INSTYTUT INFORMATYKI
UNIWERSYTETU WROCŁAWSKIEGO

INSTITUTE OF COMPUTER SCIENCE
UNIVERSITY OF WROCŁAW

ul. Joliot-Curie 15
50–383 Wrocław
Poland

Report 01/07

Artur Jeż

Conjunctive grammars can generate non-regular unary languages

April 2007

Conjunctive grammars can generate non-regular unary languages

Artur Jeż ⋆

Institute of Computer Science,
ul. Joliot-Curie 15, 50-383 Wroclaw, Poland
aje@ii.uni.wroc.pl
http://www.ii.uni.wroc.pl/~aje

Abstract. Conjunctive grammars were introduced by A. Okhotin in [1] as a natural extension of context-free grammars with an additional operation of intersection in the body of any production of the grammar. Several theorems and algorithms for context-free grammars generalize to the conjunctive case. Still some questions remained open. A. Okhotin posed nine problems concerning those grammars. One of them was a question, whether a conjunctive grammar over unary alphabet can generate only regular languages. We give a negative answer, contrary to the conjectured positive one, by constructing a conjunctive grammar for the language $\{a^{4^n} : n \in \mathbb{N}\}$. We then generalise this result—for every set of natural numbers L, such that their representation in some k-ary system is a regular set, we show that $\{a^n : n \in L\}$ is generated by a conjunctive grammar over unary alphabet and that this grammar can be efficiently computed.

Key words: Conjunctive grammars, regular languages, unary alphabet, non-regular languages

1 Introduction and background

Alexander Okhotin introduced conjunctive grammars in [1] as a simple, yet powerful extension of context-free grammars. Informally speaking, conjunctive grammars allow additional use of intersection in the body of any rule of the grammar. More formally, conjunctive grammar is defined as a quadruple $\langle \Sigma, N, P, S \rangle$ where Σ is a finite alphabet, N is a set of nonterminal symbols, S is a starting nonterminal symbol and P is a set of productions of the form:

$$A \to \alpha_1 \& \alpha_2 \& \dots \& \alpha_k, \quad \text{where } \alpha_i \in (\Sigma \cup N)^* \ .$$

Word w can be derived by this rule if and only if it can be derived from every string α_i for $i = 1, \dots, k$, and $\alpha_i = N_1 \dots N_l$ can derive word w if $w = w_1 \dots w_l$ and N_j can derive word w_j for $j = 1, \dots, l$.

We can also give semantics of conjunctive grammars with language equations that use sum, intersection and concatenation as allowed operations. Language generated by conjunctive grammar is a least solution of such equations (or rather, coordinate of the solution corresponding to the starting non-termninal S, since solution is a vector of languages).

The usage of intersection allows us to define many natural languages that are not context-free. On the other hand it can be shown [1] that languages generated by the conjunctive grammars are deterministic context-sensitive.

We give a simple example of a conjunctive grammar together with its language equations here and a formal definition in Section 2. The Reader interested in the whole theory of the conjunctive grammars should consult [1] for detailed results or [3] for shorter overview. Also work on the Boolean grammars [4], which extend conjunctive grammars by additional use of negation, may be interesting.

⋆ Research supported by MNiSW grant number N206 024 31/3826, 2006-2008.

Example 1. Let us consider conjunctive grammar $\langle \Sigma, N, P, S\rangle$ defined by:

$$\begin{aligned}
&\Sigma = \{a, b, c\} \ , \\
&N = \{S, B, C, E, A\} \ , \\
&P = \{A \to aA \mid \epsilon, C \to Cc \mid \epsilon, S \to (AE)\&(BC), \\
&\quad\quad B \to aBb \mid \epsilon, E \to bEc \mid \epsilon\} \ .
\end{aligned}$$

The language generated by this grammar is equal to $\{a^n b^n c^n : n \in \mathbb{N}\}$. The associated language equations are:

$$\begin{aligned}
L_A &= \{a\}L_A \cup \{\epsilon\} \ , \\
L_C &= \{c\}L_C \cup \{\epsilon\} \ , \\
L_S &= (L_A L_E) \cap (L_B L_C) \ , \\
L_B &= \{a\}L_B\{b\} \cup \{\epsilon\} \ , \\
L_E &= \{b\}L_E\{c\} \cup \{\epsilon\} \ .
\end{aligned}$$

Their least solution is:

$$\begin{aligned}
L_A &= a^* \ , \\
L_C &= c^* \ , \\
L_S &= \{a^n b^n c^n : n \in \mathbb{N}\} \ , \\
L_B &= \{a^n b^n : n \in \mathbb{N}\} \ , \\
L_E &= \{b^n c^n : n \in \mathbb{N}\} \ .
\end{aligned}$$

Many natural techniques and properties generalize from context-free grammars to conjunctive grammars. Among them most important are: existence of the Chomsky normal form, parsing using a modification of CYK algorithm *etc.* On the other hand many other techniques do not generalize—there is no Pumping Lemma for conjunctive grammars, they do not have bounded growth property, non-emptiness is undecidable. In particular no technique for showing that a language cannot be generated by conjunctive grammars is known; in fact, as for today, we are only able to show that languages that are not context sensitive lay outside this class of languages.

A. Okhotin in [5] gathered nine open problems for conjunctive and Boolean grammars considered to be the most important in this field. One of those problem was a question, whether conjunctive grammars over unary alphabet generate only regular languages. It is easy to show (using Pumping Lemma), that this is true in case of context-free grammars. The same result was conjectured for conjunctive grammars. We disprove this conjecture by giving conjunctive grammar for a non-regular language $\{a^{4^n} : n \in \mathbb{N}\}$.

The set $\{4^n : n \in \mathbb{N}\}$ written in binary is a regular language. This leads to a natural question, what is the relation between regular (over binary alphabet) languages and unary conjunctive languages. We prove that every regular language (written in some k-ary system) interpreted as a set of numbers can be represented by a conjunctive grammar over an unary alphabet.

1.1 Outline of the paper

The Section 2 contains the formal definition of conjunctive grammars as well as their semantics. It also gathers the needed general knowledge on the language equations. In Section 3 we present a conjunctive grammar for language $\{a^{4^n} : n \in \mathbb{N}\}$. In Section 4 we consider two topics arising from the results stated in Section 3—the number of non-terminals required to generate non-regular language and give similar construction for languages $\{a^{k^n} : n \in \mathbb{N}\}$ for any natural k. In Section 5 we further extend the main result by showing, that for any set of natural numbers L, such that the representation in k-positional system of L is regular, language $\{a^n : n \in L\}$ is generated by a conjunctive grammar over unary alphabet. In Section 6 we summarize the results and state the open problems.

