Alexander Okhotin | Christian Reitwießner

Conjunctive grammars with restricted disjunction

Turku Centre for Computer Science

TUCS Technical Report
No 915, March 2009

Conjunctive grammars with restricted disjunction

Alexander Okhotin
Academy of Finland, *and*
Department of Mathematics, University of Turku, *and*
Turku Centre for Computer Science,
Turku FIN–20014, Finland
alexander.okhotin@utu.fi
Christian Reitwießner
Julius-Maximilians-Universität Würzburg, Theoretische Informatik,
Würzburg D–97074, Germany
reitwiessner@informatik.uni-wuerzburg.de

TUCS Technical Report
No 915, March 2009

Abstract

It is shown that every conjunctive language is generated by a conjunctive grammar of a special form, in which every nonterminal A has at most one rule of the general form $A \to \alpha_1 \& \ldots \& \alpha_n$, while the rest of the rules for A must be of the type $A \to w$, where w is a terminal string. For context-free grammars, a similar property does not hold (S. A. Greibach, W. Shi, S. Simonson, "Single tree grammars", 1992).

Keywords: conjunctive grammars, single tree grammars, normal form, language equations

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

Context-free grammars are the most obvious mathematical model of syntax, which represents inductive definition of a set of strings. This is done using one Boolean operation: the *disjunction*, which is implicit in having multiple rules for one nonterminal symbol. These natural expressive means, together with efficient parsing algorithms, make context-free grammars the most practically used method of defining formal languages.

As *conjunction* of syntactical conditions is not expressible in context-free grammars, this model can be extended by allowing an explicit conjunction in the formalism of rules. The resulting extension is known as *conjunctive grammars* [8], it maintains the principle of defining a language inductively and still allows efficient parsing algorithms. At the same time, using conjunction in addition to disjunction considerably increases the expressive power of the model. Besides being able to represent many standard examples of non-context-free languages, such as $\{ a^n b^n c^n \mid n \geqslant 0 \}$ and $\{ wcw \mid w \in \{a, b\}^* \}$ [8], conjunctive grammars are notable for their non-trivial expressive power over a one-letter alphabet, studied by Jeż [4] and by Jeż and Okhotin [5, 6]. This work, in particular, led to unexpected strong results on equations over sets of numbers [7].

This paper continues the investigation of the power of Boolean operations in context-free grammars with a subclass of conjunctive grammars, in which the disjunction can be used only in the form of disjunction with a terminal string. In other words, each nonterminal A may have only one rule referring to other nonterminals, while the rest of its rules must be of the form $A \to w$, where w is a terminal string. The same restriction on the context-free grammars has been studied by Greibach et al. [3] under the name of *single tree grammars*. These grammars have quite a limited expressive power; in particular, they cannot generate the language of all palindromes. The latter language, as shown by Reitwießner [10] is not even in the union closure of single tree grammars. This means that unrestricted use of disjunction is essential for context-free grammars.

Similarly to single tree grammars, one can expect conjunctive grammars restricted to use disjunction only with terminal strings to be much weaker than conjunctive grammars of the general form. However, the results of this paper contrast this intuition, and it is shown that in fact every conjunctive grammar can be effectively transformed to an equivalent grammar with restricted disjunction. Unrestricted disjunction is thus redundant in conjunctive grammars. The form with restricted disjunction may thus be regarded as a normal form for conjunctive grammars.

The proof of this result is based upon another normal form for conjunctive grammars, the *odd normal form*, in which every nonterminal other than the start symbol generates only strings of odd length. In Section 3 it is shown how to transform every conjunctive grammar to this form. The main result of the paper, that every conjunctive language can be generated by a conjunctive grammar with restricted disjunction, is obtained in Section 4. Finally, the question of eliminating ε-rules in conjunctive grammars with restricted disjunction is addressed in Section 5: though it is not determined whether this is always possible, a construction of ε-free restricted conjunctive grammars for a subfamily of conjunctive languages including all regular languages is given.

2 Conjunctive grammars

Let us define the main operations on languages used in this paper. These are, first of all, Boolean operations: *union, intersection* and *complementation* $\overline{L} = \Sigma^* \setminus L$, as well as *concatenation*: $K \cdot L = KL = \{ uv \mid u \in K, v \in L \}$. The *quotient* of a language with a singleton is defined as follows: for all $L \subseteq \Sigma^*$ and $u \in \Sigma^*$, the languages $u^{-1}L := \{w \mid uw \in L\}$ and $Lu^{-1} := \{w \mid wu \in L\}$ are the *left* and *right* quotients of L with u, respectively. This operation is extended to languages as $K^{-1}L := \{ v \mid \exists u \in K : uv \in L \}$ and $LK^{-1} := \{ u \mid \exists v \in K : uv \in L \}$ for $K, L \subseteq \Sigma^*$.

Definition 1 (Okhotin [8]). *A conjunctive grammar is a quadruple* $G = (\Sigma, N, P, S)$*, in which* Σ *and* N *are disjoint finite nonempty sets of terminal and nonterminal symbols, respectively;* P *is a finite set of rules, each of the form*

$$A \to \alpha_1 \& \ldots \& \alpha_n \quad (\textit{with } A \in N,\ n \geqslant 1 \textit{ and } \alpha_1, \ldots, \alpha_n \in (\Sigma \cup N)^*) \qquad (1)$$

and $S \in N$ *is a nonterminal designated as the start symbol.*

Informally, a rule (1) states that if a string is generated by each α_i, then it is generated by A. This semantics can be formalized using term rewriting, which generalizes Chomsky's string rewriting.

Definition 2 ([8]). *Given a grammar* G*, consider terms over concatenation and conjunction with symbols from* $\Sigma \cup N$ *as atomic terms. The relation* $\Longrightarrow$ *of immediate derivability on the set of terms is defined as follows:*

- *Using a rule* $A \to \alpha_1 \& \ldots \& \alpha_n$*, a subterm* $A \in N$ *of any term* $\varphi(A)$ *can be rewritten as* $\varphi(A) \Longrightarrow \varphi(\alpha_1 \& \ldots \& \alpha_n)$*.*
- *A conjunction of several identical strings can be rewritten by one such string:* $\varphi(w \& \ldots \& w) \Longrightarrow \varphi(w)$*, for every* $w \in \Sigma^*$*.*

The language generated by a term φ *is* $L_G(\varphi) = \{ w \mid w \in \Sigma^*,\ \varphi \Longrightarrow^* w \}$*. The language generated by the grammar is* $L(G) = L_G(S) = \{w \mid w \in \Sigma^*,$ $S \Longrightarrow^* w\}$*.*

An equivalent definition can be given using language equations. This definition generalizes the well-known characterization of the context-free grammars by equations, due to Ginsburg and Rice [1].

Definition 3. *For every conjunctive grammar $G = (\Sigma, N, P, S)$, the associated system of language equations is a system of equations in variables N, in which each variable assumes the value of a language over Σ, and which contains the following equation for every variable A:*

$$A = \bigcup_{A \to \alpha_1 \& \ldots \& \alpha_m \in P} \bigcap_{i=1}^{m} \alpha_i \quad (\textit{for all } A \in N) \ . \tag{2}$$

Each occurrence of a symbol $a \in \Sigma$ in such a system defines a constant language $\{a\}$, while each empty string denotes a constant language $\{\varepsilon\}$. A solution of a system is a vector of languages $(\ldots, L_C, \ldots)_{C \in N}$, such that the substitution of L_C for C, for all $C \in N$, turns each equation (2) into an equality.

