Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 369 (2006) 436–441
Theoretical Computer Science
www.elsevier.com/locate/tcs

Note

A note on ambiguity of internal contextual grammars☆

Lakshmanan Kuppusamy*

Department of Computer Engineering, National Institute of Technology Karnataka, Surathkal-575 025, India

Received 14 September 2005; received in revised form 28 July 2006; accepted 2 August 2006

Communicated by A.K. Salomaa

Abstract

In this paper, we continue the study of ambiguity of internal contextual grammars which was investigated in Ilie [On ambiguity in internal contextual languages, in: C. Martin-Vide (Ed.), Second Int. Conf. on Mathematical Linguistics, Tarragona, 1996, John Benjamins, Amsterdam, 1997, pp. 29–45] and Martin-Vide et al. [Attempting to define the ambiguity in internal contextual languages, in: C. Martin-Vide (Ed.), Second Int. Conf. on Mathematical Linguistics, Tarragona, 1996, John Benjamins, Amsterdam, 1997, pp. 59–81]. We solve some open problems formulated in these papers. The main results are: (i) there are inherently 1-ambiguous languages with respect to internal contextual grammars with arbitrary choice which are 0-unambiguous with respect to finite choice, (ii) there are inherently 2-ambiguous languages with respect to internal contextual grammars with arbitrary choice which are 1-unambiguous with respect to regular choice, and (iii) there are inherently 2-ambiguous languages with respect to depth-first internal contextual grammars with arbitrary choice which are 1-unambiguous with respect to finite choice.

Keywords: Internal contextual grammars; Inherent ambiguity; Depth-first contextual grammars

1. Introduction

Contextual grammars were introduced by Marcus in 1969 [3] as 'intrinsic grammars', without auxiliary symbols, based on the fundamental linguistic operation of inserting words in given phrases, according to certain contextual dependencies [2]. More precisely, contextual grammars produce languages starting from a finite set of *axioms* and adjoining *contexts*, iteratively, according to the *selector* present in the current sentential form. As introduced in [3], if adjoining the contexts is done at the ends of the strings, the grammar is called *external*. *Internal* contextual grammars were introduced by Păun and Nguyen in 1980 [8], where the contexts are adjoined to the selector strings appearing as substrings of the string.

Generally, ambiguity for a grammar is defined as follows: given a grammar, are there words in the generated language which have two distinct derivations? For Chomsky grammars, the notion of ambiguity is clear, but defining ambiguity for contextual grammars is not so obvious since the derivation of contextual grammars consists of many components such as axioms, contexts and selectors. In [6], the notion of ambiguity was considered for the first time in this field,

☆ A preliminary version of this paper was presented in Finite State Methods and Natural Language Processing 2005 (FSMNLP'05) held in Helsinki, Finland during September 1–2, 2005 and the paper has appeared in their CD-ROM pre-proceedings.

* Tel.: +91 824 2474 000x3445.
E-mail address: laksh@tifr.res.in

0304-3975/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2006.08.009

with the ambiguity defined for external contextual grammars. Then, in [5] five types of ambiguity of internal contextual grammars were considered. Among them, the following types of ambiguity are relevant to this paper. 1-*ambiguity* which is based on the axiom and the contexts used in the derivation but not on their order; 2-*ambiguity* which is based on the axiom, the contexts and the selectors used in the derivation but not their order. Another type of ambiguity called 0-*ambiguity* was considered by Ilie in [1] by taking into account of axioms only. There are many open problems formulated in [1,5] on different aspects of ambiguity of internal contextual grammars.

In this paper, we solve some open problems on the *inherent ambiguity* of internal contextual languages. In [5], it was shown that there are inherently 1-ambiguous languages with respect to internal contextual grammars without choice; here we (im)prove the result for grammars with arbitrary choice. Further, we show that there exist inherently 2-ambiguous languages with respect to internal contextual grammars with arbitrary selectors which are 1-unambiguous with respect to regular selectors, but we cannot prove this result for internal contextual grammars with finite selectors. However, we can show that there exist inherently 2-ambiguous languages with respect to *depth-first* contextual grammars (a variant of internal contextual grammars) with arbitrary selectors which are 1-unambiguous with respect to finite selectors.

