AUTHOR QUERY FORM

ELSEVIER	**Journal:** TCS **Article Number:** 11102	**Please e-mail your responses and any corrections to:** **E-mail:** **corrections.esch@elsevier.vtex.lt**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult *http://www.elsevier.com/artworkinstructions*

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	**Query / Remark: click on the Q link to go** **Please insert your reply or correction at the corresponding line in the proof**
Q1	Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact <t.boopathy@elsevier.com> immediately prior to returning your corrections. (p. 1/ line 1)
Q2	The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly and are presented in the desired order. (p. 1/ line 13)
Q3	Please indicate which author should be marked as 'Corresponding author'. (p. 1/ line 14)

Please check this box if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

Characterization of context-free languages

M. Badano, D. Vaggione

Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Córdoba, Argentina

ARTICLE INFO

Article history:
Received 6 September 2016
Received in revised form 12 January 2017
Accepted 2 March 2017
Available online xxxx
Communicated by D. Perrin

Keywords:
Context-free language

ABSTRACT

In this note we present a simple condition upon which a formal grammar produces a context-free language.

Context-free grammars are one of the most investigated families of grammars in formal language theory. They provide a precise mechanism for describing the basic recursive structure of sentences in human language, and also have played a central role in compiler technology, as in the implementation of parsers, for example. In this note we give a characterization of context-free languages (i.e. languages generated by context free grammars), which is based on Greibach [1] normal form.

In order to state the result we revise the basic definitions. A *grammar* is a 4-tuple $G = (V, T, S, P)$, where V and T are finite sets of *variables* and *terminals*, respectively, $S \in V$ is the *start symbol* and P is a finite set of productions of the form $\alpha \to \beta$, with $\alpha, \beta \in (V \cup T)^*$ and α non-empty. We assume that V and T are disjoint. The grammar G is *context-free* if all its productions are of the form $A \to \beta$ where $A \in V$ and $\beta \in (V \cup T)^*$. A language L is *context-free* if L can be generated by a context-free grammar. Let ε denote the empty string.

Theorem 1. *Let L be a language without ε. Then L is context-free if an only if L can be generated by a grammar for which every production is of the form $\alpha \to a\beta$, where α is a non-empty string of variables, a is a terminal and β is a (possibly empty) string of variables.*

Before we prove the theorem, we need to state some notation and previous results. Let $G = (V, T, S, P)$ be a grammar. We write $\gamma_1 \underset{G}{\Rightarrow} \gamma_2$ when there exist $\lambda_1, \lambda_2 \in (V \cup T)^*$ and a production $\alpha \to \beta$ in P such that $\gamma_1 = \lambda_1\alpha\lambda_2$ and $\gamma_2 = \lambda_1\beta\lambda_2$. For $n \geq 0$, we write $\gamma_1 \overset{n}{\underset{G}{\Rightarrow}} \gamma_2$ when there exist $\alpha_1, \dots, \alpha_{n+1}$ such that

$$\gamma_1 = \alpha_1,\ \gamma_2 = \alpha_{n+1} \text{ and } \alpha_i \underset{G}{\Rightarrow} \alpha_{i+1},\ i = 1, \dots, n$$

(note that $\gamma_1 \overset{0}{\underset{G}{\Rightarrow}} \gamma_2$ iff $\gamma_1 = \gamma_2$). We use $\overset{*}{\underset{G}{\Rightarrow}}$ to denote the reflexive and transitive closure of $\underset{G}{\Rightarrow}$. As usual, we define the *language generated by* G to be

$$L(G) = \{w \in T^* : S \overset{*}{\underset{G}{\Rightarrow}} w\}.$$

E-mail addresses: mbadano@famaf.unc.edu.ar (M. Badano), vaggione@famaf.unc.edu.ar (D. Vaggione).

http://dx.doi.org/10.1016/j.tcs.2017.03.002
0304-3975/© 2017 Published by Elsevier B.V.