Every such system has at least one solution, and among them a *least solution* with respect to componentwise inclusion. This solution consists of exactly the languages generated by the nonterminals of the original conjunctive grammar: $(\ldots, L_G(C), \ldots)_{C \in N}$.

Let us give some examples of conjunctive grammars. Every language representable as an intersection of finitely many context-free languages, such as $\{ a^n b^n c^n \mid n \geqslant 0 \}$, can be straightforwardly specified using conjunction for the start symbol. It is more interesting to construct a grammar for a language not in the intersection closure of the context-free languages, such as the following.

Example 1 (Okhotin [8]). *The conjunctive grammar*

$$\begin{array}{rcl}
S & \to & C\&D \\
C & \to & aCa \mid aCb \mid bCa \mid bCb \mid c \\
D & \to & aA\&aD \mid bB\&bD \mid cE \\
A & \to & aAa \mid aAb \mid bAa \mid bAb \mid cEa \\
B & \to & aBa \mid aBb \mid bBa \mid bBb \mid cEb \\
E & \to & aE \mid bE \mid \varepsilon
\end{array}$$

generates the language $\{ wcw \mid w \in \{a,b\}^* \}$. In particular, $L(D) = \{ uczu \mid u, z \in \{a,b\}^* \}$.

The rules for D match a single symbol in the left part to the corresponding symbol in the right part using A or B, and the recursive reference to aD or bD makes the remaining symbols be compared in the same way. The intersection with the language $\{ ucv \mid u, v \in \{a,b\}^*, |u| = |v| \}$ generated by C completes the grammar.

Example 2 (Jeż [4])**.** *The following conjunctive grammar with the start symbol A_1 generates the language $\{ a^{4^n} \mid n \geqslant 0 \}$:*

$$\begin{array}{lcl} A_1 & \to & A_2A_2 \& A_1A_3 \mid a \\ A_2 & \to & A_6A_2 \& A_1A_1 \mid aa \\ A_3 & \to & A_6A_6 \& A_1A_2 \mid aaa \\ A_6 & \to & A_3A_3 \& A_1A_2 \end{array}$$

Each nonterminal A_i generates the language $\{ a^{i \cdot 4^n} \mid n \geqslant 0 \}$.

A generalization of the Chomsky normal form for conjunctive grammars is known.

Definition 4 (Binary normal form [8])**.** *A conjunctive grammar $G = (\Sigma, N, P, S)$ is in the binary normal form if every rule in P is of the form*

$$\begin{array}{l} A \to B_1C_1 \& \dots \& B_nC_n \quad (n \geqslant 1,\ B_i, C_i \in N) \\ A \to a \\ S \to \varepsilon \quad \text{(only if } S \text{ does not appear in right-hand sides of rules)} \end{array}$$

Every conjunctive grammar can be effectively transformed to a conjunctive grammar in the binary normal form generating the same language [8]. In particular, this normal form is used to obtain a simple generalization of the Cocke–Kasami–Younger parsing algorithm to conjunctive grammars, which still works in time $O(n^3)$ [8].

For context-free grammars, there is another important normal form: the *Greibach normal form* [2], in which every rule is either $A \to a\alpha$ with $\alpha \in (\Sigma \cup N)^*$, or $A \to \varepsilon$. This definition naturally carries on to conjunctive grammars. It can be said that a conjunctive grammar $G = (\Sigma, N, P, S)$ is in Greibach normal form if every rule in P is of the form

$$\begin{array}{l} A \to a\alpha_1 \& \dots \& a\alpha_n \quad (n \geqslant 1,\ \alpha_i \in N^*) \quad \text{or} \\ A \to \varepsilon. \end{array}$$

However, it is not known whether every conjunctive grammar can be transformed to this form.

Let us establish an entirely new normal form for conjunctive grammars, which will be crucial for the subsequent constructions.

3 The odd normal form

The *odd normal form* for conjunctive grammars proposed in this section has the following main property: every nonterminal (possibly except the start symbol) may only generate strings of odd length. As the parity of the length of strings is going to play an important role in all constructions below, let us introduce the notation Even $:= (\Sigma^2)^*$ and Odd $:= \Sigma(\Sigma^2)^*$ (where Σ is the implicitly assumed alphabet) for the sets of all strings of even and odd length, respectively.

Definition 5 (Odd normal form)**.** *A conjunctive grammar $G = (\Sigma, N, P, S)$ is said to be in* odd normal form *if all rules in P are of the form*

$$\begin{aligned} &A \to a && \text{with } A \in N,\ a \in \Sigma, \qquad \text{or} \\ &A \to B_1 a_1 C_1 \,\&\, \dots \,\&\, B_n a_n C_n && \text{with } n \geqslant 1,\ A, B_i, C_i \in N,\ a_i \in \Sigma \end{aligned}$$

If S does not occur in the right-hand sides of the rules, then the following two types of rules, called ***even rules****, are also allowed:*

$$\begin{aligned} &S \to aA && \text{with } a \in \Sigma,\ A \in N \\ &S \to \varepsilon \end{aligned}$$

Note that if there are no even rules in a grammar in odd normal form, then it generates a subset of Odd. Thus even rules are needed for some languages, but regardless of whether they are used, the main part of the grammar operates on odd strings only. The main step towards the transformation to the odd normal form is taking an arbitrary grammar in binary normal form and representing its operation on all strings using only odd strings.

Lemma 1. *For every conjunctive grammar $G = (\Sigma, N, P, S)$ in binary normal form there exists and can be effectively constructed a conjunctive grammar $G' := (\Sigma, N', P', S')$ in odd normal form without even rules, in which the set of nonterminals is $N' := (\Sigma \cup \{\varepsilon\}) \times N \times (\Sigma \cup \{\varepsilon\})$ and the language generated by each nonterminal (x, A, y), denoted ${}_xA_y$, is*

$$L_{G'}({}_xA_y) = x^{-1} L_G(A) y^{-1} \cap \text{Odd},$$

where $A \in N$ and $x, y \in \Sigma \cup \{\varepsilon\}$. The start symbol is $S' := {}_\varepsilon S_\varepsilon$, and hence $L(G') = L(G) \cap \text{Odd}$.

Proof. It can be assumed that G does not contain the rule $S \to \varepsilon$, since the languages $x^{-1} L_G(S) y^{-1} \cap \text{Odd}$ consist of strings of length at least one, and hence the membership of ε in $L(G)$ does not affect them.