2 Definitions and notation

2.1 Conjunctive grammars

Definition 1. *A conjunctive grammar [1] is a quadruple $G = \langle \Sigma, N, P, S \rangle$, in which Σ and N are disjoint finite non-empty sets of terminal and non-terminal symbols respectively; P is a finite set of grammar rules, each of the form*

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

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

Informally, a rule (1) states that if a word is generated by each α_i, then it is generated by A. This natural semantic of conjunctive grammar can be formalized in many ways, here we choose the formalism of language equations, which seems to be the easiest to handle in technical proofs. This definition is taken from [2].

Definition 2. *For every conjunctive grammar $G = \langle \Sigma, N, P, S \rangle$, the associated system of language equations [2] is a system of equations in variables N, in which each variable assumes a value of a language over Σ, and which contains the following equation for every variable A:*

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

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

Clearly the converse is true. For every set of language E equations with variables $A \in N$ of the form

$$A = \bigcup_i \bigcap_{j=1}^{m} \alpha_{i,j} \quad \text{, where } \alpha_{i,j} \in (A \cup \Sigma)^*$$

there exists a conjunctive grammar G with nonterminals N, such that E is a set of language equations associated with G.

2.2 Language equations

We prove theorems using mainly language equations, we gather the basic definitions and facts needed in this paper in this section. Reader familiar with the subject can skip this section.

Definition 3. *Let $(X_1, \dots, X_n)$ denote language variables. A resolved system of language equations is a system of a form*

$$X_i = \phi_i(X_1, \dots, X_n) \quad \textit{for } i = 1, \dots, n\ .$$

We abbreviate it to vector form

$$(\dots, X_i, \dots) = \phi(\dots, X_j, \dots)\ .$$

Since solutions are vectors of languages we use notation

$$(\dots, A_i, \dots) \subset (\dots, B_i, \dots)\ ,$$

meaning, that $A_i \subset B_i$ for $i = 1, \dots, n$.

We say that a language operation θ is *monotone* if

$$(\ldots, X_i, \ldots) \subset (\ldots, Y_i, \ldots) \quad \text{implies} \quad \theta(\ldots, X_i, \ldots) \subset \theta(\ldots, Y_i, \ldots) \ .$$

For example intersection, sum and concatenation are monotone while complementation is not. By basic knowledge of the language equations every resolved system of equations using only monotone operations has the least (with respect to inclusion) solution $(\ldots, S_i, \ldots)$ given by

$$(\ldots, S_i, \ldots) = \bigcup_{i=0}^{\infty} \phi^i(\ldots, \emptyset, \ldots)$$

We use a following alternative characterization of the least solution.

Lemma 1. *Let an operator ϕ associated with monotonic language equations be given and let $(\ldots, S_i, \ldots)$ be its least solution. If a vector of languages $(\ldots, X_i, \ldots)$ satisfies*

$$\phi(\ldots, X_i, \ldots) \subset (\ldots, X_i, \ldots) \quad \text{and} \quad (\ldots, X_i, \ldots) \subset (\ldots, S_i, \ldots)$$

then

$$(\ldots, X_i, \ldots) = (\ldots, S_i, \ldots) \ .$$

Proof. The least solution $(\ldots, S_i, \ldots)$ does exist and it is given by

$$(\ldots, S_i, \ldots) = \bigcup_{i=0}^{\infty} \phi^i(\ldots, \emptyset, \ldots) \ .$$

Clearly

$$(\ldots, \emptyset, \ldots) \subset (\ldots, X_i, \ldots) \subset (\ldots, S_i, \ldots) \ .$$

Since ϕ is monotone:

$$\bigcup_{i=0}^{\infty} \phi^i(\ldots, \emptyset, \ldots) \subset \bigcup_{i=0}^{\infty} \phi^i(\ldots, X_i, \ldots) \subset \bigcup_{i=0}^{\infty} \phi^i(\ldots, S_i, \ldots) \ .$$

Since $\phi(\ldots, X_i, \ldots) \subset (\ldots, X_i, \ldots)$ and by the definition of $(\ldots, S_i, \ldots)$:

$$(\ldots, S_i, \ldots) = (\ldots, X_i, \ldots) = (\ldots, S_i, \ldots) \ ,$$

Which ends the lemma. □

In particular Lemma 1 can be applied to the language equations emerging from conjunctive grammars.

3 Main result—non-regular conjunctive language over unary alphabet

Since we deal with an unary alphabet we identify word a^n with number n and work with sets of integers rather than with sets of words. The allowed operations in these equations are (set-theoretical) sum, intersection and 'addition', that is the analogon of concatenation. We write it in concatenation style for simplicity:

$$XY := \{x + y : x \in X, y \in Y\} \ .$$

Sometimes it is more convenient to use system of resolved language equations instead of grammars.

Still we use words 'grammar' and 'language', since this is the main interest of this paper.

Let us define the following sets of integers:

$$A_1 = \{1 \cdot 4^n : n \in \mathbb{N}\} ,$$
$$A_2 = \{2 \cdot 4^n : n \in \mathbb{N}\} ,$$
$$A_3 = \{3 \cdot 4^n : n \in \mathbb{N}\} ,$$
$$A_{12} = \{6 \cdot 4^n : n \in \mathbb{N}\} .$$

The indices reflect the fact that these sets written in tetrary system begin with digits $1, 2, 3, 12$, respectively and have only 0'es afterwards. We will show that those sets are the minimal solution of the equations:

$$B_1 = (B_2B_2 \cap B_1B_3) \cup \{1\} , \tag{1}$$
$$B_2 = (B_{12}B_2 \cap B_1B_1) \cup \{2\} , \tag{2}$$
$$B_3 = (B_{12}B_{12} \cap B_1B_2) \cup \{3\} , \tag{3}$$
$$B_{12} = (B_3B_3 \cap B_1B_2) . \tag{4}$$

This set of language equations can be easily transformed to a conjunctive grammar over unary alphabet (we should specify the starting symbol, say B_1). None of the sets A_1, A_3, A_3, A_{12} is a regular language over unary alphabet.

Since we often prove theorems by induction on number of digits, it is convenient to use the following notation for language (set) S:

$$S \restriction_n := \{s \in S : s \text{ has at most } n \text{ digits}\} .$$

We shall use it also for the vectors of languages (sets) with an obvious meaning.

Lemma 2. *Every solution* (S_1, S_2, S_3, S_{12}) *of equations* (1)–(4) *satisfies:*

$$(A_1, A_2, A_3, A_{12}) \subset (S_1, S_2, S_3, S_{12}).$$

Proof. We shall prove by induction on n, that

$$(A_1, A_2, A_3, A_{12}) \restriction_n \subset (S_1, S_2, S_3, S_{12}) .$$

For $n = 1$ we know that $i \in S_i$ by (1), (2) and (3). This ends induction basis.