Let $G=(V,T,S,P)$ be a grammar for which every production is of the form $\alpha\to a\beta$, where α is a non-empty string of variables, a is a terminal and β is a (possibly empty) string of variables. We will write

$$\gamma_1 \underset{G}{\Rightarrow} \gamma_2 \text{ (leftmost)}$$

when there exist $\lambda_1\in T^*$, $\lambda_2\in(V\cup T)^*$ and a production $\alpha\to\beta$ such that $\gamma_1=\lambda_1\alpha\lambda_2$ and $\gamma_2=\lambda_1\beta\lambda_2$. We write

$$\gamma_1 \overset{n}{\underset{G}{\Rightarrow}} \gamma_2 \text{ (leftmost)}$$

when there exist $\alpha_1,\dots,\alpha_{n+1}$ such that

$$\gamma_1=\alpha_1,\ \gamma_2=\alpha_{n+1} \text{ and } \alpha_i \underset{G}{\Rightarrow} \alpha_{i+1} \text{ (leftmost)},\ i=1,\dots,n.$$

Lemma 2. *Let $G=(V,T,S,P)$ be a grammar for which every production is of the form $\alpha\to a\beta$, where α is a non-empty string of variables, a is a terminal and β is a (possibly empty) string of variables. Suppose $\beta_1x_1\beta_2x_2\dots\beta_kx_k\beta_{k+1} \overset{n}{\underset{G}{\Rightarrow}} w$, with $w\in T^*$, $n\geq 1$, $k\geq 0$, $\beta_1,\dots,\beta_{k+1}\in V^*$, $x_1,\dots,x_k\in T^*-\{\varepsilon\}$. There exist $w_1,\dots,w_{k+1}\in T^*$ and $n_1,\dots,n_{k+1}\geq 0$ such that*

1. $\beta_i \overset{n_i}{\underset{G}{\Rightarrow}} w_i$ *(leftmost), for $i=1,\dots,k+1$.*
2. $\sum_{i=1}^{k+1} n_i = n$.
3. $w_1x_1w_2x_2\dots w_kx_kw_{k+1}=w$.

Proof. We proceed by induction on n. The case $n=1$ is trivial. Assume the result is valid for n and let

$$\beta_1x_1\dots\beta_kx_k\beta_{k+1} \overset{n+1}{\underset{G}{\Rightarrow}} w.$$

Then there exist $j\geq 1$ and $\delta_1,\delta_2,\alpha\in V^*$ such that $\beta_j=\delta_1\alpha\delta_2$, $\alpha\to a\gamma\ \in P$ and

$$\beta_1x_1\dots x_{j-1}\delta_1a\gamma\delta_2x_j\dots\beta_kx_k\beta_{k+1} \overset{n}{\underset{G}{\Rightarrow}} w.$$

By the inductive hypothesis, we have that there are $w_1,\dots,w_{k+2}\in T^*$ and $n_1,\dots,n_{k+2}\geq 0$ such that

1. $\beta_i \overset{n_i}{\underset{G}{\Rightarrow}} w_i$ (leftmost) for $i<j$, $\delta_1 \overset{n_j}{\underset{G}{\Rightarrow}} w_j$ (leftmost), $\gamma\delta_2 \overset{n_{j+1}}{\underset{G}{\Rightarrow}} w_{j+1}$ (leftmost) and $\beta_i \overset{n_{i+1}}{\underset{G}{\Rightarrow}} w_{i+1}$ for $i>j$ (leftmost).
2. $\sum_{i=1}^{k+2} n_i = n$.
3. $w_1x_1\dots x_{j-1}w_jaw_{j+1}x_j\dots w_{k+1}x_kw_{k+2}=w$.

So

$$\beta_j=\delta_1\alpha\delta_2 \overset{n_j}{\underset{G}{\Rightarrow}} w_j\alpha\delta_2 \overset{1}{\underset{G}{\Rightarrow}} w_ja\gamma\delta_2 \overset{n_{j+1}}{\underset{G}{\Rightarrow}} w_jaw_{j+1}$$

and then we have

$$\beta_j \overset{n_j+n_{j+1}+1}{\underset{G}{\Rightarrow}} w_jaw_{j+1} \text{ (leftmost)}.$$

The proof easily follows from this. □

Corollary 3. *If $\alpha \overset{n}{\underset{G}{\Rightarrow}} w$ then $\alpha \overset{n}{\underset{G}{\Rightarrow}} w$ (leftmost).*

Proof. It is a straightforward inductive argument. □

Proof of Theorem 1. Let L be a context-free language. We recall that a grammar G is in Greibach Normal Form if every production rule is of the form $A\to a\beta$ where $A\in V$, $a\in T$ and $\beta\in V^*$. If L is a context-free language without ε then there is a grammar in Greibach Normal Form G such that $L=L(G)$ (see [1]), which proves one direction of Theorem 1.

