# SIMPLE SEMI-CONDITIONAL ET0L GRAMMARS

Martin ŠVEC, Doctoral Degree Programme (1)
Dept. of Information Systems, FIT, BUT
E-mail: svecm@fit.vutbr.cz

Supervised by: Dr. Alexander Meduna

**ABSTRACT**

The present contribution introduces and studies simple semi-conditional ET0L grammars (SSC-ET0L grammars for short). It is demonstrated that SSC-ET0L grammars with permitting and forbidding conditions of length no more than one and two, respectively, generate all the family of recursively enumerable languages; if erasing productions are not allowed, these grammars generate precisely the family of context-sensitive languages.

## 1 INTRODUCTION

So far the language theory has defined and investigated a number of sequential grammars possessing an additional regulation mechanism based on permitting and forbidding context conditions (see [3], [4]). Introducing this regulation to the area of parallel grammars, Meduna and Švec have recently defined so-called forbidding ET0L grammars (see [5]). In this contribution, simple semi-conditional ET0L grammars, another conditional version of ET0L grammars, is presented. By analogy with sequential simple semi-conditional grammars, productions of SSC-ET0L grammars have attached no more than one context condition. We demonstrate that SSC-ET0L grammars whose permitting and forbidding conditions are of length no more than one and two, respectively, characterize all the family of recursively enumerable languages. However, if erasing productions are not allowed, SSC-ET0L grammars generate only the family of context-sensitive languages.

## 2 DEFINITIONS

This contribution assumes that the reader is familiar with the language theory (see [2]).

Let $V$ be an alphabet. $V^*$ denotes the free monoid generated by $V$ under the operation of concatenation. Let $\varepsilon$ be the unit of $V^*$ and $V^+ = V^* - \{\varepsilon\}$. Given a word, $w \in V^*$, $|w|$ represents the length of $w$ and $\mathrm{alph}(w)$ denotes the set of all symbols occuring in $w$. For every $w \in V^+$, $\mathrm{sub}(w)$ is the set of all nonempty substrings of $w$. Furthermore, $\mathrm{sub}(w,k) =$

$\{x \in \text{sub}(w) : |x| \leq k\}$. Let first$(w)$ denote the leftmost symbol of $w$. Given a pair of integers, $r$ and $s$, $\max(r,s)$ denotes the maximal value of $r$ and $s$.

A *simple semi-conditional ET0L grammar* (*SSC-ET0L grammar* for short) is defined as a $t$+3-tuple, $G = (V, T, P_1, \ldots, P_t, S)$, $t \geq 1$, where $V$, $T$, and $S$ are the total alphabet, the terminal alphabet ($T \subset V$), and the axiom ($S \in V - T$), respectively. Each $P_i$ is a finite set of productions of the form $(a \rightarrow x, \alpha, \beta)$ with $a \in V$, $x \in V^*$, $\alpha, \beta \in V^+ \cup \{0\}$, and $0 \in \{\alpha, \beta\}$, where 0 is a special symbol; if $\alpha = 0$ or $\beta = 0$, the corresponding condition is missing. SSC-ET0L grammar without erasing productions is said to be *propagating* (*SSC-EPT0L grammar* for short). $G$ has *degree* $(r,s)$, where $r$ and $s$ are natural numbers, if for every $i \in \{1, \ldots, t\}$ and $(a \rightarrow x, \alpha, \beta) \in P_i$, $\alpha \neq 0$ implies $|\alpha| \leq r$ and $\beta \neq 0$ implies $|\beta| \leq s$. Let $u, v \in V^*$, $u = a_1 a_2 \ldots a_q$, $v = v_1 v_2 \ldots v_q$, $q = |u|$, $a_j \in V$, $v_j \in V^*$, and $p_1, p_2, \ldots, p_q$ is a sequence of productions $p_j = (a_j \rightarrow v_j, \alpha_j, \beta_j) \in P_i$ for all $j = 1, \ldots, q$ and some $i \in \{1, \ldots, t\}$. If for every $\alpha_j \neq 0$, $\alpha_j \in \text{sub}(u)$, and for every $\beta_j \neq 0$, $\beta_j \notin \text{sub}(u)$, then $u$ directly derives $v$ according to $p_1, p_2, \ldots, p_q$ in $G$, denoted by $u \Rightarrow_G v\ [p_1, p_2, \ldots, p_q]$. The language of $G$ is defined as $L(G) = \{x \in T^* : S \Rightarrow_G^* x\}$. If $t = 1$ then $G$ is called an SSC-E0L grammar. If $G$ is a propagating SSC-E0L grammar then $G$ is said to be an SSC-EP0L grammar. The families of languages defined by SSC-EPT0L, SSC-ET0L, SSC-EP0L, and SSC-E0L grammars of degree $(r,s)$, and by SSC-EPT0L, SSC-ET0L, SSC-EP0L and SSC-E0L grammars of any degree are denoted by **SSC-EPT0L**$(r,s)$, **SSC-ET0L**$(r,s)$, **SSC-EP0L**$(r,s)$, **SSC-E0L**$(r,s)$, **SSC-EPT0L**, **SSC-ET0L**, **SSC-EP0L**, and **SSC-E0L**, respectively. The families of context-free, context-sensitive, and recursively enumerable languages are denoted by **CF**, **CS**, and **RE**, respectively.

