THEORIE-TAG 2023

A. Lin, G. Zetzsche (Eds.): Theorietag 2023, Kaiserslautern, 4. – 6.10.2023
RPTU & MPI-SWS Kaiserslautern, Technical Report, pp. 61–66.

Strictly Locally Testable and Resources Restricted Control Languages in Tree-Controlled Grammars

Bianca Truthe

Institut für Informatik, Universität Giessen
Arndtstr. 2, 35392 Giessen, Germany
bianca.truthe@informatik.uni-giessen.de

Abstract

Tree-controlled grammars are context-free grammars where the derivation process is controlled in such a way that every word on a level of the derivation tree must belong to a certain control language. We investigate the generative capacity of such tree-controlled grammars where the control languages are special regular sets, especially strictly locally testable languages or languages restricted by resources of the generation (number of non-terminal symbols or production rules) or acceptance (number of states). Furthermore, the set theoretic inclusion relations of these subregular language families themselves are studied.

1. Introduction

In the monograph [1] by Jürgen Dassow and Gheorghe Păun, *Seven Circumstances Where Context-Free Grammars Are Not Enough* are presented. A possibility to enlarge the generative power of context-free grammars is to introduce some regulation mechanism which controls the derivation in a context-free grammar. In some cases, regular languages are used for such a regulation. They are rather easy to handle and, used as control, they often lead to context-sensitive or even recursively enumerable languages while the core grammar is only context-free.

One such control mechanism was introduced by Karel Čulik II and Hermann A. Maurer in [13] where the structure of derivation trees of context-free grammars is restricted by the requirement that the words of all levels of the derivation tree must belong to a given regular (control) language. This model is called tree-controlled grammar.

Gheorghe Păun proved that the generative capacity of such grammars coincides with that of context-sensitive grammars (if no erasing rules are used) or arbitrary phrase structure grammars (if erasing rules are used). Thus, the question arose to what extend the restrictions can be weakened in order to obtain 'useful' families of languages which are located somewhere between the classes of context-free and context-sensitive languages.

This is an extended abstract of a paper presented at AFL 2023 ([10]).

In [2, 3, 4, 5, 9, 11, 12], many subregular families of languages have been investigated as classes for the control languages. In this paper, we continue this research with further subregular language families, especially strictly locally testable languages or languages restricted by resources of the generation (number of non-terminal symbols or production rules) or acceptance (number of states). Furthermore, the set theoretic inclusion relations of these subregular language families themselves are studied.

2. Preliminaries

By *REG*, *CF* and *CS*, we denote the set of all regular, all context-free, and all context-sensitive languages, respectively. With a derivation of a terminal word by a context-free grammar, we associate a derivation tree which has the start symbol in its root and where every node with a non-terminal $A \in N$ has as children nodes with symbols which form, read from left to right, a word w such that $A \to w$ is a rule of the grammar (if $A \to \lambda$, then the node with A has only one child node and this is labelled with λ). Nodes with terminal symbols or λ have no children. With any derivation tree t of height k and any number $0 \leq j \leq k$, we associate the word of level j and the sentential form of level j which are given by all nodes of depth j read from left to right and all nodes of depth j and all leaves of depth less than j read from left to right, respectively. Obviously, if two words w and v are sentential forms of two successive levels, then $w \Longrightarrow^* v$ holds and this derivation is obtained by a parallel replacement of all non-terminal symbols occurring in the word w.

By *MAT*, we denote the family of all languages generated by matrix grammars with appearance checking and without erasing rules; by $\mathit{MAT}_{\mathit{fin}}$, we denote the family of all such languages where the matrix grammar is of finite index ([1], [8]). By *E0L* (*ET0L*), we denote the family of all languages generated by extended (tabled) interactionless Lindenmayer systems ([7]).

By restricting the resources needed for generating or accepting their elements, we define the following families:

$$\begin{aligned}
\mathit{RL}_n^V &= \{\, L \mid L \text{ is gen. by a right-lin. grammar with at most } n \text{ non-terminal symbols} \,\}\,, \\
\mathit{RL}_n^P &= \{\, L \mid L \text{ is gen. by a right-lin. grammar with at most } n \text{ production rules} \,\}\,, \\
\mathit{REG}_n^Z &= \{\, L \mid L \text{ is acc. by a DFA with at most } n \text{ states} \,\}\,.
\end{aligned}$$

Furthermore, we consider the following restrictions for regular languages. Let L be a language over an alphabet V. With respect to the alphabet V, the language L is said to be