For induction step let us assume that

$$(A_1, A_2, A_3, A_{12}) \restriction_{n+1} \subset (S_1, S_2, S_3, S_{12}).$$

We shall prove that this is true also for $(n + 2)$.

Let us start with 4^{n+1} and S_1. By induction assumption $2 \cdot 4^n \in S_2$ and hence

$$(2 \cdot 4^n) + (2 \cdot 4^n) = 4^{n+1} \in S_2S_2 .$$

Also by induction assumption $4^n \in S_1$ and $3 \cdot 4^n \in S_3$, hence

$$(4^n) + (3 \cdot 4^n) = 4^{n+1} \in S_1S_3 ,$$

and so $4^{n+1} \in S_2S_2 \cap S_1S_3$ and by (1) we conclude that $4^{n+1} \in S_1$.

For $6 \cdot 4^n$, which is a $(n + 2)$-digit number, we can see that $3 \cdot 4^n \in S_3$, $2 \cdot 4^n \in S_2$ by induction hypothesis and $1 \cdot 4^{n+1} = 4 \cdot 4^n \in S_1$, which was proved already in induction step. Hence $6 \cdot 4^n \in S_3S_3 \cap S_1S_2$ and by (4) we get $6 \cdot 4^n \in S_{12}$.

For $2 \cdot 4^{n+1}$ note that $2 \cdot 4^n \in S_2$, $6 \cdot 4^n \in S_{12}$ and $1 \cdot 4^{n+1} \in S_1$ hence $2 \cdot 4^{n+1} \in S_1S_1 \cap S_{12}S_2$ and by (2) $2 \cdot 4^{n+1} \in S_2$.

For $3 \cdot 4^{n+1}$ notice that $2 \cdot 4^{n+1} \in S_2$, $6 \cdot 4^n \in S_{12}$ and $1 \cdot 4^{n+1} \in S_1$ hence $3 \cdot 4^{n+1} \in S_{12}S_{12} \cap S_1S_2$ and by (3) $3 \cdot 4^{n+1} \in S_3$.

This ends induction step. □

Lemma 3. *Sets* (A_1, A_2, A_3, A_{12}) *satisfy*

$$\phi(A_1, A_2, A_3, A_{12}) \subset (A_1, A_2, A_3, A_{12}) .$$

Proof. Putting this in explicit way, we have to prove that

$$A_1 \supset (A_2A_2 \cap A_1A_3) \cup \{1\} , \quad (5)$$
$$A_2 \supset (A_{12}A_2 \cap A_1A_1) \cup \{2\} , \quad (6)$$
$$A_3 \supset (A_{12}A_{12} \cap A_1A_2) \cup \{3\} , \quad (7)$$
$$A_{12} \supset (A_3A_3 \cap A_1A_2) . \quad (8)$$

Consider first (5). Let m belong to the right-hand side of (5). If $m = 1$ then the thesis is obvious. So consider

$$m \in A_2A_2 \cap A_1A_3 .$$

By definition there are numbers k, l such that $k, l \in A_2$ and $m = k + l$. Then either $k = l$ and $m \in A_1$ or $k \neq l$ and so m has exactly two non-zero digits, both being 2's. On the other hand $m \in A_1A_3$, so by definition there are $k' \in A_1$, $l' \in A_3$ such that $m = l' + k'$. And so either $m = 4k'$, if $l' = 3k'$ and m, or m has only two non-zero digits: 1 and 3. But this is a contradiction with a claim that m has only two non-zero digits, both being $2's$.

We deal with other cases in the same manner.

Consider now (6). Let m belong to the right-hand side of (6). If $m = 2$ then the thesis is obvious. So consider

$$m \in A_{12}A_2 \cap A_1A_1 .$$

Then $m \in A_1A_1$. By definition there are numbers k, l such that $k, l \in A_1$ and $m = k + l$. The sum of digits in k and l is 2. On the other hand $m \in A_{12}A_2$, so by definition there are $k' \in A_{12}$, $l' \in A_2$ such that $m = l' + k'$. The sum of digits in l' and k' is 5. Hence, since the equation $k + l = k' + l'$ is true, there must be a carrying of digits in $k' + l'$. And this is only possible when $k' = 3l'$ and so $k' + l' \in A_2$.

Consider now (7). Let m belong to the right-hand side of (7). If $m = 3$ then the thesis is obvious. So consider

$$m \in A_{12}A_{12} \cap A_1A_2 .$$

Then $m \in A_{12}A_{12}$. By definition there are numbers k, l such that $k, l \in A_{12}$ and $m = k + l$. The sum of digits of k and l is 6. On the other hand $m \in A_1A_2$, so by definition there are $k' \in A_1$, $l' \in A_2$ such that $m = l' + k'$. The sum of digits of k' and l' is 3 and so since $k + l = k' + l'$ there is a carrying of digits in $k + l$, but this is possible only when $k = l$ and clearly $k + l \in A_3$.

Consider now (8). Let m belong to the right-hand side of (8), that is

$$m \in A_3A_3 \cap A_1A_2 .$$

Then $m \in A_3A_3$. By definition there are numbers k, l such that $k, l \in A_3$ and $m = k + l$. The sum of digits of k and l is 6. On the other hand $m \in A_1A_2$, so by definition there are $k' \in A_1$, $l' \in A_2$ such that $m = l' + k'$. The sum of digits of k' and l' is 3 and so since $k + l = k' + l'$ there is a carrying of digits in $k + l$, but this is possible only when $k = l$ and clearly $k + l \in A_{12}$. □

Theorem 1. *Sets* A_1, A_2, A_3, A_{12} *are the least solution of* (1)–(4).

Proof. By Lemma 1 it is enough to show that (A_1, A_2, A_3, A_{12}) are included in every solution and that $\phi(A_1, A_2, A_3, A_{12}) \subset (A_1, A_2, A_3, A_{12})$. The former was shown in Lemma 2 and the latter in Lemma 3. □

Corollary 1. *The non-regular language* $\{a^{4^n} : n \in \mathbb{N}\}$ *can be generated by conjunctive grammar over unary alphabet.*

Corollary 2. *Conjunctive grammars over unary alphabet have more expressive power then context-free grammars.*

4 Additional results

4.1 Number of nonterminals required

The grammar described in the previous section uses four nonterminals. It can be easily converted to Chomsky normal form—we need to introduce two new nonterminals for languages $\{1\}$ and $\{2\}$, respectively. Hence grammar for language $\{4^n : n \in \mathbb{N}\}$ in Chomsky normal form requires at the most six nonterminals. It is an interesting question, which mechanisms of conjunctive grammars and how many of them are required to generate a non-regular language? How many nonterminals are required? How many of them must generate non-regular languages? How many intersections are needed? Putting this question in the other direction, are there any natural sufficient conditions for a conjunctive grammar to generate regular language?