The grammar G' is constructed as follows. For every rule

$$A \to B^{(1)}C^{(1)} \& \dots \& B^{(n)}C^{(n)} \in P,$$

each nonterminal ${}_xA_y$ with $x, y \in \Sigma \cup \{\varepsilon\}$ in the new grammar G' has all possible rules of the form

$${}_xA_y \to {}_x\alpha_y^{(1)} \& \dots \& {}_x\alpha_y^{(n)}$$

such that for every $i = 1, \ldots, n$,

$$
\begin{aligned}
{}_x\alpha_y^{(i)} \in \{{}_xB_a^{(i)} \cdot a \cdot {}_\varepsilon C_y^{(i)} \mid a \in \Sigma\} \cup & \qquad (3a) \\
\{{}_xB_\varepsilon^{(i)} \cdot a \cdot {}_a C_y^{(i)} \mid a \in \Sigma\} \cup & \qquad (3b) \\
\{{}_xB_\varepsilon^{(i)} \mid y \in L_G(C^{(i)})\} \cup & \qquad (3c) \\
\{{}_\varepsilon C_y^{(i)} \mid x \in L_G(B^{(i)})\}. & \qquad (3d)
\end{aligned}
$$

Additionally, for every ${}_xA_y \in N'$ and $a \in \Sigma$ with $xay \in L_G(A)$, the new grammar contains the rule

$$ {}_xA_y \to a. \tag{4} $$

It is easy to check that no nonterminal in G or G' generates the empty string and that all strings generated by nonterminals in N' have odd length.

Now it is claimed that for each ${}_xA_y \in N'$ and for every $w \in \Sigma^*$,

$$ w \in L_{G'}({}_xA_y) \qquad \text{if and only if} \qquad xwy \in L_G(A) \text{ and } w \in \mathrm{Odd}. $$

The proof in each direction is by induction on the length of w, and inside this induction there is another induction on $|xy|$.

$\Rightarrow$ Let $w \in L_{G'}({}_xA_y)$; it has to be proved that $xwy \in L_G(A)$. The proof will be done by induction on the length of w. More precisely, for each string, the statement is first proved for nonterminals with shorter indices. This means that the induction is actually on $3|w| + |xy|$).

Induction basis $|w| = 1$. If $w \in L_{G'}({}_xA_y)$ with $|w| = 1$, then w is either generated directly by a rule of type (4) (in which case the assertion obviously holds), or it can be generated via a "long" rule. Note that such a rule must consist entirely of unit conjuncts of the form (3c) and (3d), since all other conjuncts generate longer strings (as no nonterminal in G' generates the empty string). So let ${}_xA_y \to {}_x\alpha_y^{(1)} \& \ldots \& {}_x\alpha_y^{(n)}$ be this rule.

If $x = y = \varepsilon$, there cannot be conjuncts of type (3c) or (3d), since $\varepsilon \notin L_G(B^{(i)}), L_G(C^{(i)})$. So in this case, w can only be generated by a "short" rule, and there is nothing left to prove.

If $|x| = 1$ and $y = \varepsilon$, then there cannot be conjuncts of type (3c)), and there is a rule $A \to B^{(1)}C^{(1)} \& \ldots \& B^{(n)}C^{(n)} \in P$ such that for every $i = 1, \ldots, n$ it holds that $x \in L_G(B^{(i)})$ and ${}_x\alpha_y^{(i)} = {}_\varepsilon C_y^{(i)} = {}_\varepsilon C_\varepsilon^{(i)}$ As $w \in {}_\varepsilon C_y^{(i)}$ this means that $w \in L_G(C^{(i)})$, as we already proved. Then, of course, $xwy = xw \in L_G(B^{(i)}C^{(i)})$ and thus $xwy \in L_G(A)$.

The case for $x = \varepsilon$ and $|y| = 1$ is symmetric, so let $|x| = |y| = 1$. In this case, similarly there is a rule $A \to B^{(1)}C^{(1)} \& \ldots \& B^{(n)}C^{(n)} \in P$ in which, for every i-th conjunct, $x \in L_G(B^{(i)})$ and ${}_x\alpha_y^{(i)} = {}_\varepsilon C_y^{(i)}$, or $y \in L_G(C^{(i)})$ and ${}_x\alpha_y^{(i)} = {}_xB_\varepsilon^{(i)}$. Fix now i and, without loss of generality, assume the first of these two cases. We already proved that if $w \in L_{G'}({}_\varepsilon C_y^{(i)})$, then $wy \in L_G(C^{(i)})$, so $xwy \in L_G(B^{(i)}C^{(i)})$. As this holds for all i, we get $xwy \in L_G(A)$ and the induction basis is complete.

Induction step. Let $n \geqslant 1$ and assume that the assertion holds for all w with $|w| \leqslant n$. Let $w \in L_{G'}({}_xA_y)$ for some ${}_xA_y \in N'$ and $|w| = n+1$. Since $|w| > 1$, there must be a rule ${}_xA_y \to {}_x\alpha_y^{(1)} \mathbin{\&} \ldots \mathbin{\&} {}_x\alpha_y^{(n)}$ and $w \in L_{G'}({}_x\alpha_y^{(i)})$ for all $i = 1, \ldots, n$. Now fix i and consider the form of ${}_x\alpha_y^{(i)}$.

Assume it is of the form (3a), that is, ${}_x\alpha_y^{(i)} = {}_xB_a^{(i)} \cdot a \cdot {}_\varepsilon C_y^{(i)}$ for some $a \in \Sigma$. Then there are strings $u \in L_{G'}({}_xB_a^{(i)})$, $v \in L_{G'}({}_\varepsilon C_y^{(i)})$ such that $w = uav$. Since $1 \leqslant |u|, |v| \leqslant |w| - 2 = n - 1$, we have $xua \in L_G(B^{(i)})$ and $vy \in L_G(C^{(i)})$ by induction and thus $xwy \in L_G(B^{(i)}C^{(i)})$. The second case, ${}_x\alpha_y^{(i)} = {}_xB_\varepsilon^{(i)} \cdot a \cdot {}_aC_y^{(i)}$, works analogously.

Now there are the cases (3c) and (3d) left and we assume without loss of generality that $x \neq \varepsilon$ and ${}_x\alpha_y^{(i)} = {}_\varepsilon C_y^{(i)}$. This implies that $w \in L_{G'}({}_\varepsilon C_y^{(i)})$. Since ${}_\varepsilon C_y^{(i)}$ always has shorter indices than ${}_xA_y$, we get $wy \in L_G(C^{(i)})$ by induction. Since $x \in L_G(B^{(i)})$, we finally have $xwy \in L_G(B^{(i)}C^{(i)})$.

In all four cases, we got $xwy \in L_G(B^{(i)}C^{(i)})$. Since this holds for all $i \in \{1, \ldots, n\}$, we get $xwy \in L_G(A)$, which was asserted.

$\Leftarrow$ The other direction is now proved by induction on $|w|$ for $xwy \in L_G(A)$ (and again the statement is first proved for smaller $|xy|$ if the string length $|w|$ is the same).

Induction basis $|w| = 1$. The induction basis is clear by the rules (4).

Induction step. Assume that all four statements hold for $|w| \leqslant n$ and $n \geqslant 1$.