- *monoidal* if and only if $L = V^*$,
- *nilpotent* if and only if it is finite or its complement $V^* \setminus L$ is finite,
- *combinational* if and only if it has the form $L = V^*X$ for some subset $X \subseteq V$,
- *definite* if and only if it can be represented in the form $L = A \cup V^*B$ where A and B are finite subsets of V^*,
- *suffix-closed* (or *fully initial* or *multiple-entry* language) if and only if, for any two words $x \in V^*$ and $y \in V^*$, the relation $xy \in L$ implies the relation $y \in L$,
- *ordered* if and only if the language is accepted by some deterministic finite automaton $A = (V, Z, z_0, F, \delta)$ with an input alphabet V, a finite set Z of states, a start state $z_0 \in Z$, a set $F \subseteq Z$ of accepting states and a transition mapping δ where $(Z, \preceq)$

is a totally ordered set and, for any input symbol $a \in V$, the relation $z \preceq z'$ implies $\delta(z,a) \preceq \delta(z',a)$,
- *commutative* if and only if it contains with each word also all permutations of this word,
- *circular* if and only if it contains with each word also all circular shifts of this word,
- *non-counting* (or *star-free*) if and only if there is a natural number $k \geq 1$ such that, for every three words $x \in V^*$, $y \in V^*$, and $z \in V^*$, it holds $xy^kz \in L$ if and only if $xy^{k+1}z \in L$,
- *power-separating* if and only if, there is a natural number $m \geq 1$ such that for every word $x \in V^*$, either $J_x^m \cap L = \emptyset$ or $J_x^m \subseteq L$ where $J_x^m = \{ x^n \mid n \geq m \}$,
- *union-free* if and only if L can be described by a regular expression which is only built by product and star,
- *strictly locally k-testable* if and only if there are three subsets B, I, and E of V^k such that any word $a_1a_2\dots a_n$ with $n \geq k$ and $a_i \in V$ for $1 \leq i \leq n$ belongs to the language L if and only if $a_1a_2\dots a_k \in B$, $a_{j+1}a_{j+2}\dots a_{j+k} \in I$ for every j with $1 \leq j \leq n-k-1$, and $a_{n-k+1}a_{n-k+2}\dots a_n \in E$,
- *strictly locally testable* if and only if it is strictly locally k-testable for some natural number k.

We remark that monoidal, nilpotent, combinational, definite, ordered, union-free, and strictly locally (k-)testable languages are regular, whereas non-regular languages of the other types mentioned above exist. Here, we consider among the commutative, circular, suffix-closed, non-counting, and power-separating languages only those which are also regular.

By *FIN*, *MON*, *NIL*, *COMB*, *DEF*, *SUF*, *ORD*, *COMM*, *CIRC*, *NC*, *PS*, *UF*, SLT_k (for any natural number $k \geq 1$), and *SLT*, we denote the families of all finite, monoidal, nilpotent, combinational, definite, regular suffix-closed, ordered, regular commutative, regular circular, regular non-counting, regular power-separating, union-free, strictly locally k-testable, and strictly locally testable languages, respectively.

For any natural number $n \geq 1$, let MON_n be the set of all languages that can be represented in the form $A_1^* \cup A_2^* \cup \cdots \cup A_k^*$ with $1 \leq k \leq n$ where all A_i ($1 \leq i \leq k$) are alphabets. Obviously,

$$\mathit{MON} = \mathit{MON}_1 \subset \mathit{MON}_2 \subset \cdots \subset \mathit{MON}_j \subset \cdots .$$

A strictly locally testable language characterized by three finite sets B, I, and E as above which includes additionally a finite set F of words which are shorter than those of the sets B, I, and E is denoted by $[B,I,E,F]$.

As the set of all families under consideration, we set

$$\begin{aligned}\mathfrak{F} = \{&\mathit{FIN},\mathit{NIL},\mathit{COMB},\mathit{DEF},\mathit{SUF},\mathit{ORD},\mathit{COMM},\mathit{CIRC},\mathit{NC},\mathit{PS},\mathit{UF}\}\\ &\cup \{ \mathit{MON}_k \mid k \geq 1 \} \cup \{\mathit{SLT}\} \cup \{ \mathit{SLT}_k \mid k \geq 1 \}\\ &\cup \{ \mathit{RL}_n^V \mid n \geq 1 \} \cup \{ \mathit{RL}_n^P \mid n \geq 1 \} \cup \{ \mathit{REG}_n^Z \mid n \geq 1 \}.\end{aligned}$$

A tree-controlled grammar is a quintuple $G = (N,T,P,S,R)$ where
- (N,T,P,S) is a context-free grammar with a set N of non-terminal symbols, a set T of terminal symbols, a set P of context-free non-erasing rules (with the only exception that the rule $S \to \lambda$ is allowed if S does not occur on a right-hand side of a rule), and an axiom S,
- R is a regular set over $N \cup T$.