Suppose that $L=L(G)$ where G is a grammar such that every production rule is of the form $\alpha\to a\beta$ where $\alpha\in V^*\setminus\{\varepsilon\}$, $a\in T$ and $\beta\in V^*$. Define the following sets:

$$N_G=\{\alpha\in V^*:\alpha\to a\beta\in P \text{ for some } a\in T,\ \beta\in V^*\},$$

$$M_G=\{\beta\in V^*: \text{ there is } \alpha\in N_G \text{ such that } \beta \text{ is a prefix of } \alpha \text{ and } \beta\neq\alpha\}.$$

For $\alpha \in N_G$ and $\beta \in M_G$ let $V_{\alpha,\beta}$ be a new variable. Let $\bar{G} = (\bar{V}, T, \bar{P}, V_{S,\varepsilon})$, where

$$\bar{V} = \{V_{\alpha,\beta} : \alpha \in N_G \text{ and } \beta \in M_G\},$$

and $\bar{P} = \bar{P}_1 \cup \bar{P}_2$ where $\bar{P}_1 = \{V_{\alpha,\beta} \to a : \alpha \to a\beta \in P \text{ and } \beta \in M_G\}$ and

$$\begin{aligned}\bar{P}_2 = \{V_{\alpha,\beta} \to aV_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\ldots V_{\tau_{k-1}\beta_k,\tau_k} : k \geq 1,\\ \alpha \to a\beta_1\ldots\beta_k\beta_{k+1} \in P, \tau_i \in M_G \text{ and } \beta_i \neq \varepsilon \text{ for } i = 1,\ldots,k,\\ \beta_1 \in N_G, \tau_i\beta_{i+1} \in N_G \text{ for } i = 1,\ldots,k-1, \text{ and } \beta = \tau_k\beta_{k+1} \in M_G\}.\end{aligned}$$

The following example shows the construction of the grammar $\bar{G}$ for a given grammar G.

Example 4. Let $G = (V, T, P, S)$ where $V = \{S, A, B\}$, $T = \{a, b, c, d\}$ and P is the set of following production rules

$$\begin{aligned} S &\to aAB\\ A &\to aAB\\ AB &\to c\\ BB &\to d\\ B &\to b\end{aligned}$$

According to the previous definition we have

$$N_G = \{S, A, B, AB, BB\} \text{ and } M_G = \{\varepsilon, A, B\}.$$

The set of production rules $\bar{P}$ is given by

$V_{S,\varepsilon} \to aV_{A,\varepsilon}V_{B,\varepsilon}$	$V_{A,\varepsilon} \to aV_{A,\varepsilon}V_{B,\varepsilon}$	$V_{AB,\varepsilon} \to c$
$V_{S,\varepsilon} \to aV_{A,A}V_{AB,\varepsilon}$	$V_{A,\varepsilon} \to aV_{A,A}V_{AB,\varepsilon}$	$V_{BB,\varepsilon} \to d$
$V_{S,\varepsilon} \to aV_{A,B}V_{BB,\varepsilon}$	$V_{A,\varepsilon} \to aV_{A,B}V_{BB,\varepsilon}$	$V_{B,\varepsilon} \to b$
$V_{S,\varepsilon} \to aV_{AB,\varepsilon}$	$V_{A,\varepsilon} \to aV_{AB,\varepsilon}$	
$V_{S,A} \to aV_{A,\varepsilon}V_{B,A}$	$V_{A,A} \to aV_{A,\varepsilon}V_{B,A}$	
$V_{S,A} \to aV_{A,A}V_{AB,A}$	$V_{A,A} \to aV_{A,A}V_{AB,A}$	
$V_{S,A} \to aV_{A,B}V_{BB,A}$	$V_{A,A} \to aV_{A,B}V_{BB,A}$	
$V_{S,A} \to aV_{AB,A}$	$V_{A,A} \to aV_{AB,A}$	
$V_{S,B} \to aV_{A,\varepsilon}V_{B,B}$	$V_{A,B} \to aV_{A,\varepsilon}V_{B,B}$	
$V_{S,B} \to aV_{A,A}V_{AB,B}$	$V_{A,B} \to aV_{A,A}V_{AB,B}$	
$V_{S,B} \to aV_{A,B}V_{BB,B}$	$V_{A,B} \to aV_{A,B}V_{BB,B}$	
$V_{S,B} \to aV_{AB,B}$	$V_{A,B} \to aV_{AB,B}$	
$V_{S,B} \to aV_{A,\varepsilon}$	$V_{A,B} \to aV_{A,\varepsilon}$	

