Theoretical Computer Science 665 (2017) 13–39

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Unambiguous conjunctive grammars over a one-symbol alphabet ☆

Artur Jeż [a,*], Alexander Okhotin [b]

[a] Institute of Computer Science, University of Wrocław, ul. Joliot-Curie 15, 50-383 Wrocław, Poland
[b] St. Petersburg State University, 14th Line V.O., 29B, Saint Petersburg 199178, Russia

A R T I C L E I N F O

Article history:
Received 10 September 2015
Received in revised form 2 December 2016
Accepted 6 December 2016
Available online 2 January 2017
Communicated by M. Crochemore

Keywords:
Conjunctive grammars
Ambiguity
Language equations
Undecidability
Unary languages

A B S T R A C T

It is demonstrated that unambiguous conjunctive grammars over a unary alphabet $\Sigma = \{a\}$ have non-trivial expressive power, and that their basic properties are undecidable. The key result is that for every base of positional notation, $k \geqslant 11$, and for every one-way real-time cellular automaton operating over the alphabet of base-k digits between $\lfloor \frac{k+9}{4} \rfloor$ and $\lfloor \frac{k+1}{2} \rfloor$, the language of all strings a^n with the base-k representation of the form $1w1$, where w is accepted by the automaton, is described by an unambiguous conjunctive grammar. Another encoding is used to simulate a cellular automaton in a unary language containing almost all strings. These constructions are used to show that for every fixed unambiguous conjunctive language L_0, testing whether a given unambiguous conjunctive grammar generates L_0 is undecidable.

1. Introduction

In ordinary formal grammars, called "context-free" in the literature, the available operations are concatenation and disjunction: each rule defines a concatenation, whereas disjunction is implicit in having multiple rules for a single symbol. *Conjunctive grammars* [21] are an extension of ordinary grammars, which further allows a conjunction operation in any rules. Conjunctive grammars are more of a variant of the definition of grammars than something entirely new, as they maintain the main principle behind the context-free grammars—that of defining the structure of shorter strings by concatenating longer strings with already defined properties—and only extend the set of logical connectives used to combine syntactical conditions. In spite of the increased expressive power of conjunctive grammars, they inherit the subcubic upper bound on the parsing complexity [28], as well as other parsing techniques originally developed for ordinary grammars, such as the *generalized LR* and, recently, the *deterministic LR* [1]. These results make conjunctive grammars suitable for practical use. The work on conjunctive grammars has also led to more general grammar models further equipped with negation [23, 16] or with context operators [4].

Conjunctive grammars over a one-symbol alphabet $\Sigma = \{a\}$ were proved non-trivial by Jeż [10], who constructed a grammar for the language $\{a^{4^n} \mid n \geqslant 0\}$ and extended this construction to describe every so-called *automatic set* [2], that is, a unary language with a regular base-k representation. Subsequent work on such grammars revealed their relatively high expressive power and a number of undecidable properties [11]. Testing whether a given string a^n is generated by a grammar

☆ A preliminary version of this paper was presented at the conference on Developments in Language Theory (DLT 2013, Paris, France, 18–21 June 2013), and its extended abstract appeared in the conference proceedings.

* Corresponding author.
E-mail addresses: aje@cs.uni.wroc.pl (A. Jeż), alexander.okhotin@utu.fi (A. Okhotin).

http://dx.doi.org/10.1016/j.tcs.2016.12.009
0304-3975/

G can be done in time $|G| \cdot n(\log n)^3 \cdot 2^{O(\log^* n)}$ [29], and if n is given in binary notation, this problem is EXPTIME-complete already for a fixed grammar G [12]. Conjunctive grammars over a unary alphabet remain non-trivial even in the special case of grammars with one nonterminal symbol [13]. These results had impact on the study of language equations [17,26], being crucial to understanding their computational completeness over a unary alphabet [15,18]. They are also related to the complexity results for circuits over sets of numbers [20], as conjunctive grammars over a unary alphabet may be regarded as a generalization of those circuits.

Unambiguous conjunctive grammars [24] are an important subclass of conjunctive grammars defined by analogy with ordinary unambiguous grammars, which represent the idea of assigning a unique syntactic structure to every well-formed sentence. Little is known about their properties, besides a parsing algorithm with running time $|G| \cdot O(n^2)$, where n is the length of the input [24]; for a unary alphabet, the running time can be improved to $|G| \cdot n(\log n)^2 \cdot 2^{O(\log^* n)}$ [29]. However, all the known results on the expressive power of conjunctive grammars over a unary alphabet [10–12,30] rely upon ambiguous grammars, and it is not even known whether unambiguous grammars can describe anything non-regular.