The language $L(G)$ generated by a tree-controlled grammar $G = (N,T,P,S,R)$ consists of all such words $z \in T^*$ which have a derivation tree t where z is the word obtained by reading the leaves from left to right and the words of all levels of t – besides the last one – belong to the regular control language R.

Let $\mathcal{F}$ be a subfamily of *REG*. Then, we denote the family of languages generated by tree-controlled grammars $G = (N,T,P,S,R)$ with $R \in \mathcal{F}$ by $\mathcal{TC}(\mathcal{F})$.

In [6] (see also [1]), it has been shown that a language L is generated by a tree-controlled grammar if and only if it is generated by a context-sensitive grammar. In subsequent papers, tree-controlled grammars have been investigated where the control language belongs to some subfamily of the class *REG* ([2, 3, 4, 5, 9, 11, 12]). In this paper, we continue this research with further subregular language families. From the definition follows that the subset relation is preserved under the use of tree-controlled grammars: if we allow more, we do not obtain less.

3. Results

A summary of all the inclusion relations obtained so far is given in Figure 1. An arrow from an entry X to an entry Y depicts the inclusion $X \subseteq Y$; a solid arrow means proper inclusion; a dashed arrow indicates that it is not known whether the inclusion is proper. If two families are not connected by a directed path, then they are not necessarily incomparable. An edge label in this figure refers to the paper where the respective inclusion is proved.

