{L}(SC, RLIN, n) &\subset \mathscr{L}(SC, LIN, n).
\end{aligned}
$$

- $\mathscr{L}(SC, LIN) = \bigcup_{n=1}^{\infty} \mathscr{L}(SC, LIN, n)$
- $\mathscr{L}(SC, RLIN) = \bigcup_{n=1}^{\infty} \mathscr{L}(SC, RLIN, n)$

Theorem

$$
\begin{aligned}
&\mathscr{L}(SC, LIN) \subset \mathscr{L}(PSC),\ \mathscr{L}(CF) - \mathscr{L}(SC, LIN) \neq \emptyset,\\
&\mathscr{L}(SC, RLIN) \subset \mathscr{L}(PSC),\ \mathscr{L}(CF) - \mathscr{L}(SC, RLIN) \neq \emptyset,\\
&\mathscr{L}(SC, RLIN) \subset \mathscr{L}(SC, LIN).
\end{aligned}
$$

Main Results II

Theorem (Positive Closure Properties)

Each family $\mathscr{L}(SC, LIN, n)$ and $\mathscr{L}(SC, RLIN, n)$, where $n \geq 1$, is closed under union, reversal, homomorphism, inverse homomorphism, substitution with regular languages, concatenation with regular languages, intersection with regular languages, left and right quotient by regular languages. $\mathscr{L}(SC, LIN)$ and $\mathscr{L}(SC, RLIN)$ are closed under concatenation.

Theorem (Negative Closure Properties)

Each family $\mathscr{L}(SC, LIN, n)$, where $n \geq 1$, is not closed under concatenation with linear languages. Each family $\mathscr{L}(SC, RLIN, n)$, where $n \geq 1$, is not closed under concatenation with $\mathscr{L}(SC, RLIN, 2)$. $\mathscr{L}(SC, LIN)$ and $\mathscr{L}(SC, LIN)$ are not closed under intersection, complement and Kleene star. $\mathscr{L}(SC, LIN)$ is not closed under substitution with linear languages. $\mathscr{L}(SC, RLIN)$ is not closed under substitution with $\mathscr{L}(SC, RLIN, 2)$.

Conclusion and Open Problems

- The proof of the previous theorems is based on the proof of the equivalence of (right) linear sattered context grammars and (right) linear simple matrix grammars
- We may want to know what is the power of scattered context grammars with context-sensitive and unrestricted components; clearly:
  - $\mathscr{L}(SC, CS) = \mathscr{L}(CS)$
  - $\mathscr{L}(SC, RE) = \mathscr{L}(RE)$
- Concerning the power of scattered context grammars, there remains the original open problem:

$$\mathscr{L}(CS) - \mathscr{L}(PSC) \stackrel{?}{=} \emptyset$$