2. Basic definitions

A finite non-empty set V is called an *alphabet*. We denote by V^* the free monoid generated by V, λ the *empty string*, and V^+ the set $V^* - \{\lambda\}$. The elements of V^* are called *words* or *strings*. For more details on formal language theory, we refer to [9].

An *internal contextual grammar with choice* is a construct

$$G = (V, A, (S_1, C_1), \ldots, (S_n, C_n)), \quad n \geqslant 1, \text{ where}$$

- V is a finite *alphabet*,
- $A \subseteq V^*$ is a finite set of *axioms*,
- $S_i \subseteq V^*$, $1 \leqslant i \leqslant n$, are the sets of *selectors*,
- $C_i \subseteq V^* \times V^*$, C_i finite, $1 \leqslant i \leqslant n$, are the sets of *contexts*.

The usual derivation in the *internal mode* is defined as $x \Longrightarrow_{\text{in}} y$ iff

$$x = x_1x_2x_3, \quad y = x_1ux_2vx_3, \quad \text{for } x_1, x_2, x_3 \in V^*, \ \ x_2 \in S_i, \ \ (u, v) \in C_i,$$

for some $1 \leqslant i \leqslant n$. The language generated by the above grammar G is $L_{\text{in}}(G) = \{x \in V^* \mid w \Longrightarrow^*_{\text{in}} x, w \in A\}$, where $\Longrightarrow^*_{\text{in}}$ is the reflexive transitive closure of the relation $\Longrightarrow_{\text{in}}$.

If all the sets of selectors $S_1, \ldots, S_n$ are in a family F of languages, we say that the grammar G is with F selection. When all the selectors are empty, G is said to be internal contextual grammar *without choice*. In such a case, we can apply any context $(u, v) \in C_i$, $1 \leqslant i \leqslant n$, to any substring of the derivation as there is no specified selector.

Depth-first contextual grammars were introduced in [4]. In depth-first contextual grammars, at each derivation step (except the first one), at least one of the contexts u or v which was introduced in the previous derivation step must be a subword of the currently used selector. More formally, given a contextual grammar $G = (V, A, (S_1, C_1), \ldots, (S_n, C_n))$, $n \geqslant 1$, a *depth-first derivation* in G is a derivation $w_1 \Longrightarrow_{\text{df}} w_2 \Longrightarrow_{\text{df}} \cdots \Longrightarrow_{\text{df}} w_m$, $m \geqslant 1$, where

(i) $w_1 \in A$, $w_1 \Longrightarrow w_2$ in the usual sense,

(ii) for each $j = 2, 3, \ldots, m$, if $w_{j-1} = x_1x_2x_3$, $w_j = x_1ux_2vx_3$ ((u, v) is the context adjoined to w_{j-1} in order to get w_j), then $w_j = y_1y_2y_3$, $w_{j+1} = y_1u'y_2v'y_3$, such that $y_2 \in S_i$, $(u', v') \in C_i$, for some i, $1 \leqslant i \leqslant n$, and y_2 contains one of the contexts u or v (which was adjoined in the previous derivation) as a substring.

The set of all words generated by a grammar G in this way is denoted by $L_{\text{df}}(G)$.

Now, let us formally introduce the types of ambiguity of internal contextual grammars discussed in the introduction. Given a contextual grammar

$$G = (V, A, (S_1, C_1), \ldots, (S_n, C_n)), \quad n \geqslant 1,$$

a *derivation* δ of a word z is given by

$$\delta = w_1 \Longrightarrow_{\text{in}} w_2 \Longrightarrow_{\text{in}} \cdots \Longrightarrow_{\text{in}} w_m = z, \quad m \geqslant 1, \text{ such that}$$

$$w_1 \in A,$$

$$w_j = x_{1,j}x_{2,j}x_{3,j}, \quad x_{1,j}, x_{2,j}, x_{3,j} \in V^*,$$

$$w_{j+1} = x_{1,j}u_i x_{2,j} v_i x_{3,j}, \quad x_{2,j} \in S_i, (u_i, v_i) \in C_i, \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant m-1.$$

The sequence of axiom and contexts used for the word z is given by

$$w_1, (u_1, v_1), (u_2, v_2), \ldots, (u_{m-1}, v_{m-1})$$

and is called the *control sequence* associated to δ. The sequence

$$w_1, ((u_1, v_1), x_{2,1}), ((u_2, v_2), x_{2,2}), \ldots, ((u_{m-1}, v_{m-1}), x_{2,m-1})$$

is called the *complete control sequence* of δ (it contains the axiom, contexts used and their corresponding selectors). If we take into the consideration the contexts (and selectors) used and not the order in which they are applied, then we obtain the *unordered control sequence* (and the *unordered complete control sequence*).