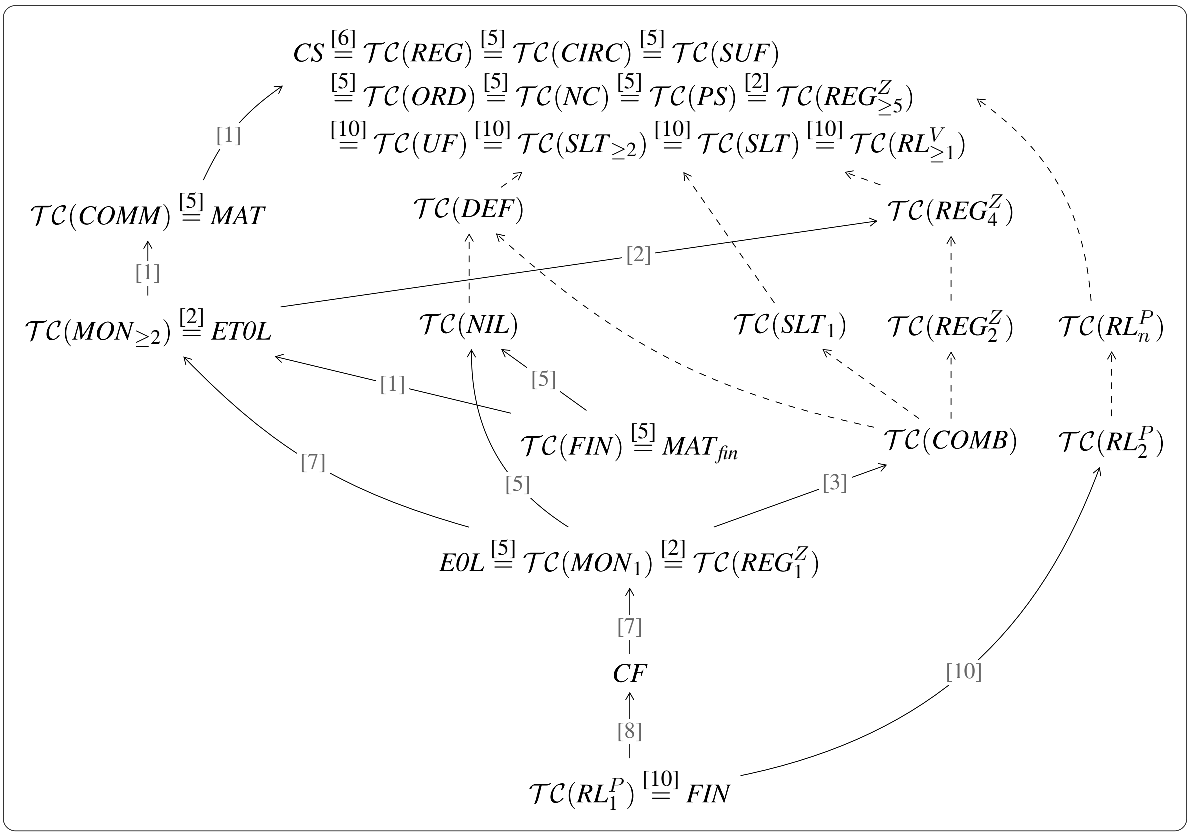

Figure 1: Hierarchy of subregularly tree-controlled language families

4. Future Research

There are several families of languages generated by tree-controlled grammars where we do not have a characterization by some other language class. The strictness of some inclusions and the incomparability of some families remain as open problems.

In the present paper, we have only considered tree-controlled grammars without erasing rules. For tree-controlled grammars where erasing rules are allowed, several results have been published already (see, e. g., [3, 11, 12]). Also in this situation, there are some open problems.

Another direction for future research is to consider other subregular language families or to relate the families of languages generated by tree-controlled grammars to language families obtained by other grammars/systems with regulated rewriting.

References

[1] J. DASSOW, GH. PĂUN, *Regulated Rewriting in Formal Language Theory*. EATCS Monographs in Theoretical Computer Science 18, Springer-Verlag, 1989.

[2] J. DASSOW, R. STIEBE, B. TRUTHE, Two collapsing hierarchies of subregularly tree controlled languages. *Theoretical Computer Science* **410** (2009) 35, 3261–3271.

[3] J. DASSOW, R. STIEBE, B. TRUTHE, Generative capacity of subregularly tree controlled grammars. *International Journal of Foundations of Computer Science* **21** (2010) 5, 723–740.

[4] J. DASSOW, B. TRUTHE, On two hierarchies of subregularly tree controlled languages. In: C. CÂMPEANU, G. PIGHIZZINI (eds.), *Descriptional Complexity of Formal Systems, 10th International Workshop, Charlottetown, Prince Edward Island, Canada, July 16–18, 2008, Proceedings*. University of Prince Edward Island, 2008, 145–156.

[5] J. DASSOW, B. TRUTHE, Subregularly tree controlled grammars and languages. In: E. CSUHAJ-VARJÚ, Z. ÉSIK (eds.), *Automata and Formal Languages, 12th International Conference, AFL 2008, Balatonfüred, Hungary, May 27–30, 2008, Proceedings*. Computer and Automation Research Institute, Hungarian Academy of Sciences, 2008, 158–169.

[6] GH. PĂUN, On the generative capacity of tree controlled grammars. *Computing* **21** (1979), 213–220.

[7] G. ROZENBERG, A. SALOMAA, *The Mathematical Theory of L Systems*. Academic Press, 1980.

[8] G. ROZENBERG, A. SALOMAA (eds.), *Handbook of Formal Languages*. Springer-Verlag, Berlin, 1997.

[9] R. STIEBE, On the complexity of the control language in tree controlled grammars. In: J. DASSOW, B. TRUTHE (eds.), *Colloquium on the Occasion of the 50th Birthday of Victor Mitrana, Otto-von-Guericke-Universitäat Magdeburg, Germany, June 27, 2008, Proceedings*. Otto-von-Guericke-Universitäat Magdeburg, Germany, 2008, 29–36.

[10] B. TRUTHE, Strictly Locally Testable and Resources Restricted Control Languages in Tree-Controlled Grammars. In: ZS. GAZDAG, SZ. IVÁN, G. KOVÁSZNAI (eds.), *16th International Conference on Automata and Formal Languages, AFL 2023, Eger, Hungary, September 5–7, 2023, Proceedings*. EPTCS 386, Open Publishing Association, 2023, 253–268.

[11] S. TURAEV, J. DASSOW, F. MANEA, M. H. SELAMAT, Language classes generated by tree controlled grammars with bounded nonterminal complexity. *Theoretical Computer Science* **449** (2012), 134–144.

[12] GY. VASZIL, On the nonterminal complexity of tree controlled grammars. In: H. BORDIHN, M. KUTRIB, B. TRUTHE (eds.), *Languages Alive – Essays Dedicated to Jürgen Dassow on the Occasion of His 65th Birthday*. LNCS 7300, Springer, 2012, 265–272.

[13] K. ČULIK II, H. A. MAURER, Tree controlled grammars. *Computing* **19** (1977) 2, 129–139.