Let now $xwy \in L_G(A)$, $w \in \text{Odd}$ and $|w| = n + 2$. Since $|xwy| \geqslant |w| \geqslant 3$, there must be a rule $A \to B^{(1)}C^{(1)} \& \ldots \& B^{(n)}C^{(n)} \in P$ such that $w \in L_G(B^{(i)}C^{(i)})$ for all $i = 1, \ldots, n$. By the construction, there can be multiple rules in P' that correspond to this rule. We now argue that for every $i = 1, \ldots, n$, we can find a suitable conjunct ${}_x\alpha_y^{(i)}$ that generates w. For this, fix i again. Then there must be strings $u, v \in \Sigma^*$ such that $xu \in L_G(B^{(i)})$, $vy \in L_G(C^{(i)})$ and $xwy = xuvy$. Note that since w has odd length, either u or v has odd length. Without loss of generality, assume that $|v|$ is odd. Since $|v|$ is strictly smaller than $|w|$ (no nonterminal generates the empty string) we get $|v| \leq |w| - 2 = n$ and thus $v \in L_{G'}(C_y^{(i)})$ by induction. For u there are two cases.

- If $x \neq \varepsilon$, then it can be that $u = \varepsilon$ and thus $x = xu \in L_G(B^{(i)})$. In this case, ${}_{\varepsilon}C_y^{(i)}$ is a possible ith conjunct ${}_x\alpha_y^{(i)}$ (cf. rule (3d)). Since $v \in L_{G'}(C_y^{(i)})$, we get $w = uv = v \in L_{G'}(C_y^{(i)}) = L_{G'}({}_x\alpha_y^{(i)})$.
- If $u \neq \varepsilon$, then $u = u'a$ for some $a \in \Sigma$ and $xu'a \in L_G(B^{(i)})$ and thus $u' \in L_{G'}({}_xB_a^{(i)})$ by induction ($|u'| \leqslant |w| - 2 = n$ and it is odd). This means that $w = u'av \in L_{G'}({}_xB_a^{(i)} \cdot a \cdot C_y^{(i)})$, so this is a possible conjunct ${}_x\alpha_y^{(i)}$.

Now we showed that for every $i = 1, \ldots, n$, there is a legal conjunct ${}_x\alpha_y^{(i)}$ in the respective rule for ${}_xA_y$ in P' that generates w, which implies that $w \in L_{G'}({}_xA_y)$.

The grammar G' constructed above is not yet in the odd normal form, because it may contain so-called *unit conjuncts*, that is, rules of the form $A \to \ldots \& B \& \ldots$. The known procedure for eliminating such conjuncts [8] is a sequence of substitutions of the bodies of all rules for B inside a rule $A \to \ldots \& B \& \ldots$. Accordingly, once these substitutions are done, the grammar G' will contain conjuncts of the form (3a) and (3b), while all conjuncts of the form (3c) and (3d) will be eliminated. Then G' will be in the odd normal form. □

The grammar constructed in Lemma 1 generates the odd subset of the given language. However, it actually encodes the entire information defined in the original grammar, and using the "even rules" allowed in the odd normal form one can generate the original language as it is.

Theorem 1. *For every conjunctive grammar there exists and can be effectively constructed a conjunctive grammar in odd normal form generating the same language.*

Proof. Let $L \subseteq \Sigma^*$ be conjunctive. Since every conjunctive language can be generated by a conjunctive grammar in binary normal form (which can be obtained effectively), there is, by Lemma 1, a conjunctive grammar $G = (\Sigma, N, P, S)$ in odd normal form wihout even rules, such that for all $a \in \Sigma$,

$$L_G(S) = L \cap \mathrm{Odd} \qquad \text{and} \qquad L_G({}_aS_\varepsilon) = a^{-1}L \cap \mathrm{Odd}.$$

The grammar $G' := (\Sigma, N \cup \{S'\}, P', S')$ with a new nonterminal S' and

$$P' := P \cup \{S' \to \varphi \mid S \to \varphi \in P\} \cup \{S' \to a\,{}_aS_\varepsilon \mid a \in \Sigma\} \cup \{S' \to \varepsilon \mid \varepsilon \in L\}$$

is in odd normal form (with even rules) and generates L:

$$\begin{aligned}
L_{G'}(S') &= L_G(S) \cup \bigcup_{a \in \Sigma} a L_G({}_aS_\varepsilon) \cup (L \cap \{\varepsilon\}) \\
&= (L \cap \mathrm{Odd}) \cup \bigcup_{a \in \Sigma} a(a^{-1}L \cap \mathrm{Odd}) \cup (L \cap \{\varepsilon\}) \\
&= (L \cap \mathrm{Odd}) \cup \bigcup_{a \in \Sigma} (aa^{-1}L \cap a\mathrm{Odd}) \cup (L \cap \{\varepsilon\}) \\
&= (L \cap \mathrm{Odd}) \cup (L \cap \Sigma\mathrm{Odd}) \cup (L \cap \{\varepsilon\}) \\
&= L
\end{aligned}$$

If $L \cap \mathrm{Even} = \varnothing$, that is, if L does not contain strings of even length, then $L_{G'}({}_aS_\varepsilon) = \varnothing$ for every $a \in \Sigma$. Unfortunately, checking this property is undecidable in the general case, but if this property holds, then the even rules can be removed without changing the generated language. □

Some corollaries can be inferred. The first one concerns Greibach normal form for conjunctive grammars. As already mentioned, it is unknown whether every conjunctive grammar can be transformed to that form. However, Theorem 1 straightforwardly implies a transformation to Greibach normal form for grammars over a one-letter alphabet.

Corollary 1 (Unary Greibach normal form)**.** *For every conjunctive grammar over a unary alphabet there exists and can be effectively constructed a conjunctive grammar in Greibach normal form generating the same language.*

Indeed, since concatenation of languages over $\{a\}$ is commutative, each term BaC in an odd normal form grammar can be equivalently replaced by aBC.

The second consequence of Theorem 1 is actually quite obvious, but nevertheless it is new:

Theorem 2. *Conjunctive languages are effectively closed under quotient with letters, and hence under quotient with finite languages.*

Proof. Let $L \subseteq \Sigma^*$ be conjunctive and fix $a \in \Sigma$. By Lemma 1, there is a conjunctive grammar $G = (\Sigma, N, P, S)$, which contains nonterminal symbols S_a and ${}_bS_a$ for all $b \in \Sigma$ that generate the languages

$$L_G(S_a) = La^{-1} \cap \mathrm{Odd} \qquad \text{and} \qquad L_G({}_bS_a) = b^{-1}La^{-1} \cap \mathrm{Odd}.$$

Construct the grammar $G' = (\Sigma, N \cup \{S'\}, P \cup P', S')$ with the following additional rules:

$$\begin{aligned}
&S' \to S_a \\
&S' \to b\,{}_bS_a && \text{(for all } b \in \Sigma) \\
&S' \to \varepsilon && \text{(if } a \in L(G))
\end{aligned}$$

Then we have $L(G') = L(G)a^{-1}$. The construction for $a^{-1}L$ is symmetric. □

4 Restricted conjunctive grammars

Now let us define a restricted subfamily of conjunctive grammars that will be studied in this paper.

Definition 6. *A* restricted conjunctive grammar *is a conjunctive grammar in which every nonterminal may have at most one rule not of the form $A \to w$, with $w \in \Sigma^*$. In other words, the rules for every nonterminal A are of the form:*

$$A \to \alpha_1 \& \ldots \& \alpha_n \mid w_1 \mid \ \ldots \ \mid w_m \quad (n \geqslant 1,\ m \geqslant 0,\ \alpha_i \in (\Sigma \cup N)^*,\ w_j \in \Sigma^*)$$

A context-free grammar satisfying this restriction is known as a *single tree grammar*, see Greibach et al. [3].

