JOURNAL OF COMPUTER AND SYSTEM SCIENCES 6, 267–282 (1972)

The Relation Between Derivations and Syntactical Structures in Phrase-Structure Grammars

JÜRGEN EICKEL

Technische Universität München, München, West Germany

AND

JACQUES LOECKX*

Technische Hogeschool Twente, Enschede, Netherlands

Received September 9, 1971

Conditions for a phrase-structure grammar (Chomsky 0-type grammar) are established which warrant that any of its derivations univocally defines a syntactical structure of the sentence. The simplifications for the particular cases of context-sensitive and context-free grammars are indicated.

1. INTRODUCTION

It is well known that, for context-free grammars, a leftmost derivation of a sentence univocally defines its syntactical tree; the construction of this tree consists of an iterative procedure during which two successive strings in the leftmost derivation are compared (see, e.g., [1, 2]).

These—or similar—properties do not hold for leftmost derivations in general phrase-structure grammars [3–5] nor for (not necessarily leftmost) derivations in context-free grammars (see, e.g., [6, 7]).

In this paper conditions for a general phrase-structure grammar are established under which each of its derivations univocally defines a syntactical structure (Sections 2–5). The simplifications for context-sensitive and context-free grammars are indicated in Sections 6 and 7. The case of leftmost derivations is discussed briefly in Section 8.

* Present address: Universität des Saarlandes, Saarbrücken, West Germany.

2. Preliminary Definitions

A *general phrase-structure grammar* is defined by a 4-tuple $(\mathbf{V}, \mathbf{T}, \mathbf{R}, Z)$:

$\mathbf{V}$ is a finite set of symbols,

$\mathbf{T}$ is a subset of $\mathbf{V}$,

$\mathbf{R}$ is a finite set of ordered pairs $\phi \to \psi$ called *rules* with ϕ in $\mathbf{V}^* - \mathbf{T}^*$, ψ in $\mathbf{V}^*$ and $\phi \neq \psi$ (see Footnotes 1, 2),

Z is an element of $\mathbf{V} - \mathbf{T}$.

A *sentence* is a string x in $\mathbf{T}^*$ for which there exists a finite sequence

$$\omega_1, \omega_2, \ldots, \omega_n \qquad (n > 1)$$

and strings σ_i, τ_i, ϕ_i and ψ_i in $\mathbf{V}^*$ $(1 \leqslant i \leqslant n - 1)$ such that

$$\omega_1 = Z, \tag{1}$$

$$\omega_n = x, \tag{2}$$

$$\omega_i = \sigma_i \phi_i \tau_i \qquad (1 \leqslant i \leqslant n-1), \tag{3}$$

$$\omega_{i+1} = \sigma_i \psi_i \tau_i \qquad (1 \leqslant i \leqslant n-1), \tag{4}$$

and

$$\phi_i \to \psi_i \text{ is in } \mathbf{R} \qquad (1 \leqslant i \leqslant n-1). \tag{5}$$

The sequence $\omega_1, \omega_2, \ldots, \omega_n$ is called a *derivation* of the sentence; each element ω_i $(1 \leqslant i \leqslant n)$ of this sequence is called a *sentential form*. The set of all sentences is the *language*.

A *syntactical structure* of a sentence [4, 7–9] is a directed graph with labeled vertices which may be informally defined as follows: with each rule

$$\alpha_1 \alpha_2 \cdots \alpha_M \to \beta_1 \beta_2 \cdots \beta_N \ (\alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_N \text{ in } \mathbf{V};\ M \geqslant 1;\ N \geqslant 0)$$

applied in a given derivation step of the sentence there corresponds a branching with M upper edges and N lower edges; the upper edges originate from M vertices labeled $\alpha_1, \alpha_2, \ldots, \alpha_M$ and the N lower edges point to N vertices labeled $\beta_1, \beta_2, \ldots, \beta_N$.

Examples of general phrase-structure grammars, derivations and syntactical structures are in Figs. 1, 4, 5 and 6.

