I F I G

Research

Report

Institut für Informatik

A Short Comment on Controlled Context-Free Grammar Derivations

Markus Holzer

IFIG Research Report 2402
July 2024

Institut für Informatik
JLU Gießen
Arndtstraße 2
35392 Giessen, Germany
Tel: +49-641-99-32141
Fax: +49-641-99-32149
mail@informatik.uni-giessen.de
www.informatik.uni-giessen.de

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

IFIG Research Report
IFIG Research Report 2402, July 2024

A Short Comment on Controlled Context-Free Grammar Derivations

Markus Holzer[1]

Institut für Informatik, Universität Giessen
Arndtstr. 2, 35392 Giessen, Germany

Abstract. We prove that the family of languages generated by regularly controlled grammars with control languages accepted by ordered automata is equal to the family of languages generated by matrix grammars. To our knowledge, this equivalence has been overlooked in the literature.

MSC Classification:
68Q42: Grammars and rewriting systems
68Q45: Formal languages and automata

Additional Key Words and Phrases: Context-Free Grammars, Controlled Derivation, Ordered Languages, Matrix Grammars and Languages.

[1] Email: holzer@informatik.uni-giessen.de

We assume the reader to be familiar with formal language theory as contained in [5]. In particular, a *regularly controlled grammar* is a quintuple $G = (N, T, S, P, R)$, where N, T, P, and S are specified as in a context-free grammar, and R is a regular set over P, the so called *control language*. The *language* $L(G)$ *generated by* G consists of all those words $w \in T^*$ such that there is a context-free derivation

$$S \Longrightarrow_{p_1} \alpha_1 \Longrightarrow_{p_2} \alpha_2 \Longrightarrow_{p_3} \cdots \Longrightarrow_{p_n} \alpha_n = w$$

with $p_1p_2p_3 \ldots p_n \in R$. Let X be a family of languages. Then we set

$$\mathcal{L}(X) = \{\, L(G) \mid G \text{ is a regularly controlled grammar with control language } R \in X \,\}.$$

Obviously, $\mathcal{L}(\{\, P^+ \mid P \text{ is an alphabet} \,\})$ is equal to CFL, the well-known family of context-free languages, and additionally, $\mathcal{L}(\mathsf{REG})$ is equal to the family of languages generated by matrix grammars with context-free rules [5].

The language families $\mathcal{L}(X)$ induced by sub-regularly controlled grammars, i. e., the control language is from a sub-regular language family X, fall into three classes depending on X—see, e. g., [1–3]: either $\mathcal{L}(X)$ is equal to

1. the family of finite languages FIN,
2. the family of context-free languages CFL, or
3. the family of matrix grammar languages, which is identical to the language family $\mathcal{L}(\mathsf{REG})$.

For instance, $\mathsf{FIN} = \mathcal{L}(\mathsf{FIN})$. Obviously for language families X and Y with $X \subseteq Y$, we have $\mathcal{L}(X) \subseteq \mathcal{L}(Y)$. For example, for the family of definite languages[1] DEF and the family of star-free languages[2] SF, we have $\mathsf{DEF} \subseteq \mathsf{SF}$, and thus $\mathcal{L}(\mathsf{DEF}) \subseteq \mathcal{L}(\mathsf{SF})$. In fact, in [1] it was shown that $\mathsf{CFL} = \mathcal{L}(\mathsf{DEF})$, while $\mathcal{L}(\mathsf{SF}) = \mathcal{L}(\mathsf{REG})$.

For the family of ordered languages ORD we know $\mathsf{DEF} \subseteq \mathsf{ORD} \subseteq \mathsf{SF}$. A language $L \subseteq \Sigma^*$ is *ordered* if and only if the language is accepted by some deterministic finite automaton $A = (Q, \Sigma, \delta, q_0, F)$ with an input alphabet Σ, a finite set Q of states, a start state $q_0 \in Q$, a set $F \subseteq Q$ of accepting states and a transition mapping δ where $(Q, \preceq)$ is a totally ordered set and, for any input symbol $a \in \Sigma$, the relation $q \preceq q'$ implies $\delta(q, a) \preceq \delta(q', a)$. To the best of our knowledge, the category into which $\mathcal{L}(\mathsf{ORD})$ falls is not known. We close this gap in the forthcoming. Before we can state our result we need some facts from [5] on matrix grammars.