The grammar in Example 2 is restricted conjunctive, while the grammar in Example 1 is not. The next example illustrates the key expressive power of these grammars.

Example 3. *The following restricted conjunctive grammar generates the set of palindromes of odd length over $\{a, b\}$:*

$$\begin{aligned} S &\to AB \& O \mid a \mid b \\ A &\to aSa \mid \varepsilon \\ B &\to bSb \mid \varepsilon \\ O &\to OOO \mid a \mid b \end{aligned}$$

Here the nonterminal O generates the set Odd, and hence S may generate only strings of odd length. Then the rule $S \to AB \& O$ generates

$$(aSa \cup \{\varepsilon\})(bSb \cup \{\varepsilon\}) \cap \mathrm{Odd} = (aSabSb \cup aSa \cup bSb \cup \{\varepsilon\}) \cap \mathrm{Odd} = aSa \cup bSb,$$

that is, it is equivalent to two rules $S \to aSa$ and $S \to bSb$. Thus the set of odd-length palindromes is generated inductively, starting from a and b.

This representation of the union of two languages actually works in the general context, as long as both languages consist of strings of odd length. As in the above example, it is sufficient to add the empty string to both languages, concatenate them and then filter out the strings of even length. This gives a way to simulate every conjunctive grammar in which every nonterminal generates a subset of Odd. Since grammars in odd normal form have this property and every conjunctive grammar can be transformed to this form, the following statement can be proved.

Lemma 2. *For every conjunctive grammar generating a subset of* Odd $\subseteq \Sigma^*$ *there exists and can be effectively constructed a restricted conjunctive grammar generating the same language.*

Proof. Assume that the conjunctive grammar G is in odd normal form without even rules (by Theorem 1). The first goal is to transform it so that for every nonterminal A there is either a unique rule of an arbitrary form $A \to \alpha_1 \& \ldots \& \alpha_n$, or two rules $A \to B \mid C$. During this transformation, the property that $L(A) \subseteq \text{Odd}$ for every nonterminal A should be retained.

Let $A \to r_1 \mid r_2 \mid \ldots \mid r_n$ be the rules for the nonterminal A. Of course, $L(r_i) \subseteq \text{Odd}$ for all i. If $n \geqslant 2$, then the rules for A are replaced with $A \to B \mid C$, where B and C are two new nonterminals with the rules $B \to r_1 \mid r_2 \mid \ldots \mid r_{n-1}$ and $C \to r_n$. Observe that iterative application of this transformation results in a grammar $G' = (\Sigma, N, P, S)$ that generates the same language as G, still has $L(A) \subseteq \text{Odd}$ for all $A \in N$, and furthermore, for every nonterminal $A \in N$ there is either a unique rule of an arbitrary form or two rules $A \to B \mid C$.

Next, construct a restricted conjunctive grammar $G'' = (\Sigma, N \cup N' \cup \{O\}, P', S)$, in which $N' = \{\, A' \mid A \in N \,\}$ is a disjoint copy of N, O is a new nonterminal, and P' contains the following rules:

$A' \to A \mid \varepsilon$	(for all $A \in N$)
$A \to \alpha_1 \& \ldots \& \alpha_n \& O$	(if $A \to \alpha_1 \& \ldots \& \alpha_n$ is the unique rule for A in P)
$A \to B'C' \& O$	(if $A \to B \mid C$ are the rules for A in P)
$O \to OOO$	
$O \to a$	(for all $a \in \Sigma$)

Here it obviously that $L_{G'}(O) = \text{Odd}$, $L_{G'}(A) \subseteq \text{Odd}$ and $L_{G'}(A') = L_{G'}(A) \cup \{\varepsilon\}$ for every $A \in N$. Assume now that the nonterminal A has the rule $A \to B'C' \& O$ in P'. Then

$$\begin{aligned} L_{G'}(A) &= \big(L_{G'}(B) \cup \{\varepsilon\}\big)\big(L_{G'}(C) \cup \{\varepsilon\}\big) \cap \text{Odd} \\ &= \big(L_{G'}(B)L_{G'}(C) \cup L_{G'}(B) \cup L_{G'}(C) \cup \{\varepsilon\}\big) \cap \text{Odd} \\ &= L_{G'}(B) \cup L_{G'}(C). \end{aligned}$$

This means that the rule $A \to B'C' \& O$ can be equivalently replaced by the rules $A \to B \mid C$. One then gets a grammar obtained from G' by changing every rule $A \to \alpha_1 \& \ldots \& \alpha_n$ (where this is the only rule for A) to $A \to \alpha_1 \& \ldots \& \alpha_n \& O$ and adding the rules for O (rules of the type $A \to B \mid C$ stay the same by the above equation, and the nonterminals A' are superfluous). Since the nonterminals in G' produce only subsets of Odd, this conjunction with O does not change the generated language and the lemma is proved. □

Lemma 2 can be used to construct a restricted conjunctive grammar for the language containing all odd strings belonging to a given conjunctive language L and no even strings. In order to get the whole language L later, it is useful to generate *all* even strings: this will be a superset of L, which could be intersected with some other languages to obtain L. The addition of all even strings is performed in the following lemma.

Lemma 3. *For every conjunctive language $L \subseteq \Sigma^*$, the language* $(L \cap \text{Odd}) \cup \text{Even}$ *is generated by a restricted conjunctive grammar.*

Proof. Let $G = (\Sigma, N, P, S)$ be a restricted conjunctive grammar generating the language $L \cap \text{Odd}$, which is given by Lemma 2. Construct a new grammar G' with the following rules:

$$\begin{array}{lllllll}
S' & \to & AB \& C \mid \varepsilon & \qquad & C & \to & CC \\
A & \to & S & & C & \to & w \quad (w \in (\Sigma \cap L) \cup \Sigma^2 \cup \Sigma^3) \\
A & \to & a \quad (a \in \Sigma) & & O & \to & OOO \\
B & \to & O \mid \varepsilon & & O & \to & a \quad (a \in \Sigma)
\end{array}$$

This concatenation AB generates the following language:

$$\big((L \cap \text{Odd}) \cup \Sigma\big) \cdot (\text{Odd} \cup \{\varepsilon\}) = (\text{Even} \setminus \{\varepsilon\}) \cup (L \cap \text{Odd}) \cup \Sigma$$

Its intersection with $L(C) = (\Sigma^+ \setminus \{\, a \in \Sigma \mid a \notin L \,\})$ produces $(\text{Even} \setminus \{\varepsilon\}) \cup (L \cap \text{Odd})$, and taking the rule $S' \to \varepsilon$ into account, the grammar generates $\text{Even} \cup (L \cap \text{Odd})$. ☐

The above construction cannot be used symmetrically to obtain the language $(L \cap \text{Even}) \cup \text{Odd}$ directly. However, the method of Lemma 3 can be elaborated to generate the following superset of L:

Lemma 4. *For every conjunctive language $L \subseteq \Sigma^*$ and for every symbol $a \in \Sigma$, the language $(L \cap a\text{Odd}) \cup \overline{a\text{Odd}}$ is generated by some restricted conjunctive grammar.*

Proof. Let L be a conjunctive language over Σ and let $a \in \Sigma$. Define $L_a := a(a^{-1}L \cap \text{Odd})$: these are all even strings in L that start with a, that is, $L_a = L \cap a\text{Odd}$. Define the following three languages:

$$\begin{aligned}
L_1 &= \{\varepsilon\} \cup (\Sigma \setminus \{a\})\Sigma^*, \\
L_2 &= L_a \cup \{\varepsilon\}, \\
L_3 &= \text{Odd} \cup \{\varepsilon\}.
\end{aligned}$$

Each of these languages has a restricted conjunctive grammar. It is not difficult to construct such grammars for L_1 and L_3. For L_2, since L is conjunctive, the language $a^{-1}L \cap \text{Odd}$ is conjunctive by Theorem 2, and therefore, by Lemma 2, there is a restricted conjunctive grammar generating this language. This grammar can be easily modified to generate L_2.