It should be noted that we are able to reduce the number of nonterminals to three, but we sacrifice Chomsky normal form and introduce also concatenations of three nonterminals in productions. This can be seen as some trade-off between number of nonterminals and length of concatenations. Consider language equations:

$$B_1 = (B_{2,12}B_{2,12} \cap B_1B_3) \cup \{1\} \ , \tag{9}$$

$$B_{2,12} = \Big((B_{2,12}B_{2,12} \cap B_1B_1) \cup \{2\}\Big) \cup \Big((B_3B_3 \cap B_{2,12}B_{2,12})\Big) \ , \tag{10}$$

$$B_3 = (B_{2,12}B_{2,12} \cap B_1B_1B_1) \cup \{3\} \ . \tag{11}$$

These are basically the same equations as (1)–(4), except that

- nonterminals B_2 and B_{12} are identified (or merged),
- conjunct B_2B_1 in (3) was changed to $B_1B_1B_1$.

Theorem 2. *The least solution of* (9)–(11) *is*

$$(A_1,\ A_2 \cup A_{12},\ A_3) \ .$$

Proof. The proof of this theorem is a slight technical modification of the proof of Theorem 1.

The main idea of the proof is to think of nonterminal $B_{2,12}$ that corresponds to the set $A_2 \cup A_{12}$ as two nonterminals: B_2 and B_{12}, corresponding to sets A_2 and A_{12}, respectively.

Firstly it is easy to check that replacing B_2B_1 with $B_1B_1B_1$ in (3) requires only small changes of proofs of Lemma 2 and Lemma 3.

Let $(S_1, S_{2,12}, S_3)$ be the least solution of the equations. We want to show, that

$$(A_1, A_2 \cup A_{12}, A_3) \subset (S_1, S_{2,12}, S_3) \ ,$$

in analogy to Lemma 2. It is enough to show, that $S_{2,12}$ is a superset of both A_2 and A_{12}. But this is obvious—in the equations we have replaced every occurrence of B_1 and B_{12} by $B_{2,12}$. This can be thought as adding new production $B_2 \to B_{12}$ and $B_{12} \to B_2$. Clearly adding the new productions does not decrease the solution.

The more interesting part is showing, that

$$\phi(A_1, A_2 \cup A_{12}, A_3) \subset (A_1, A_2 \cup A_{12}, A_3) \ .$$

Or explicitly:

$$\begin{aligned} A_1 &\supset \Big((A_2 \cup A_{12})(A_2 \cup A_{12}) \cap A_1A_3\Big) \cup \{1\} \ , \\ A_2 \cup A_{12} &\supset \Big(\big((A_2 \cup A_{12})(A_2 \cup A_{12}) \cap A_1A_1\big) \cup \{2\}\Big) \cup \Big(\big(A_3A_3 \cap (A_2 \cup A_{12})(A_2 \cup A_{12})\big)\Big) \ , \\ A_3 &\supset \Big((A_2 \cup A_{12})(A_2 \cup A_{12}) \cap A_1A_1A_1\Big) \cup \{3\} \ . \end{aligned}$$

We will show even stronger result, that is

$$A_1 \supset \Big((A_2 \cup A_{12})(A_2 \cup A_{12}) \cap A_1A_3\Big) \cup \{1\} \ , \tag{12}$$

$$A_2 \supset \Big((A_2 \cup A_{12})(A_2 \cup A_{12}) \cap A_1A_1\Big) \cup \{2\} \ , \tag{13}$$

$$A_{12} \supset A_3A_3 \cap (A_2 \cup A_{12})(A_2 \cup A_{12}) \ , \tag{14}$$

$$A_3 \supset \Big((A_2 \cup A_{12})(A_2 \cup A_{12}) \cap A_1A_1A_1\Big) \cup \{3\} \ . \tag{15}$$

These equations are similar to equations from Lemma 3, apart that on the right-hand side each A_2 and A_{12} was replaced by $A_2 \cup A_{12}$. We show, that each $A_2 \cup A_{12}$ can be replaced by exactly one A_2 or A_12 and keep the value of the right-hand side constant.

We use the fact, that if $n + m = n' + m'$ then the sum of tetrary digits of n and m is equal modulo 3 the sum of tetrary digits of n' and m'.

Consider (12). The sum of digits from A_1A_3 is 4, the sum of digits from $(A_2 \cup A_{12})(A_2 \cup A_{12})$ can be $4, 5, 6$, with 4 only for two choices of A_2. Hence we may remove the A_{12} from the equations.

Consider (13). The sum of digits from A_1A_1 is 2, the sum of digits from $(A_2 \cup A_{12})(A_2 \cup A_{12})$ can be $4, 5, 6$, with 5 only for choices of A_2 and A_{12}. Hence we may replace $(A_2 \cup A_{12})(A_2 \cup A_{12}$ by A_2A_{12}, as desired.

Consider (14). The sum of digits from A_3A_3 is 6, the sum of digits from $(A_2 \cup A_{12})(A_2 \cup A_{12})$ can be $4, 5, 6$, with 6 only for two choices of A_{12}. Hence we may remove the A_2 from the equations.

Consider (15). The sum of digits from $A_1A_1A_1$ is 3, the sum of digits from $(A_2 \cup A_{12})(A_2 \cup A_{12})$ can be $4, 5, 6$, with 6 only for two choices of A_{12}. Hence we may remove the A_2 from the equations.

Hence the (12)–(15) follow from the Lemma 3. Now by Lemma 1 the theorem follows. □

4.2 Related languages

Theorem 3. *For every natural number k, and every $i \in \{1, \dots, k-1\}$ there is a conjunctive grammar over unary alphabet generating language*

$$\{i \cdot k^n : n \in \mathbb{N}\} \ ,$$

for every $i, j \in \{1, \dots, k-1\}$ there is a conjunctive grammar over unary alphabet generating language

$$\{(ki + j) \cdot k^n : n \in \mathbb{N}\} \ .$$

Proof. For every $k > 5$ we introduce non-terminals $B_{i,j}$, where $i = 1, \dots k-1$ and $j = 0, \dots, k-1$, with intention that $B_{i,j}$ defines language of numbers beginning with digits i, j and then only zeroes in k-ary system of numbers. Then we define the productions as:

$B_{1,j} \to$	$B_{k-1,0}B_{j+1,0}$	$\& \ B_{k-2,0}B_{j+2,0}$		for $j = 0, 1, 2$	(16)
$B_{i,j} \to$	$B_{i-1,k-1}B_{j+1,0}$	$\& \ B_{i-1,k-2}B_{j+2,0}$		for $j = 0, 1, 2$ $i > 1$	(17)
$B_{i,j} \to$	$B_{i,j-1}B_{1,0}$	$\& \ B_{i,j-2}B_{2,0}$	$\& \ B_{i,0}B_{j,0}$	for $j > 2$	(18)
$B_{i,0} \to$	i				(19)

We show, using methods as in Lemma 2 and Lemma 3, that the least solution of these equations is

$$L_{i,j} = \{(k \cdot i + j) \cdot k^n : n \in \mathbb{N}\} \quad \text{for } j \neq 0 \ ,$$
$$L_{i,0} = \{i \cdot k^n : n \in \mathbb{N}\} \ .$$

For $k = 2, 3, 4, 5$ we have to sum up some languages generated in cases of $k = 8, 9, 25$, respectively. The case of $k = 1$ is trivial.