A *matrix grammar* is a quadruple $G = (N, T, M, S)$, where N, T, and S are specified as in a context-free grammar, and P is a finite set of sequences of context-free rules, i. e.,

$$M = \{m_1, m_2, \ldots m_n\}$$

and

$$m_i = (A_{i,1} \to \alpha_{i,1}, A_{i,2} \to \alpha_{i,2}, \ldots, A_{i,k_i} \to \alpha_{i,k_i}),$$

where $A_{i,j} \in N$ and $\alpha_{i,j} \in (N \cup T)^+$, for $1 \leq i \leq r$ and $1 \leq j \leq k_i$. We associate labels with the rules in the matrices, and if $A_{i,j} \to \alpha_{i,j}$ in m_i is labeled by $p_{i,j}$, the matrix m_i is described by $m_i = p_{i,1}p_{i,2} \ldots p_{i,k_i}$. The language $L(G)$ generated by the matrix grammar G can be described by control sets as follows: let $G' = (N, T, P, S, R)$ be a regularly controlled grammar, where P is the set of all rules occurring in the matrices of M. Then $L(G) = L(G')$ with $R = M^+$. In fact, one can relax the control set to M^* instead of M^+.

[1] A language $L \subseteq \Sigma^*$ is a *definite language* if and only if if it can be represented in the form $L = A \cup \Sigma^* B$ where A and B are finite subsets of Σ^*.

[2] A regular language $L \subseteq \Sigma^*$ is *star-free* or *non-counting* if and only if there is a natural number $k \geq 1$ such that, for any three words $x \in \Sigma^*$, $y \in \Sigma^*$, and $z \in \Sigma^*$, it holds $xy^kz \in L$ if and only if $xy^{k+1}z \in L$.

For matrix grammars the following normal form exists: N is the union of two disjoint sets N_1 and N_2 and M contains only matrices of length 2, i. e., $m = pq$, for any matrix $m \in M$, and the productions labeled by p rewrite a nonterminal of N_1, and the productions labeled by q replace a nonterminal of N_2. Now we are ready to prove the following theorem.

Theorem 1. $\mathcal{L}(\mathsf{ORD}) = \mathcal{L}(\mathsf{REG})$.

Proof. Since $\mathsf{ORD} \subseteq \mathsf{SF}$ and thus $\mathcal{L}(\mathsf{ORD}) \subseteq \mathcal{L}(\mathsf{REG})$ it suffices to prove that $\mathcal{L}(\mathsf{REG}) \subseteq \mathcal{L}(\mathsf{ORD})$. Since $\mathcal{L}(\mathsf{REG})$ is equal to the family of languages accepted by matrix grammars by combining the above facts we argue that for any regularly controlled grammar G with control set R there is a regularly controlled grammar G' with control set M^*, where $M = \{m_1, m_2, \ldots, m_n\}$ with $m_i = p_i q_i$ with $p_i \neq q_j$, for $1 \leq i, j \leq n$ such that $L(G) = L(G')$. We prove that M^* is an ordered language.

Consider the following deterministic finite automaton $A = (Q, P, \delta, 0, \{0\})$ with

$$Q = \{\bot\} \cup \{0, 1, \ldots, n\} \cup \{\top\}$$

and

$$P = \{\, p_i \mid 1 \leq i \leq n \,\} \cup \{\, q_i \mid 1 \leq i \leq n \,\},$$

and the transition relation is given by

$$\delta(\bot, p_j) = \delta(\bot, q_j) = \bot, \quad \text{for } 1 \leq j \leq n,$$

$$\delta(i, p_j) = \begin{cases} j, & \text{if } i = 0 \text{ and } 1 \leq j \leq n, \\ \top & \text{otherwise,} \end{cases}$$

$$\delta(i, q_j) = \begin{cases} \bot, & \text{if } i < j \text{ and } 1 \leq j \leq n, \\ 0, & \text{if } i = j \text{ and } 1 \leq j \leq n, \\ \top, & \text{if } i > j \text{ and } 1 \leq j \leq n, \end{cases}$$

$$\delta(\top, p_j) = \delta(\top, q_j) = \top, \quad \text{for } 1 \leq j \leq n,$$

where $0 \leq i \leq n$. A drawing of the automaton is shown in Figure 1. By construction it is easy to see that $L(A) = M^*$ as desired. It remains to show that there is a total order that respects the transition relation, i. e., for any input symbol $a \in P$, the relation $s \preceq s'$ implies $\delta(s, a) \preceq \delta(s', a)$. To this end we argue as follows: let

$$\bot \preceq 0 \preceq 1 \preceq \cdots \preceq n \preceq \top.$$

Then we consider the following cases—consider an arbitrary pair of states (s, s') with $s \preceq s'$:

- If $s = \bot$, then by construction $\delta(s, a) = s$, for every $a \in P$, and hence $s = \delta(s, a) \preceq \delta(s', a)$, for very $a \in P$. A similar argumentation applies for $s' = \top$.
- Neither $s = \bot$ nor $s' = \top$. Thus, let $s = i$ and $s' = j$ with $i \neq j$. By assumption $s \preceq s'$, hence $i \leq j$ together with $i \neq j$ gives us $i < j$. Then we distinguish the following sub-cases:
  - Assume $i = 0$. Then for $p_k \in P$, for $1 \leq k \leq n$, we have $\delta(s, p_k) = \delta(0, p_k) = k$, while $\delta(s', p_k) = \delta(j, p_k) = \top$. Moreover, for letter $q_k \in P$, for $1 \leq k \leq n$, by construction we have $\delta(s, q_k) = \delta(0, q_k) = \bot$. Thus for every $p \in P$ we have $\delta(s, p) \preceq \delta(s', p)$.
  - Finally, let $1 \leq i \leq n - 1$. Then $\delta(s, p_k) = \delta(s', p_k) = \top$, for every $1 \leq k \leq n$. Next, for $1 \leq k < i$ we have $\delta(s, q_k) = \delta(i, q_k) = \top = \delta(j, q_k) = \delta(s', q_k)$. For $k = i$, the construction of the automaton gives us $\delta(s, q_k) = \delta(i, q_i) = 0 \preceq \top = \delta(j, q_i) = \delta(s', q_k)$. Finally, in case $i < k$ we have $\delta(s, q_k) = \bot \preceq \delta(s', q_k)$. Thus, $\delta(s, p) \preceq \delta(s', p)$, for every $p \in P$.

This proves the stated claim. □

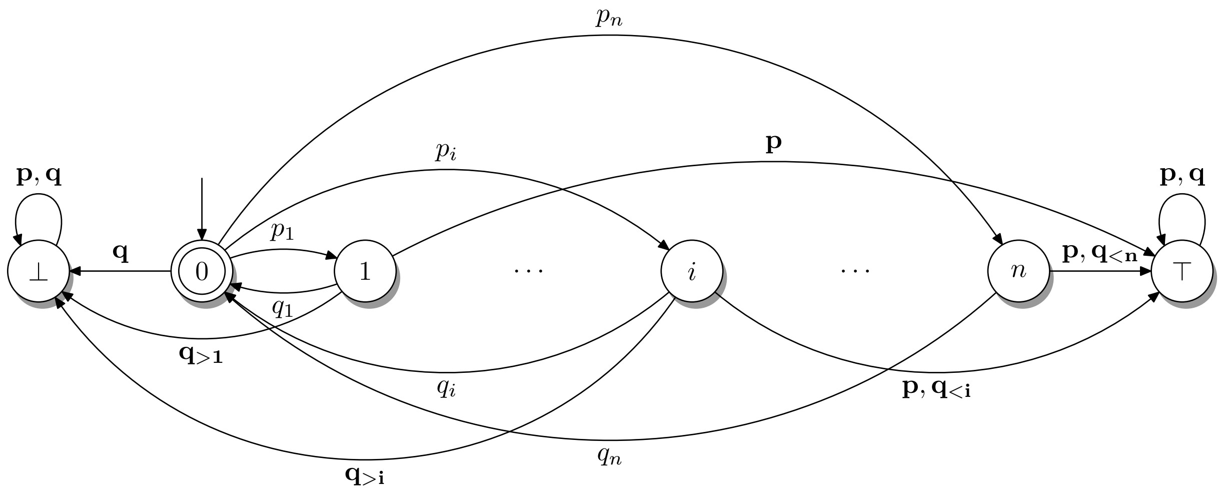

Fig. 1. Ordered automaton A. Here $\mathbf{p}$ ($\mathbf{q}$, respectively) abbreviates p_i (q_i, respectively), for $1 \leq i \leq n$. Moreover $\mathbf{q_{<i}}$ ($\mathbf{q_{>i}}$, respectively) is a short hand notation for q_j, for $1 \leq j < i$ (for $i < j \leq n$, respectively).

Thus, we have shown that the family of languages generated by regularly controlled grammars with control languages accepted by ordered automata is equal to the family of languages generated by matrix grammars. It is worth mentioning that, recently, it has been shown independently in [4] that $\mathcal{L}(\mathsf{ORD}) = \mathcal{L}(\mathsf{REG})$.

Acknowledgement. Thanks to Christian Rauch for his comments on an early draft of this paper.

References

1. J. Dassow. Subregularly controlled derivations: the context-free case. *Rostocker Mathematisches Kolloquium*, 34:61–70, 1988.
2. J. Dassow. Subregularly controlled derivations: Restrictions by syntactic parameters. In C. Martín-Vide and V. Mitrana, editors, *Where Mathematics, Computer Science, Linguistics and Biology Meet: Essays in honour of Gheorghe Păun*, pages 51–61. Springer, 2001.
3. J. Dassow. Grammars with control by ideals and codes. *Journal of Automata, Languages and Combinatorics*, 23(1-3):143–163, 2018.
4. J. Dassow. Remarks on context-free grammars with subregular control languages. *Theoretical Computer Science*, 1010:114704, September 2024.
5. J. Dassow and Gh. Păun. *Regulated Rewriting in Formal Language Theory*, volume 18 of *EATCS Monographs in Theoretical Computer Science*. Springer, 1989.

