Internal Contextual Grammars:
Minimal, Maximal, and Scattered Use of Selectors

C. Martín-Vide, A. Mateescu, J. Miquel-Vergés, Gh. Păun

cmv@astor.urv.es, mateescu@sara.cc.utu.fi, verges@lingua.fil.ub.es,
paun@sara.cc.utu.fi

Research Group in Mathematical Linguistics and Language Engineering (GRLMC)
Rovira i Virgili University, Pça. Imperial Tàrraco, 1, 43005 Tarragona, Spain

Department of Mathematics, Turku University, 20500 Turku, Finland

Institute of Mathematics of the Romanian Academy
PO Box 1 - 764, 70700 București, Romania

Abstract. We consider internal contextual grammars (with finite and with
regular selection sets) with the derivation restricted in the following way: when
a word is used as a selector, then no subword/superword of it can be used as a
selector. This can be considered both locally (with respect to the chosen selec-
tion set) or globally (with respect to all selection sets). The generative capacity
of such grammars is investigated (comparing each other the above mentioned
variants and with the class of unrestricted internal contextual grammars). Fi-
nally, grammars with selectors checked as scattered subwords of the derived
words are considered.

1. Introduction

The contextual grammars were introduced in [5] as “intrinsic grammars”, based on the
operation of adjoining contexts (pairs of words) to given strings, according to a selection
defined with respect to certain words (selectors) associated to the contexts; such a selector
must be present in the processed phrase and the adjoined context will be added at the
ends of the specified occurrence of the selector. The study of contextual grammars is quite
related to basic combinatorics on words. A series of variants were considered in [9], [13],
[14], [15]; the reader is also referred to [2], [3] for recent results in this area.

A natural restriction in using a pair (selector-word, context) is to impose some mini-
mality/maximality conditions on the selector: a context is added only to a selector which
is minimal (no subword of it can be used as a selector) or maximal (no superword of it
can be used as a selector), either with respect to the specified set of selector-words or
with respect to the whole grammar (precise definitions can be found in the following sec-
tion). The influence of these restrictions about the generative power of internal contextual
grammars is somewhat unexpected: a decrease of power is obtained in the case of regular
selectors (but not in the case of finite selectors) when the minimal restriction is used, the

1 Research partially supported by the Academy of Finland, Project 11281
global minimal use of selectors is weaker than the local one, whereas for the maximal restriction the results look different.

Another natural modification is to check the selectors not as blocks of the rewritten words but as scattered subwords. As a consequence of a result [4], the grammars with any type of selectors checked in the scattered way are equivalent with grammars with finite selectors, a quite surprising result.

2. Definitions

As usual, we denote by $V^*$ the free monoid generated by an alphabet $V$; the empty word is denoted by $\lambda$ and the length of $x \in V^*$ is denoted by $|x|$. The families of finite, regular, context-free, and context-sensitive languages are denoted by $FIN, REG, CF, CS$, respectively. Further elements of formal language theory can be found in [16].

A contextual grammar (with choice) is a construct

$$G = (V, B, (D_1, C_1), \ldots, (D_n, C_n)),$$

where $V$ is an alphabet, $B$ is a finite language over $V$, $D_i$ are languages over $V$, and $C_i$ are finite subsets of $V^* \times V^*$, $1 \leq i \leq n$, $n \geq 1$. The sets $D_i$ are called selectors and the pairs in $C_i$ are called contexts. If each $D_i, 1 \leq i \leq n$, belongs to a specified family $F$ of languages, then we say that $G$ is with $F$ choice.

Two basic derivation relations can be associated to a contextual grammar as above: for $x, y \in V^*$ we write

$$x \Rightarrow_{ex} y \textit{ iff } x \in D_i, y = uxv, \text{ for } (u, v) \in C_i, 1 \leq i \leq n,$$

$$x \Rightarrow_{in} y \textit{ iff } x = x_1 x_2 x_3, x_2 \in D_i, y = x_1 ux_2 v x_3, \text{ for } (u, v) \in C_i, 1 \leq i \leq n.$$

The first relation defines the external derivation, the second one defines the internal derivation with respect to $G$. In [5] only the external case is considered; the internal contextual grammars are introduced in [12].