Now consider the concatenation of these three languages:

$$\begin{aligned}
L_1 L_2 L_3 &= \big(\{\varepsilon\} \cup L_a \cup (\Sigma \setminus \{a\})\Sigma^* L_a \cup (\Sigma \setminus \{a\})\Sigma^*\big) \cdot (\text{Odd} \cup \{\varepsilon\}) = \\
&= \big(\{\varepsilon\} \cup L_a \cup (\Sigma \setminus \{a\})\Sigma^*\big) \cdot (\text{Odd} \cup \{\varepsilon\}) = \\
&= \{\varepsilon\} \cup L_a \cup (\Sigma \setminus \{a\})\Sigma^* \cup \text{Odd} \cup \underbrace{L_a \text{Odd}}_{\subseteq \text{Odd}} \cup \underbrace{(\Sigma \setminus \{a\})\Sigma^* \text{Odd}}_{\subseteq (\Sigma \setminus \{a\})\Sigma^*} = \\
&= L_a \cup \text{Odd} \cup (\text{Even} \setminus a\Sigma^*) = L_a \cup \overline{a\text{Odd}} = (L \cap a\text{Odd}) \cup \overline{a\text{Odd}}.
\end{aligned}$$

Using this equation, it suffices to construct a restricted conjunctive grammar for each of the three languages $(\varepsilon \cup (\Sigma \setminus \{a\})\Sigma^*)$, $(L_a \cup \varepsilon)$ and $(\mathrm{Odd} \cup \varepsilon)$, which will now be done.

1. $\varepsilon \cup (\Sigma \setminus \{a\})\Sigma^*$: Observe the restricted conjunctive grammar $G := (\Sigma, \{S, A, X\}, S, P)$ with P containing the rules $S \to AX$, $S \to \varepsilon$, $A \to b$ (for all $b \in \Sigma \setminus \{a\}$), $X \to XX$, $X \to v$ (for all $v \in \Sigma \cup \{\varepsilon\}$). The nonterminals produce the following languages: $L_G(X) = \Sigma^*$, $L_G(A) = \Sigma \setminus \{a\}$ and $L_G(S) = \varepsilon \cup (\Sigma \setminus \{a\})\Sigma^*$. So the grammar produces the desired language.

2. $\mathrm{Odd} \cup \varepsilon$: This language is obviously generated by a restricted conjunctive grammar with the rules $S \to O \mid \varepsilon$, $O \to OOO$, $O \to a$ for every $a \in \Sigma$.

From this, using the grammars for L_1, L_2 and L_3, it is easy to construct a restricted conjunctive grammar for the desired language. $\square$

It remains to intersect $|\Sigma| + 1$ languages constructed in Lemmata 3 and 4 to obtain a grammar for any conjunctive language L containing ε. This gives the main result of this paper:

Theorem 3. *Every conjunctive language is generated by a restricted conjunctive grammar.*

Proof. Let $L \subseteq \Sigma^*$ be any conjunctive language. Then, by Lemmata 3 and 4, there are restricted conjunctive grammars for the languages $(L \cap \mathrm{Odd}) \cup \mathrm{Even}$ and $(L \cap a\mathrm{Odd}) \cup \overline{a\mathrm{Odd}}$ for any $a \in \Sigma$. The intersection of these languages is

$$\big((L \cap \mathrm{Odd}) \cup \mathrm{Even}\big) \cap \bigcap_{a \in \Sigma} \big((L \cap a\mathrm{Odd}) \cup \overline{a\mathrm{Odd}}\big) = L \cup \{\varepsilon\}.$$

If $\varepsilon \in L$, this immediately gives a restricted conjunctive grammar for L. Otherwise, if $\varepsilon \notin L$, then a subsequent conjunction with a nonterminal representing Σ^+ yields the required grammar. $\square$

5 Restricted conjunctive grammars without ε-rules

The above simulation of an arbitrary conjunctive grammar by a conjunctive grammar with restricted disjunction essentially uses rules of the form $A \to \varepsilon$, known as *ε-rules*. On the other hand, it is known that conjunctive grammars of the general form do not need ε-rules, and a transformation to the binary normal form leads to their elimination. This raises the question of whether restricted conjunctive grammars without ε-rules are as powerful as conjunctive grammars of the general form.

First of all, this stronger restriction on conjunctive grammars still gives a non-trivial family. For instance, the important grammar over a unary alphabet given in Example 2 is of this form. Grammars for interesting languages over larger alphabets can be constructed as well.

Example 4. *The following restricted conjunctive grammar generates the set of all palindromes:*

$$\begin{array}{lcl@{\qquad\qquad}lcl}
S & \rightarrow & XSX\&T \mid a \mid b \mid aa \mid bb & A & \rightarrow & bE \mid a \mid b \\
T & \rightarrow & AB\&CD\&XXE & B & \rightarrow & Ea \mid a \mid b \\
E & \rightarrow & XE \mid a \mid b & C & \rightarrow & aE \mid a \mid b \\
X & \rightarrow & a \mid b & D & \rightarrow & Eb \mid a \mid b
\end{array}$$

In particular, $L(E) = \Sigma^+$, $L(A) = b\Sigma^* \cup \{a\}$, $L(B) = \Sigma^*a \cup \{b\}$, $L(C) = a\Sigma^* \cup \{b\}$, $L(D) = \Sigma^*b \cup \{a\}$, *and* $L(T) = a\Sigma^+a \cup b\Sigma^+b$.

Consider the intersection $L(AB) \cap L(CD)$ used in the rule for T:

$$\begin{aligned}
&(b\Sigma^* \cup \{a\})(\Sigma^*a \cup \{b\}) \cap (a\Sigma^* \cup \{b\})(\Sigma^*b \cup \{a\}) = \\
&\quad = \big(b\Sigma^*a \cup b\Sigma^*b \cup a\Sigma^*a \cup \{ab\}\big) \cap \big(a\Sigma^*b \cup a\Sigma^*a \cup b\Sigma^*b \cup \{ba\}\big) = \\
&\qquad\qquad\qquad\qquad\qquad\qquad = a\Sigma^*a \cup b\Sigma^*b \cup \{ab, ba\},
\end{aligned}$$

and the subsequent intersection with the set of all strings of length at least 3 produces the intended language $a\Sigma^+a \cup b\Sigma^+b$. Finally, the rule $S \rightarrow XSX\&T$ generates the language

$$\{a, b\}S\{a, b\} \cap (a\Sigma^*a \cup b\Sigma^*b) = aSa \cup bSb,$$

and hence operates as if two rules $S \rightarrow aSa$ and $S \rightarrow bSb$. This is enough to generate all palindromes inductively, starting from the base set $\{a, b, aa, bb\}$.