Observe that even when the size of $\bar{P}$ is much larger than the size of P, most of the new production rules are useless (either unreachable or unproductive).

We will prove that, for any grammar G under the hypothesis of Theorem 1, $L(G) = L(\bar{G})$. Indeed we will prove the following more general result from which $L(G) = L(\bar{G})$ follows.

Claim 5. *If $\alpha \in V^* \setminus \{\varepsilon\}$ and $w \in T^*$ then $\alpha \overset{*}{\underset{G}{\Rightarrow}} w$ iff there exist $k \geq 1$, $\beta_i \neq \varepsilon$ for $i = 1,\ldots,k$, and $\tau_1,\ldots,\tau_{k-1} \in M_G$, such that $\beta_1 \in N_G$, $\tau_i\beta_{i+1} \in N_G$ for $i = 1,\ldots,k-1$,*

$$\alpha = \beta_1\ldots\beta_k \text{ and } V_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\ldots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w$$

We remark that in the case $k=1$ the expression $V_{\beta_1,\tau_1}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w$ should be interpreted as $V_{\beta_1,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w$.

In order to prove this, first we will see that, for every n, $\alpha \overset{n}{\underset{G}{\Rightarrow}} w$ implies there exist $\beta_1,\dots,\beta_k \neq \varepsilon$ and $\tau_1,\dots,\tau_{k-1}$ with $k \geq 1$ such that

$$\alpha = \beta_1\dots\beta_k \text{ and } V_{\beta_1,\tau_1}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w,$$

and we proceed by induction on n.

If $n=1$ then we have $\alpha \underset{G}{\Rightarrow} w$, which implies $\alpha \to w \in P$. By definition we have $V_{\alpha,\varepsilon} \to w \in \bar{P}$ and then $V_{\alpha,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w$.

Now assume the result is valid for n and let $\alpha \overset{n+1}{\underset{G}{\Rightarrow}} w$. By Corollary 3, we may assume that this is a leftmost derivation. Then there are $\bar{\alpha},\delta,\gamma \in V^*$ and $\bar{w} \in T^*$ such that

$$\alpha = \bar{\alpha}\gamma,\ \bar{\alpha} \to a\delta \in P,\ w = a\bar{w} \text{ and } \delta\gamma \overset{n}{\underset{G}{\Rightarrow}} \bar{w}.$$

So, by the inductive hypothesis, there exist $\beta_1,\dots,\beta_k \neq \varepsilon$ and $\tau_1,\dots,\tau_{k-1}$ with $k \geq 1$ such that

$$\delta\gamma = \beta_1\dots\beta_k \text{ and } V_{\beta_1,\tau_1}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} \bar{w}.$$

We have here two possible situations, either δ is a prefix of β_1 with $\delta \neq \beta_1$, or β_1 is a prefix of δ.