The language generated by a grammar $G$ as above in the $\alpha \in \{in, ex\}$ mode is

$$L_\alpha(G) = \{ z \in V^* \mid w \Rightarrow_\alpha^* z, \text{ for some } w \in B \},$$

where $\Rightarrow_\alpha^*$ is the reflexive and transitive closure of $\Rightarrow_\alpha$.

Consequently, $L_\alpha(G)$ contains all words in $B$ and all the words which can be obtained starting from words in $B$ (they play the role of axioms) and adding contexts (at the ends of the word for $\alpha = ex$ and in any place for $\alpha = in$) according to the selection imposed by the pairing $(D_i, C_i)$. Examples will appear in the proofs of the following section.

Natural variants of the relations $\Rightarrow_\alpha$ defined above are the next ones:

$$x \Rightarrow_{mi} y \textit{ iff } x = x_1 x_2 x_3, y = x_1 ux_2 v x_3, \text{ for } x_2 \in D_i, (u, v) \in C_i, 1 \leq i \leq n,$$

and there are no $x'_1, x'_2, x'_3 \in V^*$ such that

$$x = x'_1 x'_2 x'_3, x'_2 \in D_i, |x'_1| \geq |x_1|, |x'_2| \geq |x_3|, |x'_3| < |x_2|;$$

$$x \Rightarrow_{Mi} y \textit{ iff } x = x_1 x_2 x_3, y = x_1 ux_2 v x_3, \text{ for } x_2 \in D_i, (u, v) \in C_i, 1 \leq i \leq n,$$

and there are no $x'_1, x'_2, x'_3 \in V^*$ such that

$$x = x'_1 x'_2 x'_3, x'_2 \in D_i, |x'_1| \leq |x_1|, |x'_2| \leq |x_3|, |x'_3| > |x_2|.$$
We call \( \Rightarrow_{ml} \) a derivation in the minimal local mode and \( \Rightarrow_{MI} \) a derivation in the maximal local mode.

When the condition \( x' \in D_i \) in the previous definitions is replaced with \( x' \in D_j, 1 \leq j \leq n \), then we call the obtained relations minimal global and maximal global, and we write \( \Rightarrow_{mg}, \Rightarrow_{Mg} \) respectively. For \( \alpha \in \{ml, mg, MI, Mg\} \) we define

\[
L_\alpha(G) = \{ z \in V^* \mid w \Rightarrow^*_\alpha z \text{ for some } w \in B \}.
\]

The families of languages of the form \( L_\alpha(G) \), for \( G \) a contextual grammar with \( F \) choice, are denoted by \( CL_\alpha(F) \), \( \alpha \in \{in, ex, ml, mg, MI, Mg\} \). Here we consider \( F \in \{FIN, REG\} \). Ten families of languages are obtained in this way.

3. The power of grammars with minimal restrictions

The following relations are obvious:

**Lemma 1.** \( CL_\alpha(FIN) \subseteq CL_\alpha(REG) \) for all \( \alpha \in \{in, ml, mg, MI, Mg\} \).

By straightforward constructions (the minimality/maximality of selectors, which are members of regular sets, can be checked at the level of context-sensitive grammars), we obtain

**Lemma 2.** \( CL_\alpha(REG) \subseteq CS \) for all \( \alpha \in \{in, ml, mg, MI, Mg\} \).

Moreover, we have

**Lemma 3.** \( CL_{mg}(F) \subseteq CL_{ml}(F) \subseteq CL_{in}(F) \), \( F \in \{FIN, REG\} \).

**Proof.** Consider a grammar \( G = (V, B, (D_1, C_1), \ldots, (D_n, C_n)) \). If some \( D_i \) contains two words \( z, z' \) such that \( z \) is a proper subword of \( z' \), then \( z' \) is never used in a locally minimal derivation, hence it can be eliminated without modifying the generated language. Similarly when \( z \in D_i \) and \( z' \in D_j \) for the case of globally minimal derivations. Formally, we can construct the grammars

\[
G_g = (V, B, (D_1', C_1), \ldots, (D_n', C_n)),
\]

\[
G_l = (V, B, (D_1'', C_1), \ldots, (D_n'', C_n)),
\]

with

\[
D_i' = D_i - \bigcup_{j=1}^{n} (V^+D_jV^* \cup V^*D_jV^+), \quad 1 \leq i \leq n,
\]

\[
D_i'' = D_i - (V^+D_iV^* \cup V^*D_iV^+), \quad 1 \leq i \leq n.
\]