Lemma 5. *The family of languages generated by restricted conjunctive grammars without ε-rules is closed under union with finite sets, concatenation and intersection.*

Proof. The closure under concatenation and under intersection is obvious. For the union with finite sets, let $F \subseteq \Sigma^+$ be finite and let $G = (\Sigma, N, P, S)$ be a restricted conjunctive grammar without ε-rules. The grammar $(\Sigma, N \cup \{S'\}, P \cup \{S' \rightarrow S\} \cup \{S' \rightarrow w \mid w \in F\}, S')$ with the new nonterminal S' is restricted, does not contain ε-rules and obviously generates $L(G) \cup F$. $\square$

Lemma 6. *Any finite and co-finite language can be generated by a restricted conjunctive grammar without ε-rules.*

Proof. Since the empty set can obviously be generated by such a grammar and the respective class of languages is closed under union with finite sets by Lemma 5, we get the first part of the assertion.

Let now $L \subseteq \Sigma^+$ be co-finite. Then there is some $k \geq 1$ such that $L \cap \Sigma^k\Sigma^+ = \Sigma^k\Sigma^+$ and there exists a finite set F such that $L = \Sigma^k\Sigma^+ \cup F$. Obviously, $\Sigma^k\Sigma^+$ can be generated by a restricted conjunctive grammar without ε-rules. Since L is the union of $\Sigma^k\Sigma^+$ with the finite set F, there is also a restricted conjunctive grammar without ε-rules for L after Lemma 5. $\square$

The following theorem exhibits a significant subfamily of conjunctive grammars that can be simulated by restricted conjunctive grammars without ε-rules. This form generalizes deterministic Greibach normal form.

Theorem 4. *Let $G = (\Sigma, N, P, S)$ be a conjunctive grammar without ε-rules, in which there is a disjoint partition of its nonterminals $N = N_S \cup N_L \cup N_R$ into* simple, left *and* right *nonterminals, respectively, such that:*

- *for every $A \in N_L$ and for every $a \in \Sigma$ there is at most one rule $A \to a\alpha_1\& \dots \& a\alpha_n$ with $n \geqslant 1$ and $\alpha_i \in N^+$, and all complex rules for A are of this form;*
- *for every $A \in N_R$ and for every $a \in \Sigma$ there is at most one rule $A \to \alpha_1 a\& \dots \&\alpha_n a$ with $n \geqslant 1$ and $\alpha_i \in N^+$, and all complex rules for A are of this form;*
- *every $A \in N_S$ has at most one complex rule.*

Then there exists (and can be effectively constructed) a restricted conjunctive grammar without ε-rules that generates the same language.

For instance, the grammar in Example 1 can be easily transformed to fit this statement with $N_L = \{D, A, B\}$ and $N_R = \varnothing$. Therefore, there is a restricted conjunctive grammar without ε-rules for $\{\, wcw \mid w \in \{a, b\}^* \,\}$.

Proof. Let $G = (\Sigma, N, P, S)$ be a conjunctive grammar of the stated form. Construct a grammar $G' := (\Sigma, N', P', S)$ such that

$$\begin{aligned} N' := N \cup & \{A_a \mid A \in N_R \cup N_L,\ a \in \Sigma\} \\ & \cup \{X_a \mid a \in \Sigma\} \cup \{{}_aX \mid a \in \Sigma\} \cup \{T\} \cup N_2 \end{aligned}$$

(where N_2 is a set of auxiliary nonterminals that will not be explicitly described) and the set P' contains the following rules. *Simple* nonterminals have the same rules as in P:

$$\begin{aligned} & A \to \alpha_1\& \dots \&\alpha_n && (A \in N_S,\ A \to \alpha_1\& \dots \&\alpha_n \in P) \\ & A \to w && (A \to w \in P,\ w \in \Sigma^+) \end{aligned}$$

Left and *right* nonterminals have the following rules:

$$
\begin{array}{ll}
A \to T \& \mathop{\&}_{a \in \Sigma} (X_a \cdot A_a) & (A \in N_L) \\
A \to T \& \mathop{\&}_{a \in \Sigma} (A_a \cdot {}_aX) & (A \in N_R) \\
A \to w & (A \to w \in P,\ w \in \Sigma^+) \\
A \to w & (w \in L_G(A),\ |w| = 2),
\end{array}
$$

that is, all simple rules from P are are retained, all two-symbol strings are explicitly generated, and the unique long rule is simulated by the nonterminals A_a as follows:

$$
\begin{array}{ll}
A_a \to b & (A \in N_R \cup N_L,\ a, b \in \Sigma) \\
A_a \to \alpha_1 \& \ldots \& \alpha_n & (A \in N_R \cup N_L,\ a \in \Sigma; \\
 & \quad \text{if } A \to a\alpha_1 \& \ldots a\alpha_n \in P \text{ or } A \to \alpha_1 a \& \ldots \alpha_n a \in P)
\end{array}
$$

Additionally, the rules for the nonterminals X_a and ${}_aX$ (for every $a \in \Sigma$) and T are constructed so that $L_{G'}(X_a) = \{a\} \cup (\Sigma \setminus \{a\})\Sigma^*$, $L_{G'}({}_aX) = \{a\} \cup \Sigma^*(\Sigma \setminus \{a\})$ and $L_{G'}(T) = \Sigma^+ \setminus \Sigma^2$. This can be done by Lemmata 5 and 6 using the additional nonterminals from the set N_2.

Note that because of the restrictions on G, there is at most one complex rule for every nonterminal $A \in N'$, and thus G' is of the restricted form without ε-rules. The correctness of the construction is stated as follows:

For every $A \in N$ and $a \in \Sigma$ it holds that

1. $L_{G'}(A_a) = \begin{cases} \Sigma \cup \bigcap_{i=1}^{n} L_G(\alpha_i) & \text{if } P \text{ contains a rule } A \to a\alpha_1 \& \ldots \& a\alpha_n \text{ or } A \to \alpha_1 a \& \ldots \& \alpha_n a \text{ with } n \geqslant 1,\ \alpha_i \in N^+ \\ \Sigma & \text{otherwise.} \end{cases}$

   (provided that $A \notin N_S$)

2. $L_{G'}(A) = L_G(A)$ for all $A \in N$.

The claim is verified by showing that these languages form the least solution of the system of language equations corresponding to G'. Note that no nonterminal in G' generates the empty string and G' does not contain unit conjuncts. Then the system of language equations corresponding to G' is known to have a unique ε-free solution, which is the least solution of the system. Since the above languages and the previously mentioned languages generated by T, X_a and ${}_aX$ do not contain ε, it only remains to verify these solutions by substitution.

For every $A \in N_R \cup N_L$ and $a \in \Sigma$, consider the equation for A_a in the system corresponding to G'. If there are no complex rules for A in G, then the only rules for A_a in G' are the rules $A_a \to b$ for all $b \in \Sigma$. so the equation for A_a is of the form

$$A_a = \bigcup_{b \in \Sigma} \{b\},$$

which is turned to an equality $\Sigma = \Sigma$ under the given substitution.