Case δ is a prefix of β_1 and $\delta \neq \beta_1$. Let $\varphi \in V^* \setminus \{\varepsilon\}$ be such that $\beta_1 = \delta\varphi$. Observe that, in this case, $\gamma = \varphi\beta_2\dots\beta_k$. We take

$$\begin{aligned} k' &= k+1 \\ \beta_1' &= \bar{\alpha} \\ \tau_1' &= \delta \\ \beta_2' &= \varphi \\ \beta_i' &= \beta_{i-1} \text{ for } i = 3,\dots,k+1 \\ \tau_i' &= \tau_{i-1} \text{ for } i = 2,\dots,k \end{aligned}$$

So we have

$$\alpha = \bar{\alpha}\gamma = \beta_1'\varphi\beta_2\dots\beta_k = \beta_1'\beta_2'\beta_3'\dots\beta_{k+1}'$$

and since $\bar{\alpha} \to a\delta$ and $\delta \in M_G$ (δ is a prefix of β_1 and $\delta \neq \beta_1$)

$$V_{\beta_1',\tau_1'} \to a \in \bar{P}$$

Also note that $V_{\tau_1'\beta_2',\tau_2'} = V_{\beta_1,\tau_1}$ and then

$$V_{\beta_1',\tau_1'}V_{\tau_1'\beta_2',\tau_2'}V_{\tau_2'\beta_3',\tau_3'}\dots V_{\tau_k'\beta_{k+1}',\varepsilon} \overset{1}{\underset{\bar{G}}{\Rightarrow}} aV_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} a\bar{w} = w.$$

Case β_1 is a prefix of δ. If $\delta = \beta_1\dots\beta_k$, then $\gamma = \varepsilon$ and $\alpha = \bar{\alpha}$. Since $\bar{\alpha} \to a\delta \in P$, we have

$$\alpha \to a\beta_1\dots\beta_k \in P$$

which implies

$$V_{\alpha,\varepsilon} \to aV_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \in \bar{P}$$

and then we have

$$V_{\alpha,\varepsilon} \underset{\bar{G}}{\Rightarrow} aV_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} a\bar{w} = w.$$

If $\delta \neq \beta_1\dots\beta_k$, then $\gamma \neq \varepsilon$. Since β_1 is a prefix of δ and $\gamma \neq \varepsilon$ we have that $k \geq 2$ and there is $1 \leq j \leq k-1$ and $\varphi,\psi \in V^*$ with $\psi \neq \varepsilon$ such that $\delta = \beta_1\dots\beta_j\varphi$ and $\beta_{j+1} = \varphi\psi$. We take

$$\begin{aligned} \beta_1' &= \bar{\alpha} \\ \tau_1' &= \tau_j\varphi \\ \beta_2' &= \psi \\ \beta_i' &= \beta_{j+i-1} \text{ for } i = 3,\dots,k-j+2 \\ \tau_i' &= \tau_{j+i-1} \text{ for } i = 2,\dots,k-j+1 \end{aligned}$$

Observe that $\tau_j\varphi$ is a proper prefix of $\tau_j\beta_{j+1}$ so we have $\tau_1' \in M_G$. Since $\bar{\alpha} \to a\beta_1\dots\beta_j\varphi$, by definition we have

Please cite this article in press as: M. Badano, D. Vaggione, Characterization of context-free languages, Theoret. Comput. Sci. (2017), http://dx.doi.org/10.1016/j.tcs.2017.03.002

$$V_{\beta'_1,\tau'_1} \to aV_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\dots V_{\tau_{j-1}\beta_j,\tau_j} \in \bar{P}$$

and then

$$V_{\beta'_1,\tau'_1}V_{\tau'_1\beta'_2,\tau'_2}\dots V_{\tau'_{k-1}\beta'_k,\varepsilon} \underset{\bar{G}}{\Rightarrow} aV_{\beta_1,\tau_1}\dots V_{\tau_{j-1}\beta_j,\tau_j}V_{\tau_j\varphi\psi,\tau_{j+1}}\dots V_{\tau'_{k-1}\beta'_k,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w.$$

To see the other direction we will prove by induction on n that if there are $\beta_1,\dots,\beta_k \neq \varepsilon$ and $\tau_1,\dots,\tau_{k-1}$ such that

$$V_{\beta_1,\tau_1}V_{\tau_1\beta_2,\tau_2}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{n}{\underset{\bar{G}}{\Rightarrow}} w$$

then $\beta_1\dots\beta_k \overset{*}{\underset{G}{\Rightarrow}} w$.