Clearly, if \( D_i \in F \), then \( D_i' \in F, D_i'' \in F \) (\( REG \) is closed under difference). Moreover,

\[
L_{mg}(G) = L_{ml}(G_g) = L_{in}(G_g), \quad \text{hence } CL_{mg}(F) \subseteq CL_{ml}(F), \quad \text{and } L_{ml}(G) = L_{in}(G_l), \quad \text{hence } CL_{ml}(F) \subseteq CL_{in}(F).
\]

For finite selectors we also have

**Lemma 4.** \( CL_{in}(FIN) \subseteq CL_{ml}(FIN) \).
Proof. Consider a grammar $G = (V, B, (D_1, C_1), \ldots, (D_n, C_n))$ with $D_i = \{x_{i,1}, \ldots, x_{i,k_i}\}, 1 \leq i \leq n$. We construct the grammar

$$G' = (V, B, (\{x_{1,1}\}, C_1), (\{x_{1,2}\}, C_1), \ldots, (\{x_{1,k_1}\}, C_1),
\ldots\ldots\ldots\ldots
(\{x_{n,1}\}, C_n), (\{x_{n,2}\}, C_n), \ldots, (\{x_{n,k_n}\}, C_n)).$$

The equality $L_{ml}(G') = L_{in}(G)$ is obvious (for grammars $H$ with singleton selectors we have $L_{ml}(H) = L_{in}(H)$).

Corollary. $CL_{in}(FIN) = CL_{ml}(FIN)$.

In the case of regular selectors, such a relation is not true:

Lemma 5. $CL_{in}(REG) \neq CL_{ml}(REG)$.

Proof. Consider the grammar

$$G = (\{a, b, c\}, \{abc\}, (b^+c, \{(ab, c)\})).$$

The language $L_{in}(G)$ is included in $\{a, b\}^+c^+$ and it contains all words of the form $a^m b^m c^m$ for $m \geq 1$.

Assume that $L_{in}(G) = L_{ml}(G')$ for some $G' = (\{a, b, c\}, B, (D_1, C_1), \ldots, (D_n, C_n))$. In order to generate words $a^m b^m c^m$ with arbitrarily large $m$, we need contexts of the form $(a^i b^j, c^i)$. Indeed, if we have a context $(a^i b^j, b^kc^j), i = j + k, k \geq 1$, then the associated selector must be some $D \subseteq b^*$ and in derivations only the minimal element of $D$, $b^*$, can be used; from a word $a^p b^p c^p$ with large enough $p$ we can produce $a^p b^b b^k c^i b^{p-s} c^p$, which is not in $L_{in}(G)$, a contradiction.

The selector of a context $(a^i b^j, c^i)$ (the case of contexts of the form $(a^i, b^j c^i)$ is similar) must be of the form $b^j c^k, j \geq 0, k \geq 0$. If $k = 0$, then again we can produce words with mixed occurrences of $b$ and $c$. Thus, we have to use selectors of the form $b^j c^k$ with $k \geq 1$. The set of such words which are incomparable is finite; if two words are comparable, then the greater one is useless and can be removed without modifying the generated languages. In conclusion, we can assume that all selectors of contexts $(a^i b^j, c^i)$ are finite. This implies that the set of words $a^m b^m c^m$ generated by $G'$ is finite, because the set of subwords $b^p$ of such words is finite, a contradiction.

As it is expected, in all cases grammars with regular selectors are more powerful than grammars with finite selectors.

Lemma 6. $CL_{mg}(REG) \neq CL_{in}(FIN)$.

Proof. Consider the grammar

$$G = (\{a, b\}, \{ab, \}, \{a^+b, \{(a, a)\}\}, (b^+a, \{(b, b)\})).$$

We obtain

$$L_{mg}(G) = L_{in}(G) = \{a^n b^n a^n b^n \mid n, m \geq 1\}.$$

In [13] (Theorem 7) it is proved that this language is not in $CL_{in}(FIN)$.