Assume that there is a rule $A \to a\alpha_1 \& \dots a\alpha_n$ (or $A \to \alpha_1 a \& \dots \alpha_n a$) in G. Then G' contains an additional rule $A_a \to \alpha_1 \& \dots \& \alpha_n$ with $\alpha_i \in N^+$, and accordingly the equation for A_a is

$$A_a = \bigcap_{i=1}^{n} \alpha_n \cup \bigcup_{b \in \Sigma} \{b\}.$$

Under the substitution $B = L_G(B)$ for each $B \in N$, the right-hand side of this equation has the value $\bigcap_{i=1}^{n} L_G(\alpha_i) \cup \Sigma$, and thus the equality holds.

The equation for any $A \in N_S$ in G' is the same as in G, and thus the substitution $B = L_G(B)$ for all $B \in N$ turns it into an equality. Consider the equation for any *left* nonterminal $A \in N_L$ in G', which is of the form

$$A = \Big(T \cap \bigcap_{a \in \Sigma} X_a \cdot A_a\Big) \cup \bigcup_{A \to w \in P} \{w\} \cup \bigcup_{w \in \Sigma^2 \cap L_G(A)} \{w\}$$

Under the substitution $X_a = \{a\} \cup (\Sigma \setminus \{a\})\Sigma^*$ and $A_a = L_{A,a}$ for any $L_{A,a}$ with $\Sigma \subseteq L_{A,a} \subseteq \Sigma^+$, the subexpression $\bigcap_{a \in \Sigma} X_a \cdot A_a$ evaluates to

$$\begin{aligned} \bigcap_{a \in \Sigma} \big[\{a\} \cup (\Sigma \setminus \{a\})\Sigma^*\big] L_{A,a} &= \bigcap_{a \in \Sigma} \big[aL_{A,a} \cup (\Sigma \setminus \{a\})\Sigma^* L_{A,a}\big] = \\ &= \bigcap_{a \in \Sigma} \big[aL_{A,a} \cup (\Sigma \setminus \{a\})\Sigma^+\big] = \bigcup_{a \in \Sigma} aL_{A,a}. \end{aligned}$$

Let $\Sigma_A \subseteq \Sigma$ be the set of all terminal symbols a, for which there is a long rule $A \to a\alpha_{a,1} \& \dots \& a\alpha_{a,n_a} \in P$. Now the first part of the right-hand side of the equation for A in G', i.e. $\Big(T \cap \bigcap_{a \in \Sigma} X_a \cdot A_a\Big)$, under the substitution $T = \Sigma^+ \setminus \Sigma^2$, $X_a = \{a\} \cup (\Sigma \setminus \{a\})\Sigma^*$ for all $a \in \Sigma$, $A_a = \Sigma \cup \bigcap_{i=1}^{n_a} L_G(\alpha_{a,i})$ for $a \in \Sigma_A$ and $A_a = \Sigma$ for $a \in \Sigma \setminus \Sigma_A$ evaluates to

$$\begin{aligned} \Big((\Sigma^+ \setminus \Sigma^2) \cap \big((\Sigma \setminus \Sigma_A)\Sigma \cup \bigcup_{a \in \Sigma_A} a\big(\Sigma \cup \bigcap_{i=1}^{n_a} L_G(\alpha_{a,i})\big)\big)\Big) &= \\ = \big(\bigcup_{a \in \Sigma_A} \bigcap_{i=1}^{n_a} aL_G(\alpha_{a,i})\big) \setminus \Sigma^2. \end{aligned}$$

Thus the value of the entire right-hand side of the equation for A in G' is

$$\big(\bigcup_{a \in \Sigma_A} \bigcap_{i=1}^{n_a} aL_G(\alpha_{a,i})\big) \setminus \Sigma^2 \cup \{\, w \mid A \to w \in P \,\} \cup \big(L_G(A) \cap \Sigma^2\big) = L_G(A).$$

The case of *right* nonterminals is handled symmetrically.

This shows, in particular, that the least solution of the system corresponding to G' has $S = L_G(S)$, that is, $L_{G'}(S) = L_G(S)$, which proves the theorem. $\square$

Corollary 2. *Every regular language $L \subseteq \Sigma^+$ is restricted conjunctive without ε-rules.*

Proof. Let $L \subseteq \Sigma^+$ be regular via the deterministic finite automaton $A = (\Sigma, Q, q_0, \delta, F)$. Then the grammar $G := (\Sigma, N, P, S)$ with $N = \{ A_q \mid q \in Q \}$, $P = \{A_q \to aA_{\delta(q,a)} \mid q \in Q, a \in \Sigma\} \cup \{A_q \to a \mid \delta(q, a) \in F\}$ and $S = A_{q_0}$ generates L and is in the form required by Theorem 4. Then the theorem implies that there is a restricted conjunctive grammar without ε-rules for L. $\square$

The exact expressive power of conjunctive grammars with restricted disjunction and without ε-rules is left as an open question to study. In particular, it would be interesting to investigate it in the case of a unary alphabet: perhaps they can generate all unary conjunctive languages. For larger alphabets, these grammars likely generate a proper subfamily of conjunctive languages.

Acknowledgements

Supported by the Academy of Finland under grant 118540.

References

[1] S. Ginsburg, H. G. Rice, "Two families of languages related to ALGOL", *Journal of the ACM*, 9 (1962), 350–371.

[2] S. A. Greibach, "A new normal-form theorem for context-free phrase structure grammars", *Journal of the ACM*, 12 (1965), 42–52.

[3] S. A. Greibach, W. Shi, S. Simonson, "Single tree grammars", in: J. D. Ullman (Ed.), *Theoretical Studies in Computer Science*, Academic Press, 1992, 73–99.

[4] A. Jeż, "Conjunctive grammars generate non-regular unary languages", *International Journal of Foundations of Computer Science*, 19:3 (2008), 597–615.

[5] A. Jeż, A. Okhotin, "Conjunctive grammars over a unary alphabet: undecidability and unbounded growth", *Theory of Computing Systems*, to appear.

[6] A. Jeż, A. Okhotin, "Complexity of solutions of equations over sets of natural numbers", *25th Annual Symposium on Theoretical Aspects of Computer Science* (STACS 2008, Bordeaux, France, 21–23 February, 2008), 373–383.

[7] A. Jeż, A. Okhotin, "On the computational completeness of equations over sets of natural numbers", *Automata, Languages and Programming* (ICALP 2008, Reykjavík, Iceland, July 6–13, 2008), part II, LNCS 5126, 63–74.

[8] A. Okhotin, "Conjunctive grammars", *Journal of Automata, Languages and Combinatorics*, 6:4 (2001), 519–535.

[9] A. Okhotin, "Unambiguous Boolean grammars", *Information and Computation*, 206 (2008), 1234–1247.

[10] C. Reitwießner, *Formal Languages Defined by Recurrent Circuits*, Diplomarbeit, University of Würzburg, 2007.

TURKU
CENTRE *for*
COMPUTER
SCIENCE

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku
- Department of Information Technology
- Department of Mathematical Sciences

Åbo Akademi University
- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration
- Institute of Information Systems Sciences

ISBN 978–952–12–2151–4
ISSN 1239-1891