## 3 RESULTS

Let us investigate the generative power of SSC-ET0L grammars with erasing productions. First, we establish a normal form of recursively enumerable languages.

**Lemma 1.** *Every $L \in$ **RE** can be generated by a phrase-structure grammar, $G = (N_{CF} \cup N_{CS} \cup T, T, P, S)$, where $N_{CF}$, $N_{CS}$, and $T$ are pairwise disjoint alphabets, and every production in $P$ is either of the form $AB \rightarrow AC$ or $A \rightarrow x$, where $B \in N_{CS}$, $A, C \in N_{CF}$, $x \in \{\varepsilon\} \cup N_{CS} \cup T \cup N_{CF}^2$.*

*Proof.* To get the required forms of productions, modify Penttonen normal form of phrase-structure grammars, see [6]. □

The following lemma proves that every recursively enumerable language can be defined by an SSC-E0L grammar of degree $(1,2)$.

**Lemma 2.** **RE** $\subseteq$ **SSC-E0L**$(1,2)$.

*Proof.* Let $G = (N_{CF} \cup N_{CS} \cup T, T, P, S)$ be a phrase-structure grammar of the form of Lemma 1. Then, let $V = N_{CF} \cup N_{CS} \cup T$ and $m$ be the cardinality of $V$. Let $f$ be an arbitrary (but fixed) bijection from $V$ to $\{1, \ldots, m\}$ and $f^{-1}$ be the inverse of $f$. Set

$$\begin{aligned} M &= \{\#\} \cup \{\langle A,B,C\rangle : AB \rightarrow AC \in P, A, C \in N_{CF}, B \in N_{CS}\} \cup \\ &\quad \{\langle A,B,C,i\rangle : AB \rightarrow AC \in P, A, C \in N_{CF}, B \in N_{CS}, 1 \leq i \leq m+2\} \\ W &= \{[A,B,C] : AB \rightarrow AC \in P, A, C \in N_{CF}, B \in N_{CS}\}. \end{aligned}$$

Next, construct an SSC-E0L grammar of degree $(1,2)$, $G' = (V', T, P', S')$, where $V' = V \cup M \cup W \cup \{S'\}$ ($V$, $M$, $W$, and $\{S'\}$ are pairwise disjoint). The set of productions, $P'$, is defined in the following way:

1. add $(S' \to \#S, 0, 0)$ to $P'$;
2. if $A \to x \in P$, $A \in N_{CF}$, $x \in \{\varepsilon\} \cup N_{CS} \cup T \cup N_{CF}^2$, add $(A \to x, \#, 0)$ to $P'$;
3. for all $AB \to AC \in P$, $A, C \in N_{CF}$, $B \in N_{CS}$, add these productions to $P'$:
   (a) $(\# \to \langle A,B,C \rangle, 0, 0)$, $(B \to [A,B,C], \langle A,B,C \rangle, 0)$,
   $(\langle A,B,C \rangle \to \langle A,B,C,1 \rangle, 0, 0)$, $([A,B,C] \to [A,B,C], 0, \langle A,B,C,m+2 \rangle)$;
   (b) $(\langle A,B,C,i \rangle \to \langle A,B,C,i+1 \rangle, 0, f^{-1}(i)[A,B,C])$, where $1 \leq i \leq m$, $i \neq f(A)$;
   (c) $(\langle A,B,C,f(A) \rangle \to \langle A,B,C,f(A)+1 \rangle, 0, 0)$;
   (d) $(\langle A,B,C,m+1 \rangle \to \langle A,B,C,m+2 \rangle, 0, [A,B,C]^2)$,
   $(\langle A,B,C,m+2 \rangle \to \#, 0, \langle A,B,C,m+2 \rangle [A,B,C])$;
   (e) $([A,B,C] \to C, \langle A,B,C,m+2 \rangle, 0)$;
4. for all $X \in V$, add $(X \to X, 0, 0)$ to $P'$;
5. add $(\# \to \#, 0, 0)$ and $(\# \to \varepsilon, 0, 0)$ to $P'$.