Corollary. All inclusions $CL_{\alpha}(FIN) \subset CL_{\alpha}(REG), \alpha \in \{in, ml, mg\}$, are strict.
For regular selectors, the first inclusion in Lemma 3 is proper:

**Lemma 7.** $CL_{ml}(FIN) - CL_{mg}(REG) \neq \emptyset$.

**Proof.** Consider the language

$$L = \{a^n b^m c^m a^n \mid n, m \geq 1\}.$$  

It is in the family $CL_{ml}(FIN)$ because it is generated by the grammar

$$G = (\{a, b, c\}, \{abcba\}, (c, \{(b, b)\}), (abcba, \{(a, a)\})).$$

However, $L \notin CL_{mg}(REG)$. Indeed, assume that $L = L_{mg}(G')$ for some $G' = (\{a, b, c\}, B, (D_1, C_1), \ldots, (D_n, C_n))$. In order to generate words $a^n b^m c^m a^n$ with arbitrarily large $m$ we need contexts of the form $(b^i, b^i)$ associated to words $b^p c^q$ with $p, q \geq 0$. In order to generate words as above with arbitrarily large $n$ we need contexts of the form $(a^i, a^i)$ associated to words $a^s b^g c^t a^i$; we must have $s + t \geq 1$, otherwise the symbols $a, b$ can be mixed. If in these words we have $(p, q) < (g, g)$ (componentwise), then the contexts increasing the number of $a$ occurrences cannot be used in the global minimal mode. In a derivation we first use a context $(a^i, a^i)$, associated with $a^s b^g c^t a^i$, and then a context $(b^i, b^i)$, associated with $b^p c^q$, then we must have $(g, g) = (p, q)$ and $s = t = 0$; then words with arbitrarily many subwords $ab$ can be generated, a contradiction. Therefore we have to use first the contexts of the form $(b^i, b^i)$ and then the contexts $(a^i, a^i)$. The number of selectors $b^p c^q$ which can be used is finite (the set of such incomparable words is finite), hence the values of $g$ in $a^s b^g c^t a^i$, selectors for $(a^i, a^i)$, must be finite, too. Consequently, words $a^n b^m c^m a^n$ with arbitrarily large $n$ can be generated only for bounded $m$. However, $L$ contains words with both $n$ and $m$ arbitrarily large, a contradiction. \hfill $\Box$

![Diagram 1](Diagram 1)

Synthesizing the above results, we obtain
Theorem 1. The relations in Diagram 1 hold; the arrows indicate strict inclusions, the dotted arrow indicates an inclusion not known to be proper. Families not linked by a path in this diagram are not necessarily incomparable.

Proof. All relations follow from the previous lemmas, excepting the inclusion $REG \subset CL_{in}(FIN)$, which is proved in [3].

Open problems. Is $REG$ included into $CL_{mg}(FIN)$?

4. The power of grammars with maximal restrictions

The same proof as for Lemma 4 shows that

Lemma 8. $CL_{in}(FIN) \subseteq CL_{Ml}(FIN)$.

It is an open problem whether the converse inclusion holds true. As it is expected, a counterpart of the corollary of Lemma 6 is true also for the maximal case.

Lemma 9. $CL_{\alpha}(REG) - CL_{\alpha}(FIN) \neq \emptyset, \alpha \in \{Ml, Mg\}$.

Proof. For the grammar $G$ in the proof of Lemma 6 we have $L_{Ml}(G) = L_{Mg}(G) = \{a^n b^m a^n b^m | n, m \geq 1\}$, a language which cannot be generated by a grammar with finite selectors (with finite selectors we cannot generate words $a^n b^m a^n b^m$ with both $n$ and $m$ arbitrarily large).

Lemma 10. $CL_{in}(FIN) - CL_{Mg}(REG) \neq \emptyset$.

Proof. The language $L = \{a^n c b^n c a^n | n, m \geq 1\}$ is in $CL_{in}(FIN)$, but it cannot be generated by a grammar in the global maximal mode. The proof is similar to that used for proving that the language $L$ in the proof of Lemma 7 is not in $CL_{mg}(REG)$, but more technical. We leave the details to the reader.