The proof is analogous to the proof of Theorem 1 and it follows by Lemma 4 and Lemma 5 below, which are proved in the same way as Lemma 2 and Lemma 3. The Reader not interested in the technical details may skip them. □

Lemma 4. *Every solution* $(\ldots, S_{i,j}, \ldots)$ *of equations* (16)–(19) *satisfies*

$$(\ldots, S_{i,j}, \ldots) \supset (\ldots, L_{i,j}, \ldots) \ .$$

Proof. We proceed on induction on n, proving that

$$(\ldots, S_{i,j}, \ldots) \supset (\ldots, L_{i,j}, \ldots) \restriction_n \ .$$

For fixed n we consider words in order induced by their meaning as numbers, that is we consider w before w' if number denoted by w is smaller than number denoted by w'.

For $n = 1$ the thesis is clear by (19)—every solution must include the one-digit elements from $(\ldots, L_{i,0}, \ldots)$.

For $n \geq 2$ consider word $w = ij0^m$. If $i = 1$ and $j < 3$ then we use (16). By induction assumption $(k-1)0^m \in S_{k-1,0}$ and $(j+1)0^m \in S_{j+1,0}$. Adding

$$(k-1)0^m + (j+1)0^m = 1j0^m \in S_{k-1,0}S_{j+1,0} \ .$$

By induction assumption $(k-2)0^m \in S_{k-2,0}$ and $(j+2)0^m \in S_{j+2,0}$. Adding

$$(k-2)0^m + (j+2)0^m = 1j0^m \in S_{k-2,0}S_{j+2,0} \ ,$$

hence $1j0^m \in S_{1,j}$, by (16).

The second case, for $i > 1$ and $j < 3$ is similar—we use (17). Word $(i-1)(k-1)0^m$ denotes a number smaller than w, hence it was already considered, therefore $(i-1)(k-1)0^m \in S_{i-1,k-1}$. By induction assumption $(j+1)0^m \in S_{j+1,0}$. Adding

$$(i-1)(k-1)0^m + (j+1)0^m = ij0^m \in S_{i-1,k-1}S_{j+1,0} \ .$$

Word $(i-1)(k-2)0^m$ denotes a number smaller than w, hence it was already considered, therefore $(i-1)(k-2)0^m \in S_{i-1,k-2}$. By induction assumption $(j+2)0^m \in S_{j+2,0}$. Adding

$$(i-1)(k-2)0^m + (j+2)0^m = ij0^m \in S_{i-1,k-2}S_{j+2,0} \ ,$$

hence $ij0^m \in S_{i,j}$, by (17).

The last case, for $j > 2$—we use (18). Word $i(j-1)0^m$ denotes number smaller than w, hence it was already considered, therefore $i(j-1)0^m \in S_{i,j-1}$. By induction assumption $10^m \in S_{1,0}$. Adding

$$i(j-1)0^m + 10^m = ij0^m \in S_{i,j-1}S_{1,0} \ .$$

Word $i(j-2)0^m$ denotes number smaller than w, hence it was already considered, therefore $i(j-2)0^m \in S_{i,j-2}$. By induction assumption $20^m \in S_{2,0}$. Adding

$$i(j-2)0^m + 20^m = ij0^m \in S_{i,j-2}S_{2,0} \ .$$

Word $i00^m$ denotes number smaller than w, hence it was already considered, therefore $i00^m \in S_{i,0}$. By induction assumption $j0^m \in S_{j,0}$. Adding

$$i00^m + j0^m = ij0^m \in S_{i,0}S_{j,0} \ ,$$

hence $ij0^m \in S_{i,j}$, by (18).

Lemma 5. *Languages* $(\ldots, L_{i,j}, \ldots)$ *satisfy*

$$\phi(\ldots, L_{i,j}, \ldots) \subset (\ldots, L_{i,j}, \ldots) \ .$$

Proof. Consider first (16). Let $n \in L_{k-1,0}$, $n' \in L_{j+1,0}$ and $m \in L_{k-2,0}$, $m' \in L_{j+2,0}$ such that

$$n + n' = m + m' \ .$$

If there is a carrying of 1 in one of the sums $n+n$ or $m+m'$ then the result belongs to the $L_{1,j}$ and we are done. So we consider the case, when then is no carrying in both sums. Hence $n+n'$ has two non-zero digits—1 and $(k-1)$. Sum $m+m'$ has two non-zero digits—2 and $(k-2)$. Since $1 < 2 < k-2 < k-1$ we conclude that $n+n' \neq m+m'$.

Consider now (17). Let $n \in L_{i-1,k-1}$, $n' \in L_{j+1,0}$ and $m \in L_{i-1,k-2}$, $m' \in L_{j+2,0}$ such that

$$n + n' = m + m' \ .$$

The sum of digits of n plus the sum of digits of n' is $i+j+k-1$, the same as for m, m'. And so either there is no carrying of digits in both sums or there is a carrying in both sums (it is easy to see that in both cases it is not possible to have two carryings).

If there is no carrying then the multiset of non-zero digits is $\{i-1, k-1, j+1\}$ for n, n' and $\{i-1, k-2, j+2\}$ for m, m'. They must equal and so $\{k-1, j+1\} = \{k-2, j+2\}$, which is not possible, since $j+1 < j+2 \leq k-2 < k-1$.

If a carrying of 1 in one of the sums $n+n$ or $m+m'$ in the 'desired position', that is in $n+n'$ digit $(j+1)$ adds up with $(k-1)$ or in $m+m'$ digit $(j+2)$ adds up with $(k-2)$ then the result is as desired. And so we have to deal with the case, when both carryings are in different position. But this is possible only when $(j+1)$ adds up with $(i-1)$ in $n+n'$ and $(j+2)$ adds up with $(i-1)$ in $m+m'$. But those results give different second digit.