[1] $\mathbf{V}^*$ denotes the free monoid generated by $\mathbf{V}$

[2] We exclude the case of rules $\phi \to \phi$.

3. A Precise Formulation of the Problem

Associate with an arbitrarily given general phrase-structure grammar $G = (\mathbf{V}, \mathbf{T}, \mathbf{R}, Z)$ the set of triples $\mathbf{S} \subseteq \mathbf{V}^* \times \mathbf{V}^* \times \mathbf{R}$ and the function $F : \mathbf{S} \to \mathbf{V}^* \times \mathbf{V}^*$ defined as follows:

$$\mathbf{S} = \{(\sigma, \tau, \phi \to \psi) \mid \sigma\phi\tau \text{ and } \sigma\psi\tau \text{ are sentential forms}\}.$$

$$F : \mathbf{S} \to \mathbf{V}^* \times \mathbf{V}^* : F(\sigma, \tau, \phi \to \psi) = (\sigma\phi\tau, \sigma\psi\tau).$$

The grammar G is said to be *resolvable* when its associated function F is injective. It is the goal of this paper to establish the sufficient and necessary conditions for a grammar to be resolvable.

Clearly, if a general phrase-structure grammar is resolvable, each derivation, say $\omega_1, \omega_2, \ldots, \omega_n$, univocally defines a syntactical structure; this structure may be obtained in a straightforward way from the sequence of triplets $F^{-1}(\omega_1, \omega_2)$, $F^{-1}(\omega_2, \omega_3), \ldots, F^{-1}(\omega_{n-1}, \omega_n)$ (see Footnote 3), where F^{-1} denotes the inverse function of F.

Note that the converse is not true: if each derivation univocally determines a syntactical structure the grammar is not necessarily resolvable; this fact is illustrated by Fig. 1.

4. Two Lemmas

4.1. Lemma 1. *If x, y, z are strings (over a given vocabulary) satisfying*

$$xz = zy$$

then

$$x^k z = zy^k$$

for all $k \geqslant 0$.[4]

Proof. Clearly, the lemma is true for $k = 0$. Suppose now that the lemma is true for $k - 1$. Then

$$\begin{aligned} x^k z &= xx^{k-1}z \\ &= xzy^{k-1} \\ &= zyy^{k-1} \\ &= zy^k. \end{aligned}$$

[3] Such a sequence of triplets is what Griffiths [3] calls a derivation; see also [7].

[4] Per definition $x^0 = \epsilon$ and $x^k = xx^{k-1}$; ϵ denotes the empty string.

(a) $G = (\{A, B, a\},$
$\{a\},$
$\{A \to BB, B \to \epsilon\},$
$A)$

Note: ϵ denotes the empty string.

(b) A
BB
B
ϵ

(c)

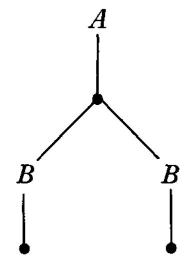

FIG. 1. (a) is a general phrase-structure (more precisely: context-free) grammar; (b) is a derivation of the sentence ϵ which is, by the way, the unique sentence of this grammar; (c) is the syntactical structure of this sentence. As $F(\epsilon, B, B \to \epsilon) = (BB, B) = F(B, \epsilon, B \to \epsilon)$, the grammar G is not resolvable; the derivation (b) nevertheless defines a single syntactical structure, viz., (c).

4.2. LEMMA 2. *If x, y, z are strings (over a given vocabulary) satisfying*

$$xz = zy$$

and if x and y are nonempty, there exist a string u, a nonempty string v and an integer $n \geqslant 0$ such that

$$x = uv,$$

$$y = vu,$$

and

$$z = (uv)^n u.$$

Proof. As x is nonempty there exists a (univocally defined) integer $n \geqslant 0$ such that