Surprisingly (different from the case of minimal restriction), we have

Lemma 11. $CL_{Mg}(REG) - (CL_{in}(REG) \cup CL_{Ml}(REG)) \neq \emptyset$.

Proof. The language $L = a^+ \cup \{a^n b^n | n \geq 1\}$, is not in $CL_{in}(REG)$ [9]; similarly, we can see that $L \notin CL_{Ml}(REG)$ (the contexts used for generating words $a^n$ with arbitrarily large $n$ can be used also for generating from $a^n b^n$ words of the form $a^p b^q$ with $p > q$, which are not in $L$). On the other hand, we have $L = L_{Mg}(G)$ for $G = (\{a, b\}, \{a, ab\}, (a, ((\lambda, a))), (a^+ b, \{(a, b)\}))$ (in the globally maximal mode, the context $(\lambda, a)$ cannot be used for rewriting words $a^n b^n$; moreover, the selector $ab$ does not appear in words of $a^+$.)

Lemma 12. $CL_{ex}(F) \subseteq CL_{Mg}(F), F \in \{FIN, REG\}$.

Proof. For $F = FIN$ the assertion is obvious, because $CL_{ex}(FIN) = FIN$.

Consider a grammar $G = (V, B, (D_1, C_1), \ldots, (D_n, C_n))$ with regular selectors $D_i, 1 \leq i \leq n$, and construct $G' = (V, B, (D_1, C_1), \ldots, (D_n, C_n), (D_{n+1}, \{(\lambda, \lambda)\}))$, where

$$D_{n+1} = V^* (\bigcup_{i=1}^n D_i) V^*.$$
If \( x_1x_2x_3 \xrightarrow{M_\delta} x_1u_2v_3 \) in \( G' \), then we must have \( x_1 = \lambda, x_3 = \lambda \), otherwise the production \((D_{n+1}, \{(\lambda, \lambda)\})\) must be used (it changes nothing, but prevents the use of other productions, when \( x_1 \neq \lambda \)).

Therefore, \( L_{M_\delta}(G') = L_{\text{ex}}(G) \), and \( CL_{\text{ex}}(\text{REG}) \subseteq CL_{M_\delta}(\text{REG}) \).

\[ \begin{diagram}
  \text{CS} & & \\
  \downarrow & & \\
  CL_{\text{in}}(\text{REG}) & CL_{M_1}(\text{REG}) & CL_{M_\delta}(\text{REG}) \\
  \downarrow & & \downarrow \\
  CL_{M_1}(\text{FIN}) & CF & \\
  \downarrow & & \\
  CL_{\text{in}}(\text{FIN}) & CL_{M_\delta}(\text{FIN}) & CL_{\text{ex}}(\text{REG}) \\
  \downarrow & & \downarrow \\
  \text{REG} & & \\
  \downarrow & & \\
  FIN = CL_{\text{ex}}(FIN) & & \\
\end{diagram} \]

Diagram 2

**Theorem 2.** The relations in Diagram 2 hold. The arrows have the same meaning as in Diagram 1. The families \( CL_{M_\delta}(\text{REG}) \) and \( CF \) are incomparable with families \( CL_{\text{in}}(F) \) and \( CL_{M_1}(F) \), \( F \in \{\text{FIN, REG}\} \).

**Proof.** Synthesize the relations in the previous lemmas.

\[ \square \]

5. **Scattered use of selectors**

For a contextual grammar \( G = (V, B, (D_1, C_1), \ldots, (D_n, C_n)) \) we can also consider the relation

\[ x \xrightarrow{s_c} y \text{ iff } x = x_1x_2x_2\ldots x_{k-1}x_kx_{k+1}, y = x_1u_2x_2x_2\ldots x_{k-1}x_ku_{k+1}, \]

for \( k \geq 1, x_j, y_j \in V^* \) for all \( j \), and

\[ z_1z_2\ldots z_k \in D_i(u, v) \in C_i, \text{ for some } i, 1 \leq i \leq n. \]

Hence the context \((u, v)\) is adjoined "around" the selector \( z_1z_2\ldots z_k \) which is present in \( x \) as a scattered subword. Note that each \( z_j, 1 \leq j \leq k \), can be empty, hence \( u, v \) are not necessarily adjacent to symbols in the selector.