Consider now (18). Let $n \in L_{i,j-1}$, $n' \in L_{1,0}$ and $m \in L_{i,j-2}$, $m' \in L_{2,0}$ and $p \in L_{i,0}$, $p' \in L_{j,0}$ such that

$$n + n' = m + m' = p + p' \ .$$

Again, the sum of digits is the same for each pair and so either there is no carrying in all of the sums or exactly one carrying (clearly $p+p'$ cannot have two carryings). Since digits $i, j-1, 1$ from $n+n'$ are all non-zero, then $n+n'$ has at least two non-zero digits. Since $p+p'$ has at the most two non-zero digits, then there are exactly two non-zero digits in the result.

Suppose now that there was a carrying and we ended up with two non-zero digits. Then $p+p'$ has 1 as its first digit and $(i+j-k)$ as the second an then only 0'es. On the other hand $n+n'$ has 1 as the first digit, 0 as the second and $j-1 > 0$ as the third, contradiction.

And so there is no carrying. If in at least one sum the adding is as desired, then we end up with a good result. If no adding is as desired, then we have the following multisets of digits: $\{i+1, j-1\}$ for $n+n'$, $\{i+2, j-2\}$ for $m+m'$ and $\{i, j\}$ for $p+p'$. Contradiction.

The case of (19) is obvious. □

5 Regular languages over k-ary alphabet

Now we deal with major generalisation of the Theorem 1 and Theorem 3. We deal with languages $\{a^n : n \in L\}$, where L is some regular language (written in some k-ary system). To simplify the notation, let $\Sigma_k = \{0, \ldots, k-1\}$. From the following on we consider regular languages over Σ_k for some k that do not have words with leading 0, since this is meaningless in case of numbers.

Definition 4. *Let $w \in \Sigma_k^*$ be a word. We define its unary representation as*

$$f_k(w) = \{a^n : w \textit{ read as } k\textit{-ary number is } n\} \ .$$

When this does not lead to confusion, we also use f_k applied to languages with an obvious meaning.

The following fact shows that it is enough to consider the k parameter in f_k that are 'large enough'.

Lemma 6. *For every $k = l^n$, $n > 0$ and every unary language L language $f_k^{-1}(L)$ is regular if and only if language $f_l^{-1}(L)$ is regular.*

Proof. An automaton over alphabet Σ_k can clearly simulate reading a word written in l-ary system and *vice-versa.* □

In the following we shall use 'big enough' k, say $k \geq 100$. We claim, that for regular L language $f_k(L)$ is conjunctive.

We now define the conjunctive grammar for fixed regular language $L \subset \Sigma_k^*$ without leading 0. Let

$$M = \langle \Sigma_k, Q, \delta, F, q_0 \rangle$$

be the (non-deterministic) automaton that recognizes L^r.

We define conjunctive grammar $G = \langle \{a\}, N, P, S \rangle$ over unary alphabet with:

$$N = \{A_{i,j,q}, A_{i,j} : 1 \leq i < k,\ 0 \leq j < k,\ q \in Q\} \cup \{S\} \ .$$

The intended solution is

$$L(A_{i,j}) = \{n \ :\ f_k^{-1}(n) = ij0^k \text{ for some natural } k\} \ , \quad (20)$$
$$L(A_{i,j,q}) = \{n : f_k^{-1}(n) = ijw,\ \delta(q_0, w^r, q)\} \ , \quad (21)$$
$$L(S) = f_k(L) \ . \quad (22)$$

We denote sets defined in (21) by $L_{i,j,q}$.

From Theorem 3 we know, that $A_{i,j}$ can be defined by conjunctive grammars, and so we focus only on productions for $A_{i,j,q}$. In those equations x, q, q' satisfy $\delta(q', x, q)$.

$$A_{i,j,q} \to \ \&_{n=0}^{3}\ A_{i,n}A_{j-n,x,q'} \quad \text{for } j > 3,\ i \neq 0, \text{ every } x, q' \ , \quad (23)$$
$$A_{i,j,q} \to \ \&_{n=1}^{4}\ A_{i-1,j+n}A_{k-n,x,q'} \quad \text{for } j < 4,\ i \neq 0, 1 \text{ and every } x, q' \ , \quad (24)$$
$$A_{1,j,q} \to \ \&_{n=1}^{4}\ A_{k-n,0}A_{j+n,x,q'} \quad \text{for } j < 4 \text{ and every } x, q' \ , \quad (25)$$
$$A_{i,j,q_0} \to \ k \cdot i + j \ , \quad (26)$$
$$S \to \ (L \cap \Sigma_k) \cup \bigcup_{\substack{i,j,q:\\ \delta(q,ji) \cap F \neq \emptyset}} A_{i,j,q} \ . \quad (27)$$

We shall prove, that $(\ldots, L_{i,j,q}, \ldots)$ are the least solution of those language equations. The case of $L(S)$ in (22) will then easily follow. We identify the equations with productions (23)–(25).

Lemma 7. *Every solution $(\ldots, X_{i,j,q}, \ldots)$ of language equations defining G for non-terminals $(\ldots, A_{i,j,q}, \ldots)$ satisfies*

$$(\ldots, L_{i,j,q}, \ldots) \subset (\ldots, X_{i,j,q_i}, \ldots) \ .$$

Proof. We prove by induction that for every $n > 1$

$$L_{i,j,q} \restriction_n \subset X_{i,j,q}.$$

When $|ijw| = 2$ then this is obvious by rule (26).

Suppose we have proven the Lemma for $k < n$, we prove it for n. Choose any $ijw \in L_{i,j,q} \restriction_n$. Let $w = xw'$. Let p be a state such that $\delta(q_0, w'^r, p)$ and $\delta(p, x, q)$ (choose one if there are many).

Suppose $j > 3$ and consider words

$$jxw' \in L_{j,x,p} \restriction_{n-1} \subset X_{j,x,p} \ ,$$
$$(j-1)xw' \in L_{j-1,x,p} \restriction_{n-1} \subset X_{j-1,x,p} \ ,$$
$$(j-2)xw' \in L_{j-2,x,p} \restriction_{n-1} \subset X_{j-2,x,p} \ ,$$
$$(j-3)xw' \in L_{j-3,x,p} \restriction_{n-1} \subset X_{j-3,x,p} \ .$$

All the '$\subset$' come from the induction hypothesis. Adding $i00^{|w'|+1}$, $i10^{|w'|+1}$, $i20^{|w'|+1}$, $i30^{|w'|+1}$, respectively, gives ijw in all cases, and so by (23) $ijw \in X_{i,j,q}$.

Other cases (which use productions (24), (25)) are proved analogously.