A contextual grammar G is said to be 0-*ambiguous* if there exist at least two different axioms $w_1, w_2 \in A$, $w_1 \neq w_2$, such that they both derive the same word z, i.e., $w_1 \Longrightarrow^+ z, w_2 \Longrightarrow^+ z$. A contextual grammar G is said to be 1-*ambiguous* (2-*ambiguous*) if there are two derivations in G having different unordered control sequences (unordered complete control sequences) and derive the same word. When we consider the order in control sequences, and in complete control sequences, we can define 3-*ambiguous* and 4-*ambiguous* grammars, but we do not discuss them here. Also, we omit the details about 5-*ambiguity*, which is another type of ambiguity based on the whole derivation.

A grammar which is not i-ambiguous, for some $i = 0, 1, 2, 3, 4, 5$, is said to be i-*unambiguous*. A language L is *inherently* i-*ambiguous* if every grammar G generating L is i-ambiguous. A language L for which an i-unambiguous grammar exists is called i-*unambiguous* and if the grammar is with arbitrary choice, we say that L is i-unambiguous (or inherently i-ambiguous, when every grammar G which generates L is i-ambiguous) with respect to arbitrary choice.

The ambiguity of internal contextual grammars was studied in [1,5,7]. The main results from these papers and the open problems we address here are the following.

Result 1. *There are inherently* 1-*ambiguous languages with respect to internal contextual grammars without choice which are* 0-*unambiguous with respect to internal contextual grammars with finite choice*.

Open problem 1. *Are there inherently* 0-*ambiguous languages with respect to internal contextual grammars without choice which are* 1-*unambiguous with respect to internal contextual grammars with finite choice*?

Result 2. *For each* $(i, j) \in \{(5, 4), (4, 3), (4, 2), (3, 2), (3, 1)\}$, *there are inherently* i-*ambiguous languages with respect to internal contextual grammars with arbitrary selection which are* j-*unambiguous with respect to internal contextual grammars with finite selection*.

Open problem 2. *For each* $(i, j) \in \{(1, 0), (2, 1)\}$, *are there inherently* i-*ambiguous languages with respect to internal contextual grammars with arbitrary choice which are* j-*unambiguous with respect to internal contextual grammars with finite choice*?

3. Results

In this section, we present our results which are solutions to the open problems mentioned in the previous section.

Theorem 1. *There are inherently* 0-*ambiguous languages with respect to internal contextual grammars without choice which are* 1-*unambiguous with respect to internal contextual grammars with finite selectors*.

Proof. In [1], it is proved that the language $L_1 = \{a, b\}^+$ is inherently 0-ambiguous with respect to internal contextual grammars without choice.

To prove that L_1 is 1-unambiguous with respect to finite selectors, consider the grammar $G_1 = (\{a, b\}, \{a, b\}, \{(\{a, b\}, \{(\lambda, a), (\lambda, b)\})\})$. It is obvious that $L(G_1) = L_1$. As a and b appear only on the right side of the contexts, all words can be generated from left to right in a unique way. Hence G_1 is 1-unambiguous with finite choice. □

We next solve the second open problem for the case $(i, j) \in (1, 0)$ mentioned in the previous section.