We denote by \( L_{s_c}(G) \) the generated language and by \( CL_{s_c}(F), F \in \{\text{FIN, REG}\} \), the corresponding families of languages. Obviously, we have

**Lemma 13.** \( CL_{s_c}(\text{FIN}) \subseteq CL_{s_c}(\text{REG}) \).

**Lemma 14.** \( CL_{s_c}(\text{REG}) \subseteq CL_{\text{in}}(\text{REG}) \).
Proof. Let \( G = (V, B, (D_1, C_1), \ldots, (D_n, C_n)) \) be an internal contextual grammar with regular \( D_1, \ldots, D_n \). Construct the grammar

\[
G' = (V, B, (D_1 \uplus V^*, C_1), \ldots, (D_n \uplus V^*, C_n)),
\]

where \( \uplus \) is the shuffle operation: if \( x, y \in V^* \), then

\[
x \uplus y = \{ x_1y_1 \ldots x_ky_k \mid k \geq 1, x = x_1x_2 \ldots x_k, y = y_1y_2 \ldots y_k, x_i, y_i \in V^*, 1 \leq i \leq k \}.
\]

The family \( \text{REG} \) is closed under shuffle and, obviously, \( L_{\text{sc}}(G) \subseteq L_{\text{in}}(G') \). \( \square \)

Using the idea in the previous proof, we obtain the following surprising relation (the converse of Lemma 13).

**Lemma 15.** \( \text{CL}_{\text{sc}}(\text{REG}) \subseteq \text{CL}_{\text{sc}}(\text{FIN}) \).

**Proof.** For a grammar \( G = (V, B, (D_1, C_1), \ldots, (D_n, C_n)) \) with regular sets \( D_1, \ldots, D_n \), we construct \( G' = (V, B, (D'_1, C_1), \ldots, (D'_n, C_n)) \) as above, \( D'_i = D_i \uplus V^* \), \( 1 \leq i \leq n \), with \( L_{\text{sc}}(G) = L_{\text{in}}(G') \).

According to [4], for every language \( D \subseteq V^* \) (irrespective of the place in Chomsky hierarchy) there is a finite language \( E \subseteq D \) such that \( D \subseteq E \uplus V^* \). Consequently, \( D \uplus V^* \subseteq (E \uplus V^*) \uplus V^* = E \uplus V^* \). The converse inclusion follows from \( E \subseteq D \), hence \( D \uplus V^* = E \uplus V^* \).

Let \( E_1, \ldots, E_n \) be the finite languages associated in this way to \( D_1, \ldots, D_n \) and construct the grammar \( G'' = (V, B, (E_1, C_1), \ldots, (E_n, C_n)) \). We have \( L_{\text{sc}}(G'') = L_{\text{in}}(G') = L_{\text{sc}}(G) \), which completes the proof. \( \square \)

**Corollary.** \( \text{CL}_{\text{sc}}(\text{REG}) = \text{CL}_{\text{sc}}(\text{FIN}) \).

**Lemma 16.** \( ab^+a \not\in \text{CL}_{\text{sc}}(\text{REG}) \).

**Proof.** Assume that \( ab^+a = L_{\text{sc}}(G) \) for some \( G = (\{a, b\}, B, (D_1, C_1), \ldots, (D_n, C_n)) \). We have \( B \subseteq ab^+a \) and there is \( i \) such that \( C_i \) contains a context \( (u, v) \) with \( uv \in B^+ \) and \( (D_i, C_i) \) is effectively used in a derivation \( x \rightarrow y \). Clearly, \( x \) is of the form \( ab^+a \). Irrespective which is the used word of \( D_i \), we can also derive \( x \rightarrow uzv \). As \( uzv \) is either of the form \( bw \) or of the form \( wb \), this is a parasitic word, a contradiction. \( \square \)

**Corollary.** (i) The inclusion in Lemma 14 is strict. (ii) \( \text{REG} \) is incomparable with \( \text{CL}_{\text{sc}}(\text{FIN}) \).