Suppose $j < 4$ and $i > 1$, consider words

$$\begin{aligned}(k-1)xw' &\in L_{k-1,x,p} \restriction_{n-1} \subset X_{k-1,x,p} \ ,\\ (k-2)xw' &\in L_{k-2,x,p} \restriction_{n-1} \subset X_{k-2,x,p} \ ,\\ (k-3)xw' &\in L_{k-3,x,p} \restriction_{n-1} \subset X_{k-3,x,p} \ ,\\ (k-4)xw' &\in L_{k-4,x,p} \restriction_{n-1} \subset X_{k-4,x,p} \ .\end{aligned}$$

All the '$\subset$' come from the induction hypothesis. Adding $(i-1)(j+1)0^{|w'|+1}$, $(i-1)(j+2)0^{|w'|+1}$, $(i-1)(j+3)0^{|w'|+1}$, $(i-1)(j+4)0^{|w'|+1}$, respectively, gives ijw in all cases, and so by (24) $ijw \in X_{i,j,q}$.

Suppose $j < 4$ and $i = 1$. Consider words

$$\begin{aligned}(j+1)xw' &\in L_{j+1,x,p} \restriction_{n-1} \subset X_{j+1,x,p} \ ,\\ (j+2)xw' &\in L_{j+2,x,p} \restriction_{n-1} \subset X_{j+2,x,p} \ ,\\ (j+3)xw' &\in L_{j+3,x,p} \restriction_{n-1} \subset X_{j+3,x,p} \ ,\\ (j+4)xw' &\in L_{j+4,x,p} \restriction_{n-1} \subset X_{j+4,x,p} \ .\end{aligned}$$

All the '$\subset$' come from the induction hypothesis. Adding $(k-1)0^{|w'|+1}$, $(k-2)0^{|w'|+1}$, $(k-3)0^{|w'|+1}$, $(k-4)0^{|w'|+1}$, respectively, gives $1jw$ in all cases, and so by (25) $1jw \in X_{1,j,q}$. □

Lemma 8. *Languages $L_{i,j,q}$ satisfy*

$$\phi(\ldots, L_{i,j,q}, \ldots) \subset (\ldots, L_{i,j,q}, \ldots) \ .$$

Proof. We proceed by induction on number of digits in w. We first proof, that it in fact has the desired two first digits. Then we shall deal with the q index.

We begin with (23). Suppose w belong to the right-hand side. We shall show that it also belongs to the left-hand side. Consider the possible positions of the two first digits of each conjunct. Notice, that if j is on the position one to the right of i, then the digits are as desired. And so we may exclude this case from our consideration. The Table 1 summarizes the results.

	i and j are together	j is leading	i is leading
$A_{i,0}A_{j,xq'}$	$(i+j), x$	$j, x\langle +i\rangle$	$i, 0$
$A_{i,1}A_{j-1,xq'}$	$(i+j-1), (x+1)$	$(j-1), x\langle +i\rangle$	$i, 1$
$A_{i,2}A_{j-2,xq'}$	$(i+j-2), (x+2)$	$(j-2), x\langle +i\rangle$	$i, 2$
$A_{i,3}A_{j-3,xq'}$	$(i+j-3), (x+3)$	$(j-3), x\langle +i\rangle$	$i, 3$

Table 1. The possible leading digits of numbers resulting from adding in (23)

The Table 1 has some drawbacks:

– some digits sum up to k (or more) and influence another digit (by carrying 1),
– in the second column i may be or may be not on the same position as x, but we deal with those two cases together,
– in the second column there may be an add up to k somewhere to the right, and hence we can add 1 to x.

This possibilities in the second point were marked in the table by writing $\langle +i\rangle$.

If we want the intersection to be non-empty we have to choose four items from the Table 1, no two of them in the same row. We show, that this is not possible. We say that some choices *fit*, if the digits included in the table are the same in those choices.

First of all, no two elements in the third column fit. They have fixed digits and they clearly are different.

Suppose now that we choose two elements from the first column. We show that if in one of them $(i + j - z)$ sums up to k (or more) then the same thing happens in the second choice. If $i+j-z \geq k$ (perhaps by additional 1 carried from the previous position) then the first digit is 1. In the second element the first digit can be 1 (if there is a carrying of 1) or at least $i+j-z'$, but the latter is not possible, since $i + j - z' > 1$. In both cases it is not possible to fit the digits on the column with x.

It is not possible to choose three elements from the second column. For the sake of contradiction assume that we have three fitting choices in this column. As previously we may argue, that either in all of them in the leading digit (that is with $j - z$ for some z) there is an adding to k and carrying one to the (newly created) position or in all of them there is no adding to k in the leading position.

Suppose they add up. Since $j < k$ then this is possible only for the first row. In particular it is not possible to have two choices when there is an add-up. Suppose they do not add up. Then if there are three fitting choices, then in one of them we must increase the value of the second digit by at least 2. But the maximal value carried from the previous position is 1. Contradiction. As a consequence if there are two fitting choices then the first digit is in range $(j - 3, j)$.

And so we know, that if there are four fitting choices, then exactly one of them is in the first column, one in the third column and two in the middle column. The third column always begins with i. In the first column the leading digit is at least $i + j - 3 > i$ or it is 1. Hence $i = 1$. And so the choices in the second column begin with 1 as well. Hence $j < 4$, which is a contradiction.

We now move to (24). The Table 2 summarizes the possible first two digits:

	i and $k-z$ are together	$k-z$ is leading	$i-1$ is leading
$A_{i-1,j+1}A_{k-1,x,q'}$	$(k+i-2),(1+j+x)$	$(k-1),x\langle +i\rangle$	$i-1,j+1$
$A_{i-1,j+2}A_{k-2,x,q'}$	$(k+i-3),(2+j+x)$	$(k-2),x\langle +i\rangle$	$i-1,j+2$
$A_{i-1,j+3}A_{k-3,x,q'}$	$(k+i-4),(3+j+x)$	$(k-3),x\langle +i\rangle$	$i-1,j+3$
$A_{i-1,j+4}A_{k-4,x,q'}$	$(k+i-5),(4+j+x)$	$(k-4),x\langle +i\rangle$	$i-1,j+4$

Table 2. The possible leading digits of numbers resulting from adding in (24)

As before we may argue, that if there are some fitting entries in some column then on their leading position digits sum up to k in all choices or in all choices do not sum up to k.

We cannot have two choices from the third column (the second digits do not match). We can have at the most two from the second column (to obtain three we would have to carry at least 2 to the first digit of one of them and this is not possible).

For the same reason there can be at the most two choices from the first column. But if there are two choices from the first column then we cannot match the positions with x. Hence there is at the most one choice from the first column.

And so we have one choice from the first column, one from the third and two from the second. Since the third and the second column match, then i is a big digit, at least $k-3$. But in such a case in the first column we have at least $k + k - 3 - 5 > k$ and so the leading digit is 1. Contradiction.