Institut für Informatik

Justus-Liebig-Universität Giessen
Arndtstr. 2, 35392 Giessen, Germany

Recent Reports

(Further reports are available at `www.informatik.uni-giessen.de`.)

H. Gruber, M. Holzer, C. Rauch *On Pumping Preserving Homomorphisms and the Complexity of the Pumping Problem*, Report 2401, June 2024.

M. Holzer, *Comments on Monoids Induced by NFAs*, Report 2301, March 2023.

S. Beier, M. Holzer, *Semi-Linear Lattices and Right One-Way Jumping Finite Automata*, Report 1901, April 2019.

M. Kutrib, T. Worsch, *Self-Verifying Cellular Automata*, Report 1803, April 2018.

S. Beier, M. Holzer, *Properties of Right One-Way Jumping Finite Automata*, Report 1802, March 2018.

B. Truthe, *Hierarchy of Subregular Language Families*, Report 1801, February 2018.

M. Holzer, M. Hospodár, *On the Magic Number Problem of the Cut Operation*, Report 1703, October 2017.

M. Holzer, S. Jakobi, *A Note on the Computational Complexity of Some Problems for Self-Verifying Finite Automata*, Report 1702, April 2017.

S. Beier, M. Holzer, M. Kutrib, *On the Descriptional Complexity of Operations on Semilinear Sets*, Report 1701, April 2017.

M. Holzer, S. Jakobi, M. Wendlandt, *On the Computational Complexity of Partial Word Automata Problems*, Report 1404, May 2014.

H. Gruber, M. Holzer, *Regular Expressions From Deterministic Finite Automata, Revisited*, Report 1403, May 2014.

M. Kutrib, A. Malcher, M. Wendlandt, *Deterministic Set Automata*, Report 1402, April 2014.

M. Holzer, S. Jakobi, *Minimal and Hyper-Minimal Biautomata*, Report 1401, March 2014.

J. Kari, M. Kutrib, A. Malcher (Eds.), *19th International Workshop on Cellular Automata and Discrete Complex Systems AUTOMATA 2013 Exploratory Papers*, Report 1302, September 2013.

M. Holzer, S. Jakobi, *Minimization, Characterizations, and Nondeterminism for Biautomata*, Report 1301, April 2013.

A. Malcher, K. Meckel, C. Mereghetti, B. Palano, *Descriptional Complexity of Pushdown Store Languages*, Report 1203, May 2012.

M. Holzer, S. Jakobi, *On the Complexity of Rolling Block and Alice Mazes*, Report 1202, March 2012.

M. Holzer, S. Jakobi, *Grid Graphs with Diagonal Edges and the Complexity of Xmas Mazes*, Report 1201, January 2012.

H. Gruber, S. Gulan, *Simplifying Regular Expressions: A Quantitative Perspective*, Report 0904, August 2009.

M. Kutrib, A. Malcher, *Cellular Automata with Sparse Communication*, Report 0903, May 2009.

M. Holzer, A. Maletti, *An $n \log n$ Algorithm for Hyper-Minimizing States in a (Minimized) Deterministic Automaton*, Report 0902, April 2009.

H. Gruber, M. Holzer, *Tight Bounds on the Descriptional Complexity of Regular Expressions*, Report 0901, February 2009.

M. Holzer, M. Kutrib, and A. Malcher (Eds.), *18. Theorietag Automaten und Formale Sprachen*, Report 0801, September 2008.

M. Holzer, M. Kutrib, *Flip-Pushdown Automata: Nondeterminism is Better than Determinism*, Report 0301, February 2003

M. Holzer, M. Kutrib, *Flip-Pushdown Automata: $k + 1$ Pushdown Reversals are Better Than k*, Report 0206, November 2002

M. Holzer, M. Kutrib, *Nondeterministic Descriptional Complexity of Regular Languages*, Report 0205, September 2002

H. Bordihn, M. Holzer, M. Kutrib, *Economy of Description for Basic Constructions on Rational Transductions*, Report 0204, July 2002

M. Kutrib, J.-T. Löwe, *String Transformation for n-dimensional Image Compression*, Report 0203, May 2002

A. Klein, M. Kutrib, *Grammars with Scattered Nonterminals*, Report 0202, February 2002

A. Klein, M. Kutrib, *Self-Assembling Finite Automata*, Report 0201, January 2002