If $n = 1$ then $k = 1$ and $V_{\beta_1,\varepsilon} \to w \in \bar{P}$ which implies $\beta_1 \to w \in P$ and then $\beta_1 \overset{*}{\underset{G}{\Rightarrow}} w$.

Now assume the result is valid for n and suppose there exist $\beta_1,\dots,\beta_k$ and $\tau_1,\dots,\tau_{k-1}$ such that

$$V_{\beta_1,\tau_1}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{n+1}{\underset{\bar{G}}{\Rightarrow}} w$$

Without loss of generality, we may assume that the above derivation is leftmost. So there exist $a \in T$, $\bar{w} \in T^*$ and $\zeta \in \bar{V}^*$ such that $w = a\bar{w}$, and

$$V_{\beta_1,\tau_1} \to a\zeta \in \bar{P}, \tag{1}$$

$$\zeta V_{\tau_1\beta_2,\tau_2}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{n}{\underset{\bar{G}}{\Rightarrow}} \bar{w}. \tag{2}$$

If $\zeta = \varepsilon$ then by definition of $\bar{P}$ we have

$$\beta_1 \to a\tau_1 \in P,$$

and by the inductive hypothesis on (2)

$$\tau_1\beta_2\dots\beta_k \overset{*}{\underset{G}{\Rightarrow}} \bar{w}.$$

Then we have

$$\beta_1\beta_2\dots\beta_k \underset{G}{\Rightarrow} a\tau_1\beta_2\dots\beta_k \overset{*}{\underset{G}{\Rightarrow}} a\bar{w} = w.$$

If $\zeta \neq \varepsilon$ then there are $\beta'_1,\dots,\beta'_m \neq \varepsilon$ and $\tau'_1,\dots,\tau'_m \in V^*$ such that $\zeta = V_{\beta'_1,\tau'_1}\dots V_{\tau'_{m-1}\beta'_m,\tau'_m}$ and

$$V_{\beta_1,\tau_1} \to aV_{\beta'_1,\tau'_1}\dots V_{\tau'_{m-1}\beta'_m,\tau'_m} \in \bar{P}$$

which implies that there is β'_{m+1} such that

$$\beta_1 \to a\beta'_1\dots\beta'_m\beta'_{m+1} \in P \text{ and } \tau_1 = \tau'_m\beta'_{m+1}.$$

Then we can rewrite (2) as

$$V_{\beta'_1,\tau'_1}\dots V_{\tau'_{m-1}\beta'_m,\tau'_m}V_{\tau'_m\beta'_{m+1}\beta_2,\tau_2}\dots V_{\tau_{k-1}\beta_k,\varepsilon} \overset{n}{\underset{\bar{G}}{\Rightarrow}} \bar{w}.$$

By the inductive hypothesis we have

$$\beta'_1\dots\beta'_m(\beta'_{m+1}\beta_2)\dots\beta_k \overset{*}{\underset{G}{\Rightarrow}} \bar{w}$$

So

$$\beta_1\dots\beta_k \underset{G}{\Rightarrow} a\beta'_1\dots\beta'_m\beta'_{m+1}\beta_2\dots\beta_k \overset{*}{\underset{G}{\Rightarrow}} a\bar{w} = w$$

which concludes the proof of Claim 5.

Now by Claim 5 we have

$$L(G) = \{w \in T^* : S \overset{*}{\underset{G}{\Rightarrow}} w\} = \{w \in T^* : V_{S,\varepsilon} \overset{*}{\underset{\bar{G}}{\Rightarrow}} w\} = L(\bar{G})$$

and we have proved Theorem 1. □

References

[1] S.A. Greibach, A new normal-form theorem for context-free phrase structure grammars, J. ACM 12 (1) (1965) 42–52.

Please cite this article in press as: M. Badano, D. Vaggione, Characterization of context-free languages, Theoret. Comput. Sci. (2017), http://dx.doi.org/10.1016/j.tcs.2017.03.002