Consider the last possibility, the (25).The Table 3 summarizes the possible first two digits:

	$k-z$ is leading	$j+z$ is leading
$A_{k-1,0}A_{j+1,x,q'}$	$(k-1),0\langle +j+1\rangle$	$(j+1),x\langle +k-1\rangle$
$A_{k-2,0}A_{j+2,x,q'}$	$(k-2),0\langle +j+2\rangle$	$(j+2),x\langle +k-2\rangle$
$A_{k-3,0}A_{j+3,x,q'}$	$(k-3),0\langle +j+3\rangle$	$(j+3),x\langle +k-3\rangle$
$A_{k-4,0}A_{j+4,x,q'}$	$(k-4),0\langle +j+4\rangle$	$(j+4),x\langle +k-4\rangle$

Table 3. The possible leading digits of numbers resulting from adding in (25)

In the first column there are no two fitting choices—since in the second digit in this column is at the most $j+4<k$ and there is no carrying to the first digit. And clearly the first digits are different. So we have to choose at least one element from the second column. It leading digit is at the most $j+4+2$—carrying of more than 2 from the previous position is not possible. But $j+6<k-4$ and hence the first digit do not match. Contradiction.

We now take the indices denoting states of the automaton into our consideration. Consider production:

$$A_{i,j,q} \to A_{i,0}A_{j,x,q'} \ \& \ A_{i,1}A_{j-1,x,q'} \ \& \ A_{i,2}A_{j-2,x,q'} \& \ A_{i,3}A_{j-3,x,q'}$$

and some w belonging to the right-hand side. We have already proved that $w = ijw'$. Consider conjunct $A_{i,0}A_{j,x,q'}$. Consider $jxw'' \in L_{j,x,q'}$ that was used in derivation of w. Note that $w' = xw''$. By definition of $L_{j,x,q'}$ we obtain $\delta(q_0, w''^r, q')$ and by definition of the (23) we obtain $\delta(q', x, q)$, and so $\delta(q_0, w''^r x, q)$. But $w = ijxw''$, therefore it belongs to the left-hand side. The cases of other productions is proved in the same way.

Consider production:

$$A_{i,j,q} \to A_{i-1,j+1}A_{k-1,x,q'} \ \& \ A_{i-1,j+2}A_{k-2,x,q'} \ \&\& \ A_{i-1,j+3}A_{k-3,x,q'} \& \ A_{i-1,j+4}A_{k-4,x,q'}$$

and some w belonging to the right-hand side. We have already proved that $w = ijw'$. Consider conjunct $A_{i-1,j+1}A_{k-1,x,q'}$. Consider $(k-1)xw'' \in L_{k-1,x,q'}$ that was used in derivation of w. Note that $w' = xw''$. By definition of $L_{k-1,x,q'}$ we obtain $\delta(q_0, w''^r, q')$ and by definition of the (24) we obtain $\delta(q', x, q)$, and so $\delta(q_0, w''^r x, q)$. But $w = ijxw''$, therefore it belongs to the left-hand side.

Consider production:

$$A_{1,j,q} \to A_{k-1,0}A_{j+1,x,q'} \ \& \ A_{k-2,0}A_{j+2,x,q'} \ \&\& \ A_{k-3,0}A_{j+3,x,q'} \ \& \ A_{k-4,0}A_{j+4,x,q'}$$

and some w belonging to the right-hand side. We have already proved that $w = 1jw'$. Consider conjunct $A_{k-1,0}A_{j+1,x,q'}$. Consider $(j+1)xw'' \in L_{j+1,x,q'}$ that was used in derivation of w. Note that $w' = xw''$. By definition of $L_{j+1,x,q'}$ we obtain $\delta(q_0, w''^r, q')$ and by definition of the (25) we obtain $\delta(q', x, q)$, and so $\delta(q_0, w''^r x, q)$. But $w = 1jxw''$, therefore it belongs to the left-hand side. □

Theorem 4. *For every natural $k > 1$ and every regular $L \subset \Sigma_k^*$ without words with leading 0 language $f_k(L)$ is a conjunctive unary language.*

Proof. By Lemma 6 it is enough to consider 'big' k, say $k > 100$.

Lemma 7 and Lemma 8 assume that $k > 100$ and they guarantee that the sets $(\ldots, L_{i,j,q_i}, \ldots)$ fulfil the assumptions of Lemma 1. Now the only thing left is to see that (27) defines S properly. If $w \in L$ then either $|w| = 1$ and hence $w \in S \cap \Sigma_k$ or $|w| \geq 2$ and hence $w = ijw'$. Let q be such that

$$\delta(q_0, w'^r, q) \quad \text{and} \quad \delta(q, ji) \in F \ .$$

Clearly such state exists, since $w \in L$ and so automata recognizing l^r has some intermediate state q. but this means that $w \in A_{i,j,q}$ and $S \to A_{i,j,q}$. □

6 Conclusions and open problems

The main result of this paper is an example of a conjunctive grammar over unary alphabet generating non-regular language. This grammar has six nonterminal symbols in Chomsky normal form. Number of nonterminals could be reduced to three if we consider a grammar that is not in a Chomsky normal form. It remains an open question, how many nonterminals, intersection *etc.* are required to generate a non-regular language. In particular, can we give natural sufficient conditions for a conjunctive grammar to generate a regular language? Also, no non-trivial algorithm for recognizing conjunctive languages over unary alphabet is known. An obvious modification of the CYK algorithm requires quadratic time and linear space. Can those bounds be lowered? Closure under complementation of conjunctive languages (both in general and and in case of unary alphabet) remains unknown, with conjectured negative answer.

The second important result is a generalisation of the previous one: for every regular language $R \subset \{0, ..., k-1\}^*$ treated as set of k-ary numbers language $\{a^n : \exists_w \in R\ w$ read as a number is $n\}$ is a conjunctive unary language.

Acknowledgments. The author would like to thank Tomasz Jurdziński and Krzysztof Loryś for introducing to the subject and helpful discussion, Sebastian Bala and Marcin Bieńkowski for helpful discussion and Alexander Okhotin, who suggested the study of generalisation to unary representations of all regular languages (Theorem 4).

References

1. Okhotin, A. Conjunctive grammars. *Journal of Automata, Languages and Combinatorics*, 6:4:519–535, 2001.
2. Okhotin, A. Conjunctive grammars and systems of language equations. *Programming and Computer Software.*, 28:243–249, 2002.
3. Okhotin, A. An overview of conjunctive grammars. *Formal Language Theory Column. Bulletin of the EATCS*, 79:145–163, 2003.
4. Okhotin, A. Boolean grammars. *Information and Computation*, 194:1:19–48, 2004.
5. Okhotin, A. Nine open problems on conjunctive and boolean grammars. *TUCS Technical Report*, 794, 2006.