$$| x^n | \leqslant | z | < | x^{n+1} |,$$

where $| \phi |$ denotes the length of string ϕ. But according to Lemma 1

$$x^{n+1} z = z y^{n+1}. \tag{1}$$

Hence there exists a (nonempty) string v such that

$$x^{n+1} = zv \tag{2}$$

and

$$0 < | v | \leqslant | x |.$$

Hence there exists a string u such that

$$x = uv \tag{3}$$

and

$$0 \leqslant | u | < | x |.$$

Substituting (3) into (2) yields

$$z = (uv)^n u. \tag{4}$$

Substituting (3) and (4) into (1) yields

$$y = vu.$$

5. The Main Result

5.1. Theorem 1. *A general phrase-structure grammar* $(\mathbf{V}, \mathbf{T}, \mathbf{R}, Z)$ *is resolvable if and only if none of the following four conditions is satisfied*:

(i) *there exist strings* $\gamma, \delta, \sigma, \tau, \phi, \psi$ *in* $\mathbf{V}^*$ *such that* $\gamma\delta \neq \epsilon$, $\phi \rightarrow \psi$ *and* $\gamma\phi\delta \rightarrow \gamma\psi\delta$ *are in* $\mathbf{R}$, $\sigma\gamma\phi\delta\tau$ *and* $\sigma\gamma\psi\delta\tau$ *are sentential forms*;

(ii) *there exist strings* $\gamma, \delta, \sigma, \tau, \phi, \psi$ *in* $\mathbf{V}^*$ *such that* $\gamma\delta \neq \epsilon$, $\phi\delta \rightarrow \psi\delta$ *and* $\gamma\phi \rightarrow \gamma\psi$ *are in* $\mathbf{R}$, $\sigma\gamma\phi\delta\tau$ *and* $\sigma\gamma\psi\delta\tau$ *are sentential forms*;

(iii) *there exist strings* $\gamma, \delta, \sigma, \tau, \psi_1, \psi_2$ *in* $\mathbf{V}^*$ *and an integer* $k \geqslant 0$ *such that* $\gamma^k\delta\psi_1\psi_2 \neq \epsilon$, $\gamma\delta\psi_1 \rightarrow \psi_1$ *and* $\psi_2\delta\gamma \rightarrow \psi_2$ *are in* $\mathbf{R}$, $\sigma\psi_2(\delta\gamma)^{k+1}\delta\psi_1\tau$ *and* $\sigma\psi_2(\delta\gamma)^k\delta\psi_1\tau$ *are sentential forms*;

(iv) *there exist strings* $\gamma, \delta, \sigma, \tau, \phi_1, \phi_2$ *in* $\mathbf{V}^*$ *and an integer* $k \geqslant 0$ *such that* $\phi_1 \rightarrow \delta\gamma\phi_1$ *and* $\phi_2 \rightarrow \phi_2\gamma\delta$ *are in* $\mathbf{R}$, $\sigma\phi_2(\gamma\delta)^{k+1}\gamma\phi_1\tau$ *and* $\sigma\phi_2(\gamma\delta)^k\gamma\phi_1\tau$ *are sentential forms.*

Proof. Suppose that the grammar is not resolvable, i.e., that there exist two different arguments, say $(\sigma_1, \tau_1, \phi_1 \rightarrow \psi_1)$ and $(\sigma_2, \tau_2, \phi_2 \rightarrow \psi_2)$ for which the function F defined in Section 3 has the same value. Then

$$\sigma_1\phi_1\tau_1 = \sigma_2\phi_2\tau_2, \tag{1}$$

$$\sigma_1\psi_1\tau_1 = \sigma_2\psi_2\tau_2, \tag{2}$$

$$(\sigma_1 \neq \sigma_2) \qquad \text{or} \qquad (\tau_1 \neq \tau_2). \tag{3}$$

In order to study the relations (1)–(3) in more detail, four cases are distinguished[5]

case 1:	$\mid \sigma_1 \mid \geqslant \mid \sigma_2 \mid$	and	$\mid \tau_1 \mid \geqslant \mid \tau_2 \mid$,
case 2:	$\mid \sigma_1 \mid \geqslant \mid \sigma_2 \mid$	and	$\mid \tau_1 \mid \leqslant \mid \tau_2 \mid$,
case 3:	$\mid \sigma_1 \mid \leqslant \mid \sigma_2 \mid$	and	$\mid \tau_1 \mid \geqslant \mid \tau_2 \mid$,
case 4:	$\mid \sigma_1 \mid \leqslant \mid \sigma_2 \mid$	and	$\mid \tau_1 \mid \leqslant \mid \tau_2 \mid$.