**Proof.** \( ab^+a \in \text{REG} \subseteq \text{CL}_{\text{in}}(\text{REG}) \) and \( L = \{ a^nb^n \mid n \geq 1 \} \in \text{CL}_{\text{sc}}(\text{FIN}) \): we have \( L = L_{\text{sc}}(G) \) for \( G = (\{a, b\}, \{ab\}, \{(ab), \{(a, b)\}) \). \( \square \)

**Lemma 17.** \( \text{CL}_{\text{sc}}(\text{FIN}) - \text{CL}_{\text{in}}(\text{FIN}) \neq \emptyset \).

**Proof.** Let us consider the grammar

\[
G = (\{a, b, c, d\}, \{\lambda\}, (\{\lambda\}, \{(a, b), (c, d)\})).
\]

The language \( L_{\text{sc}}(G) \) is not in \( \text{CL}_{\text{in}}(\text{FIN}) \). Assume the contrary, and take \( G' = (\{a, b, c, d\}, B, (D_1, C_1), \ldots, (D_n, C_n)) \) such that \( L_{\text{in}}(G') = L_{\text{sc}}(G) \) and \( D_1, \ldots, D_n \) are finite. Consider the words of the form \( a^pb^q d^r d^s \), \( p, q \geq 1 \). All of them are in \( L_{\text{sc}}(G) \). If \( w \rightarrow a^p c^q b^p d^q \) in the grammar \( G' \), for \( w \neq a^p c^q b^p d^q \), then we must have \( w = a^p c^q b^p d^q \) for some \( p' \leq p, q' \leq q \),
and at least one of these inequalities is strict. Then the used selector must be of length
at least \( \min(p', q') \). Because the sets \( C_1, \ldots, C_n \) are finite, there is a constant \( t \) such that
\( p - p' < t, q - q' < t \). Consequently, because the sets \( D_1, \ldots, D_n \) are finite, derivations as
above are possible only for bounded \( p', q' \), hence for bounded \( p, q \). The language \( L_{sc}(G) \)
contains words \( a^p c^q b^r d^s \) with arbitrarily large \( p, q, a, b, c, d \), a contradiction.

**Corollary.** \( CL_{sc}(FIN) \) is incomparable with \( CL_{in}(FIN) \) and \( CF \).

**Proof.** We know already that \( CL_{in}(FIN) - CL_{sc}(FIN) \neq \emptyset \) (Lemma 16). Moreover,
\( L_{sc}(G) \) above is not context-free: \( L_{sc}(G) \cap a^p c^q b^r d^s = \{a^n c^m b^n d^m \mid n, m \geq 1\} \), which is
not context-free.

Summarizing the previous lemmas, we obtain

**Theorem 3.** (i) \( CL_{sc}(REG) = CL_{sc}(FIN) \subseteq CL_{in}(REG) \). (ii) The family \( CL_{sc}(FIN) \)
is incomparable with \( CL_{in}(FIN) \), \( REG \) and \( CF \).

Up-to-dating note. The grammars with minimal and maximal use of selectors were
investigated in a series of subsequent papers, [6], [7], [8], etc, where most of the problems
left open here are solved. A comprehensive presentation can be found in the forthcoming
monograph [11]. For instance, one knows in this moment that:

1. \( REG \subseteq CL_{mg}(FIN) \cap CL_{mg}(FIN) \).
2. \( CL_{mg}(FIN) \) contains non-semilinear languages.
3. Most of families in Diagrams 1 and 2 which are not linked by a path in these diagrams
are incomparable.
4. All the three basic non-context-free constructions in natural and artificial languages
(reduplication, multiple agreements and crossed dependencies), lead to languages
\( \{xzx \mid x \in \{a, b\}^*\}, \{a^n b^n c^n \mid n \geq 1\}, \{a^n b^n c^n d^n \mid n, m \geq 1\} \), respectively which are in
\( CL_{Ml}(REG) \cap CL_{Mg}(REG) \). Moreover, the corresponding grammars are
very simple. This considerably increases the interest for these types of grammars as
models of natural language syntax.

**References**

Mathematical Aspects of Natural and Formal Languages (Gh. Păun, ed.), World Sci.


languages by internal contextual languages, Theoretical Computer Science, to appear.

3 (1952), 326 – 336.