Theorem 2. *There are inherently* 1*-ambiguous languages with respect to internal contextual grammars with arbitrary choice which are* 0*-unambiguous with respect to internal contextual grammars with finite selection.*

Proof. Consider the language $L_2 = \{a^n ca^n \mid n \geqslant 0\} \cup \{ba^i ca^j d \mid i, j \geqslant 0\}$, and examine an arbitrary grammar G_2 generating this language. Considering the first part of the language, G_2 must have a context of the form (a^r, a^r), $r \geqslant 1$ (and the corresponding selector will be of the form $a^{k_1} ca^{k_2}$, $k_1, k_2 \geqslant 0$). Considering the second part of the language, G_2 must have contexts of the form (λ, a^s) and (a^t, λ), $s, t \geqslant 1$ (and their corresponding selectors will be of the form ba^{k_3} or $ba^{k_4} ca^{k_5}$ and $a^{k_6} d$ or $a^{k_7} ca^{k_8} d$, $k_i \geqslant 0$, $3 \leqslant i \leqslant 8$, respectively).

Now, we claim that this arbitrary grammar G_2 is 1-ambiguous. Consider the word $ba^n a^{rst} ca^{rst} a^m d \in L_2$ for large n and m. This word can be derived from $ba^n ca^m d$, in two ways: either by adjoining the context (a^r, a^r) for st times or by adjoining the contexts (λ, a^s) for rt times and (a^t, λ) for rs times. Therefore, there exists two different unordered control sequences such that one will have the context (a^r, a^r) and the other will have the contexts (λ, a^s), (a^t, λ), both derive the same word in L_2. It follows that, G_2 is 1-ambiguous. Hence, L_2 is inherently 1-ambiguous with respect to arbitrary selectors.

In order to prove that L_2 is 0-unambiguous with respect to finite selectors, consider the following grammar:

$$G_2' = (\{a, b, c, d\}, \{c, bcd\}, \{(c, (a, a)), (b, (\lambda, a)), (d, (a, \lambda))\}).$$

It is easy to see that $L(G_2') = L_2$ and G_2' is 0-unambiguous since any word in L_2 can be derived by only one of the axioms in G_2'. □

Theorem 3. *There exist inherently* 2*-ambiguous languages with respect to internal contextual grammars with arbitrary selectors which are* 1*-unambiguous with respect to internal contextual grammars with regular selectors.*

Proof. Consider the *crossed agreement* language

$$L_3 = \{a^n b^m c^n d^m \mid n, m \geqslant 1\}.$$

It is easy to see that any grammar G_3 which generates L_3 has the contexts of the form (a^i, c^i), (b^j, d^j), $i, j \geqslant 1$, and their corresponding selectors are of the form $a^{p_1} b^+ c^{p_2}$ and $b^{q_1} c^+ d^{q_2}$, $p_1, p_2, q_1, q_2 \geqslant 1$, respectively. Set $p = p_1 + p_2$, $q = q_1 + q_2$ and consider the word $a^p b^q c^p d^q \in L_3$, where p and q are very large. Now, the word $a^{p+i} b^{q+j} c^{p+i} d^{q+j} \in L_3$, can be derived from $a^p b^q c^p d^q$ in two distinct derivations δ_1, δ_2 which differ by their selectors: in one derivation, we have the unordered complete control sequence $\{(a^{p_1} b^q c^{p_2}, (a^i, c^i)), (b^{q_1} c^{p+i} d^{q_2}, (b^j, d^j))\}$ and in the another derivation, we have the unordered complete control sequence $\{(b^{q_1} c^p d^{q_2}, (b^j, d^j)), (a^{p_1} b^{q+j} c^{p_2}, (a^i, c^i))\}$. Note that the selectors appearing in δ_1 and δ_2 are distinct from each other (irrespective of their order of appearance), but the contexts are the same. Hence, L_3 is inherently 2-ambiguous with respect to arbitrary selectors.

In order to show that L_3 is 1-unambiguous with respect to regular selectors, consider the following grammar:

$$G_3' = (\{a, b, c, d\}, \{abcd\}, \{(ab^+c, (a, c)), (bc^+d, (b, d))\}).$$

Obviously, the grammar G_3' is 1-unambiguous. (Note that the language L_3 cannot be generated by a grammar with finite selectors.) □

Unlike internal contextual grammars, for the case of depth-first contextual grammars, we can show that there are inherently 2-ambiguous languages with respect to arbitrary selectors which are 1-unambiguous with respect to finite selectors.