This paper sets off by presenting the first example of an unambiguous conjunctive grammar that describes a non-regular unary language. This is the same language $\{a^{4^n} \mid n \geqslant 0\}$, yet the grammar describing it, given in Section 3, must operate more carefully than the known ambiguous grammar. This example is then extended to a construction of unambiguous conjunctive grammars generating all languages of the form $L_{k,c} = \{a^{c \cdot k^n} \mid n \geqslant 0\}$, with $k \geqslant 2$ and $c \in \{k, k+1, \dots, k^2-1\}$; in other words, these are languages of all strings a^n with the base-k notation of n of the form $(ij0^*)_k$, where i and j are base-k digits. These languages serve as building blocks in the subsequent constructions.

Then the paper proceeds with reimplementing, using unambiguous grammars, the main general method for constructing conjunctive grammars over a unary alphabet. This method involves simulating a one-way real-time cellular automaton [6,9,22] over an input alphabet $\Sigma_k = \{0, 1, \dots, k-1\}$ of base-k digits, by a grammar generating all strings a^n with the base-k representation of n accepted by the cellular automaton. The known construction of such conjunctive grammars [11] always produces ambiguous concatenations. This paper defines a different simulation, under the assumption that the input alphabet of the cellular automaton is not the entire set of base-k digits, but rather some subset of size around $\frac{k}{4}$. With this restriction, the automaton can be simulated, so that all concatenations in the grammar remain unambiguous.

The simulation of a cellular automaton presented in Section 4 produces languages that grow exponentially fast, that is, the length of the n-th shortest string in the language is exponential in n. These languages have *density 0*, in the sense that the fraction of strings of length up to ℓ belonging to these languages tends to 0. As the concatenation of any two unary languages of non-zero density is always ambiguous, this limitation of the given construction might appear to be inherent to unambiguous conjunctive grammars. Nevertheless, a method for describing unary languages of non-zero density by an unambiguous conjunctive grammar is established in the next Section 5. The construction is based on the representation of a^* as an unambiguous concatenation of several languages of density 0, given by Enflo et al. [8]. In particular, some unary languages of density 1 can be described.

The results in Sections 4–5 are weaker than the original results on ambiguous conjunctive grammars over a one-symbol alphabet [11]. It remains unknown whether every unary language with its base-k representation recognized by a trellis automaton is therefore defined by an unambiguous conjunctive grammar. Nevertheless, the constructions in this paper turn out to be sufficient for proving uniform undecidability results for unambiguous conjunctive grammars, presented in Section 6. Consider that for ordinary grammars, testing for equality to the empty set is decidable, whereas equality to the set of all strings over a multi-symbol alphabet is undecidable; for the unambiguous case of ordinary grammars, testing equality to any regular language is decidable [31]. In contrast, for unambiguous conjunctive grammars, for every fixed language L_0 (over an arbitrary alphabet) described by some unambiguous conjunctive grammar, it is proved that testing whether a given unambiguous conjunctive grammar describes L_0 is undecidable.

In the last Section 7, it is shown that unambiguous conjunctive grammars can describe some sparse unary languages that grow arbitrarily fast—the same result as for the general case of conjunctive grammars [11].

2. Conjunctive grammars and ambiguity

In ordinary formal grammars (Chomsky's "context-free"), each rule defines all strings representable as a *concatenation* of substrings with the given properties. In conjunctive grammars, a rule is a *conjunction* of such concatenations, meaning all strings that can be represented as each of the listed concatenations at the same time.

Definition 1 *(Okhotin [21]).* A conjunctive grammar is a quadruple $G = (\Sigma, N, R, S)$, in which Σ and N are disjoint finite non-empty sets of terminal and nonterminal symbols respectively; R is a finite set of rules of the form

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

and $S \in N$ is a nonterminal designated as the initial symbol. A grammar is called *linear conjunctive*, if each α_i in each rule $(*)$ contains at most one nonterminal symbol.

Informally, a rule $(*)$ in a conjunctive grammar means that every string described by each *conjunct* α_i is therefore described by A. One way of formalizing this understanding is by *language equations*, in which the conjunction is interpreted as intersection of languages, as follows.