Let us explain how $G'$ works. During the simulation of a derivation from $G$, every sentential form starts with an auxiliary symbol from $M$, called master. This symbol determines current simulation mode and controls the next derivation step. Initially, the master is set to #. In this mode, $G'$ simulates context-free productions (see (2)); notice that symbols from $V$ can always be rewritten to itself by (4). To start the simulation of a non-context-free production of the form $AB \to AC$, $G'$ rewrites the master to $\langle A,B,C \rangle$. In the following step, chosen occurences of $B$ are rewritten to $[A,B,C]$; all other productions except (4) are blocked. At the same time, the master is rewritten to $\langle A,B,C,i \rangle$ with $i = 1$. Then, $i$ is incremented by one as long as $i$ is less or equal to the cardinality of $V$. Simultaneously, master's conditions test that for every $i$ such that $f(i) \neq A$, no $f(i)$ appears as the left neighbor of any occurence of $[A,B,C]$. Finally, $G'$ checks that there are no two adjoining $[A,B,C]$ and $[A,B,C]$ does not appear as the right neighbor of the master. At this point, the left neighbors of $[A,B,C]$ are necessarily equal to $A$ and every occurence of $[A,B,C]$ is rewritten to $C$. In the same derivation step, the master is rewritten to #.

Observe that in every derivation step, the master enables only a given subset of productions according to the current mode. Indeed, it is not possible to combine context-free and non-context-free simulation modes. Furthermore, no two different non-context-free productions can be simulated at the same time. The simulation ends when # is erased by $(\# \to \#, 0, 0)$. After that, no next production modifying the sentential form can be used.

To establish $L(G) = L(G')$, we should formally prove that for every $w \in T^*$,

$$S \Rightarrow_G^* w \quad \text{if and only if} \quad S' \Rightarrow_{G'}^* w.$$

However, due to limited number of pages of the contribution, the proof is omitted. $\square$

**Lemma 3.** **SSC-ET0L**$(r,s) \subseteq$ **RE** *for any* $r,s \geq 0$.

*Proof.* Of course, this lemma follows from Church's thesis. However, let us demonstrate an effective algorithm proving the inclusion.

For $r = 0$ and $s = 0$, we have **SSC-ET0L**$(0,0) =$ **ET0L** $\subset$ **RE**. Next, assume that $r$ and/or $s$ are nonzero. Let $L$ be a language generated by an SSC-ET0L grammar $G = (V, T, P_1, \ldots, P_t, S)$ of degree $(r,s)$, for some $r,s \geq 0$, $r+s > 0$, $t \geq 1$. Set $k = \max(r,s)$. Let $M = \{x \in V^+ : |x| \leq k\}$. For every $P_i$, $1 \leq i \leq t$, set $\mathrm{cf}(P_i) = \{a \to z : (a \to z, \alpha, \beta) \in P_i,\ a \in V,\ z \in V^*\}$. Then, set $N_F = \{\langle X, x\rangle : X \subseteq M,\ x \in M\}$, $N_T = \{\lfloor X \rfloor : X \subseteq M\}$, $N_B = \{\lceil Q \rceil : Q \subseteq \mathrm{cf}(P_i),\ 1 \leq i \leq t\}$, $V' = N_F \cup N_T \cup N_B \cup \{\rhd, \lhd, \$, S'\}$. Construct the phrase-structure grammar $G' = (V', T, P', S')$ with the finite set of productions $P'$ defined as follows:

1. add $S' \to \rhd\langle\emptyset, \varepsilon\rangle S \lhd$ to $P'$;
2. for all $X \subseteq M$, $x \in (V^k \cup \{\varepsilon\})$, and $y \in V^k$, add $\langle X, x\rangle y \to y\langle X \cup \mathrm{sub}(xy,k), y\rangle$ to $P'$;
3. for all $X \subseteq M$, $x \in (V^k \cup \{\varepsilon\})$, $y \in V^*$, $|y| \leq k$, add $\langle X, x\rangle y\lhd \to y\lfloor X \cup \mathrm{sub}(xy,k)\rfloor\lhd$;
4. for all $X \subseteq M$ and $Q \subseteq \mathrm{cf}(P_i)$, $i = 1, \ldots, t$, such that for every $a \to z \in Q$, there exists $(a \to z, \alpha, \beta) \in P_i$, where $\alpha \neq 0$ implies $\alpha \in X$ and $\beta \neq 0$ implies $\beta \notin X$, add $\lfloor X \rfloor\lhd \to \lceil Q \rceil\lhd$ to $P'$;
5. for every $Q \subseteq \mathrm{cf}(P_i)$ for some $i \in \{1, \ldots, t\}$, $a \in V$, and $z \in V^*$ such that $a \to z \in Q$, add $a\lceil Q \rceil \to \lceil Q \rceil z$ to $P'$;
6. for all $Q \subseteq \mathrm{cf}(P_i)$ for some $i = \{1, \ldots, t\}$, add $\rhd\lceil Q \rceil \to \rhd\langle\emptyset, \varepsilon\rangle$ to $P'$;
7. add $\rhd\langle\emptyset, \varepsilon\rangle \to \$$, $\$a \to a\$$ for every $a \in T$, and $\$\lhd \to \varepsilon$ to $P'$.

According to the definition of $P'$, it can be shown that every successful derivation in $G'$ is of the form $S' \Rightarrow_{G'} \rhd\langle\emptyset, \varepsilon\rangle S\lhd \Rightarrow^+_{G'} \rhd\langle\emptyset, \varepsilon\rangle x\lhd \Rightarrow_{G'} \$x\lhd \Rightarrow^{|x|}_{G'} x\$\lhd \Rightarrow_{G'} x$ such that $x \in T^*$ and during $\rhd\langle\emptyset, \varepsilon\rangle S\lhd \Rightarrow^+_{G'} \rhd\langle\emptyset, \varepsilon\rangle x\lhd$, every sentential form $w$ satisfies $w \in \{\rhd\}H^+\{\lhd\}$, where $H \subseteq V' - \{\rhd, \lhd, \$, S'\}$. Next, let $x \Rightarrow^{\oplus}_{G'} y$ denote a derivation $x \Rightarrow^+_{G'} y$ such that $x = \rhd\langle\emptyset, \varepsilon\rangle u\lhd$, $y = \rhd\langle\emptyset, \varepsilon\rangle v\lhd$, $u, v \in V^*$, and there is no other occurence of a string of the form $\rhd\langle\emptyset, \varepsilon\rangle z\lhd$, $z \in V^*$, during $x \Rightarrow^+_{G'} y$. Then, it holds that for every $u, v \in V^*$, $\rhd\langle\emptyset, \varepsilon\rangle u\lhd \Rightarrow^{\oplus}_{G'} \rhd\langle\emptyset, \varepsilon\rangle v\lhd$ if and only if $u \Rightarrow_G v$. Informally, this statement tells us that every derivation step $u \Rightarrow_G v$ in $G$ is simulated by a derivation $\rhd\langle\emptyset, \varepsilon\rangle u\lhd \Rightarrow^{\oplus}_{G'} \rhd\langle\emptyset, \varepsilon\rangle v\lhd$ in $G'$. The simulation consists of two phases. During the first, forward phase, $G'$ scans $u$ to get all nonempty substrings of length $k$ or less. By repeatedly using productions $\langle X, x\rangle y \to y\langle X \cup \mathrm{sub}(xy,k), y\rangle$, $X \subseteq M$, $x, y \in V^k$, the occurence of a symbol with form $\langle X, x\rangle$ is moved towards the end of the sentential form. Simultaneously, the substrings of $u$ are collected in $X$. The second, backward phase simulates rewriting of all symbols in $u$ in parallel. From the above observations, the reader can prove that $S' \Rightarrow^+_{G'} \rhd\langle\emptyset, \varepsilon\rangle x\lhd$ if and only if $S \Rightarrow^*_G x$, for all $x \in V^*$. Then, according to the form of successful derivations, we get for each $x \in T^*$, $S' \Rightarrow^+_{G'} x$ if and only if $S \Rightarrow^*_G x$, and the lemma holds. □