Theorem 4. *There are inherently* 2*-ambiguous languages with respect to depth-first internal contextual grammars with arbitrary selectors which are* 1*-unambiguous with respect to internal contextual grammars with finite selectors.*

Proof. Consider the language

$$L_4 = \{a^n ba^m \mid n \geqslant m \geqslant 1\} \cup \{ba^k d \mid k \geqslant 1\}.$$

Let G_4 be an arbitrary contextual grammar which generates the language L_4 under the depth-first derivation. Consider the first part of L_4. Obviously, we need the contexts of the form (a^p, a^p), (a^q, λ), $p, q \geqslant 1$. Then, all the selectors must have a subword b. Otherwise, if a's are the only selector, then, more a's can be introduced in the right of b than in the left of b, and we get a string which is not in the language. At the same time, b alone cannot be a selector, because whenever (a^p, a^p) is introduced, the derivation of the next step should contain either the left context a^p or the right context a^p which was introduced in the previous step. So, the possible selectors are the subsets of the languages a^+b, ba^+ and a^+ba^+. But no subset of ba^+ can be a selector, because, if $(\{ba^k, k \geqslant 1\}, \{(a^p, a^p), (a^q, \lambda)\})$ is in G_4, then this can be applied to the axiom of the second part of L_4 and we get a word of the form $a^i ba^j d$, $i, j \geqslant 1$, not in L_4. Also, to generate the strings of the form $a^r ba^s$ where r, s are arbitrarily large, we need the context (a^p, a^p) and in order to cover any of the last introduced context a^p, we need arbitrarily long selector strings. Therefore, for the first part of the language, the possible selectors are the infinite subsets of the languages a^+, a^+ba^+, and the possible contexts are of the form $\{(a^p, a^p), (a^{q_1}, a^{q_2}), (a^q, \lambda)\}$, $p, q_1, q \geqslant 1$, $q_2 \geqslant 0$, $q_1 \geqslant q_2$.

Now, consider the second part of the language. The strings of the form ba^l, $l \geqslant 0$, cannot have the context of the form (λ, a^t), $t \geqslant 1$. Otherwise, applying this to the axiom of the first part of the language generates more a's in the right of b than in the left of b, which leads to a word not in L_4. Therefore, for the second part of the language, the possible selector and the corresponding context are of the form $(a^h d, (a^t, \lambda))$, $h \geqslant 0$, $t \geqslant 1$.

Now we claim that this grammar G_4 is 2-ambiguous. Consider the word $a^n ba^m \in L_4$ with n is arbitrarily larger than m, so that the context of the form (a^q, λ) is necessarily used when generating this word. Now, whenever, we apply the context (a^q, λ) in a derivation, we will have two choices to cover the last introduced context: either the left context a^q can be included in the selector $a^{i_1}b$, $i_1 \geqslant 1$ or we can include the right context λ to the selector $a^{i_2}b$, $i_2 \geqslant 1$ (where the occurrences of a not necessarily contain the last introduced left context a^q). Therefore, we can have two different selectors to choose for the next derivation step: one selector $a^{i_1}b$ covers the inserted left context a^q and the other selector $a^{i_2}b$ ($i_1 \neq i_2$) covers the inserted right context λ. Obviously, both the selectors derive the same word. If G_4 is having the other selector $a^r ba^s$, we can follow the similar argument. Therefore, the complete control sequences will have two different selectors (but the contexts are the same) which derive the same word in L_4. It follows that G_4 is 2-ambiguous. Hence, L_4 is inherently 2-ambiguous with respect to depth-first contextual grammars with arbitrary selectors.

To prove that L_4 is 1-unambiguous with respect to finite selectors, consider the following grammar:

$$G'_4 = (\{a, b, d\}, \{aba, aaba, bad\}, \{(\{aab, aba\}, \{(a, a), (a, \lambda)\}), (d, (a, \lambda))\}).$$

It is easy to see that $L_{\mathrm{df}}(G'_4) = L_4$. First we shall make sure that G'_4 is 2-ambiguous with respect to depth-first derivation. Assume that the word $a^j \underline{a}aba\underline{a}a^j \in L_4$, $j \geqslant 0$, was derived from the axiom under depth-first mode and the last selector used was aba (the underlined letters are the contexts which were introduced in the previous derivation step). The only possible selector for the next derivation step which contains one of the previous introduced context is $\underline{a}ab$. Now assume that we want to apply the context (a, λ). After (a, λ) is applied in the next derivation step by using the selector aab, we will have the word $a^j \underline{a}aab\underline{\lambda}aaa^j$. Now the next selector should contain one of the context $\underline{a}$ or $\underline{\lambda}$. Then we can have two choices in choosing the selector: either we can choose $aab\underline{\lambda}$ or $ab\underline{\lambda}a$. Hence, the unordered complete control sequences will have two different selectors aab and aba. Therefore, G'_4 is 2-ambiguous. As G'_4 has no other alternative contexts for (a, a) and (a, λ), G'_4 is 1-unambiguous. □

4. Final remarks

In this paper, we have solved some of the open problems listed in [1,5]. The open problem 2 mentioned in Section 2 for the case $(i, j) \in (2, 1)$ has not been completely solved for internal contextual grammars (in the sense that we cannot find a 1-unambiguous language with respect to finite choice which is inherently 2-ambiguous), but was solved for

depth-first contextual grammars. With the results presented in this paper, we now obtain a clear picture of the existence of inherently ambiguous languages (of all types of ambiguity) for internal contextual grammars.

So far, the results available in the literature were interested to know whether there exists any language which is inherently i-ambiguous ($i = 1, 2, 3, 4, 5$) with respect to arbitrary selectors, but not $(i - 1)$-ambiguous with respect to finite selectors. On the other hand, the following problem considers the question within the level of i-ambiguity itself.

Assume that G is an i-unambiguous ($i = 1, 2, 3, 4, 5$) grammar with arbitrary choice which generates a language L. If L can be generated by a grammar with finite choice (say G'), then is G' i-unambiguous or not? The answer to this problem is not trivial unless G itself is with finite choice. When the answer is negative (i.e., all G' are i-ambiguous), the language L becomes inherently i-ambiguous with respect to finite selectors, but i-unambiguous with respect to arbitrary selectors. In consequence, the problem can be rephrased as follows.

Open problem 3. *Are there inherently i-ambiguous languages* ($i = 1, 2, 3, 4, 5$) *with respect to internal contextual grammars with finite choice*, *which are i-unambiguous with respect to arbitrary* (*which is not finite*) *choice*?

Acknowledgements

I thank Prof. Kamala Krithivasan and Dr. Krishna S., for their suggestions and comments. I acknowledge my special thanks to the anonymous referee for his numerous useful corrections in the earlier versions of the paper.

References

[1] L. Ilie, On ambiguity in internal contextual languages, in: C. Martin-Vide (Ed.), Second Int. Conf. on Mathematical Linguistics, Tarragona, 1996, John Benjamins, Amsterdam, 1997, pp. 29–45.
[2] S. Marcus, Algebraic Linguistics, Analytical Models, Academic Press, New York, 1967.
[3] S. Marcus, Contextual Grammars, Rev. Roumaine Math. Pures Appl. 14 (1969) 1525–1534.
[4] C. Martin-Vide, J. Miquel-Verges, Gh. Păun, Contextual grammars with depth-first derivation, 10th Twente Workshop on Language Technology, Algebraic Methods in Language Processing, Twente, 1995, pp. 225–233.
[5] C. Martin-Vide, J. Miguel-Verges, Gh. Păun, A. Salomaa, Attempting to define the ambiguity in internal contextual languages, in: C. Martin-Vide (Ed.), Second Int. Conf. on Mathematical Linguistics, Tarragona, 1996, John Benjamins, Amsterdam, 1997, pp. 59–81.
[6] Gh. Păun, Contextual Grammars, The Publishing House of the Romanian Academy of Sciences, Bucuresti, 1982.
[7] Gh. Păun, Marcus Contextual Grammars, Kluwer Academic Publishers, Dordrecht, 1997.
[8] Gh. Păun, X.M. Nguyen, On the inner contextual grammars, Rev. Roumaine Math. Pures Appl. 25 (1980) 641–651.
[9] A. Salomaa, Formal Languages, Academic Press, New York, 1973.