Inclusions established in Lemmas 2 and 3 result in the following theorem:

**Theorem 1.** $\mathbf{SSC\text{-}E0L}(1,2) = \mathbf{SSC\text{-}ET0L}(1,2) = \mathbf{SSC\text{-}E0L} = \mathbf{SSC\text{-}ET0L} = \mathbf{RE}$.

*Proof.* From Lemmas 2 and 3, $\mathbf{RE} \subseteq \mathbf{SSC\text{-}E0L}(1,2)$ and $\mathbf{SSC\text{-}ET0L}(r,s) \subseteq \mathbf{RE}$ for any $r,s \geq 0$. By the definitions, we have $\mathbf{SSC\text{-}E0L}(1,2) \subseteq \mathbf{SSC\text{-}ET0L}(1,2) \subseteq \mathbf{SSC\text{-}ET0L}$ and $\mathbf{SSC\text{-}E0L}(1,2) \subseteq \mathbf{SSC\text{-}E0L} \subseteq \mathbf{SSC\text{-}ET0L}$. Thus, the theorem holds. □

By analogy with Lemmas 1, 2, and 3, the reader can establish that propagating SSC-E0L grammars of degree $(1,2)$ generate the family of context-sensitive languages:

**Theorem 2.** $\mathbf{CS} = \mathbf{SSC\text{-}EPT0L}(1,2) = \mathbf{SSC\text{-}EP0L}(1,2) = \mathbf{SSC\text{-}EPT0L} = \mathbf{SSC\text{-}EP0L}$.

Theorems 1 and 2 imply the following relationships of investigated language families:

$$\begin{array}{c}
\mathbf{CF} \\
\subset \\
\mathbf{SSC\text{-}EP0L}(0,0) = \mathbf{SSC\text{-}E0L}(0,0) = \mathbf{EP0L} = \mathbf{E0L} \\
\subset \\
\mathbf{SSC\text{-}EPT0L}(0,0) = \mathbf{SSC\text{-}ET0L}(0,0) = \mathbf{EPT0L} = \mathbf{ET0L} \\
\subset \\
\mathbf{SSC\text{-}EP0L}(1,2) = \mathbf{SSC\text{-}EPT0L}(1,2) = \mathbf{SSC\text{-}EP0L} = \mathbf{SSC\text{-}EPT0L} = \mathbf{CS} \\
\subset \\
\mathbf{SSC\text{-}E0L}(1,2) = \mathbf{SSC\text{-}ET0L}(1,2) = \mathbf{SSC\text{-}E0L} = \mathbf{SSC\text{-}ET0L} = \mathbf{RE}.
\end{array}$$


## ACKNOWLEDGEMENTS

This contribution was worked out under the research intent of the Ministry of Education of the Czech Republic no. CEZ: J22/98:262200012.


## REFERENCES


[1] Meduna, A., Gopalaratnam, A.: On Semi-Conditional Grammars with Productions Having either Forbidding or Permitting Conditions, Acta Cybernetica 11 (1994), 307–323.

[2] Meduna, A.: Automata and Languages: Theory and Applications, Springer, London, 2000.

[3] Meduna, A., Švec, M.: Reduction of Simple Semi-Conditional Grammars with Respect to the Number of Conditional Productions, Acta Cybernetica 15 (2002) 353–360.

[4] Meduna, A., Švec, M.: Descriptional Complexity of Generalized Forbidding Grammars, Intern. J. of Computer Math. 80(1) (2003), 11–17.

[5] Meduna, A., Švec, M.: Forbidding ET0L Grammars (submitted).

[6] Penttonen, M.: One-Sided and Two-Sided Context in Formal Grammars, Inform. Contr. 25 (1974), 371–392.

[7] Salomaa, A.: Formal Languages, Academic Press, New York, 1973.