As cases 3 and 4 are turned into cases 2 and 1, respectively, by permutation of the indices, only these latter cases are to be taken into consideration.

Case 1. $\mid \sigma_1 \mid \geqslant \mid \sigma_2 \mid$ and $\mid \tau_1 \mid \geqslant \mid \tau_2 \mid$.

The relations (1) and (2) imply the existence of strings γ and δ in $\mathbf{V}^*$ such that

$$\sigma_1 = \sigma_2 \gamma$$

and

$$\tau_1 = \delta \tau_2 .$$

Substitution into (1), (2) and (3) yields

$$\phi_2 = \gamma \phi_1 \delta,$$
$$\psi_2 = \gamma \psi_1 \delta,$$
$$\gamma \delta \neq \epsilon.$$

Case 2. $\mid \sigma_1 \mid \geqslant \mid \sigma_2 \mid$ and $\mid \tau_1 \mid \leqslant \mid \tau_2 \mid$.

The relations (1) and (2) now imply the existence of strings γ and δ in $\mathbf{V}^*$ such that

$$\sigma_1 = \sigma_2 \gamma \tag{4}$$

and

$$\tau_2 = \delta \tau_1 . \tag{5}$$

Substitution into (1), (2) and (3) yields

$$\gamma \phi_1 = \phi_2 \delta, \tag{1'}$$
$$\gamma \psi_1 = \psi_2 \delta, \tag{2'}$$
$$\gamma \delta \neq \epsilon. \tag{3'}$$

Now, (1′) implies

$$\mid \gamma \mid + \mid \phi_1 \mid = \mid \phi_2 \mid + \mid \delta \mid.$$

Hence it is sufficient to distinguish between the two cases labeled 2.1 and 2.2.

[5] Note that the case $\mid \sigma_1 \mid = \mid \sigma_2 \mid$, for instance, is covered twice; while not being prejudicial to the correctness of the results this method simplifies their expression. A similar remark holds for subsequent case distinctions.

Case 2.1. $|\gamma| \geqslant |\phi_2|$ and $|\delta| \geqslant |\phi_1|$.

The relation (1′) now implies the existence of strings η and θ in $\mathbf{V}^*$ such that

$$\gamma = \phi_2\eta,$$
$$\delta = \theta\phi_1 .$$

Substitution into (1′) and (2′) yields

$$\eta = \theta,$$
$$\phi_2\eta\psi_1 = \psi_2\eta\phi_1 . \tag{2''}$$

Now, (2″) implies

$$|\phi_2| + |\psi_1| = |\phi_1| + |\psi_2|.$$

Hence it is sufficient to distinguish between the cases labeled 2.1.1 and 2.1.2.

Case 2.1.1. $|\phi_2| \geqslant |\psi_2|$ and $|\phi_1| \geqslant |\psi_1|$.

The relation (2″) now implies the existence of strings κ and λ in $\mathbf{V}^*$ such that

$$\phi_2 = \psi_2\kappa, \tag{6}$$
$$\phi_1 = \lambda\psi_1 , \tag{7}$$

Substitution into (2″) yields

$$\kappa\eta = \eta\lambda.$$

But κ and λ are nonempty because $\phi_1 \to \psi_1$ and $\phi_2 \to \psi_2$ are rules and hence $\phi_1 \neq \psi_1$ and $\phi_2 \neq \psi_2$. Lemma 2 is therefore applicable and there exist strings μ, ν in $\mathbf{V}^*$ and an integer $k \geqslant 0$ such that

$$\kappa = \mu\nu,$$
$$\lambda = \nu\mu,$$
$$\eta = (\mu\nu)^k\mu.$$

Substitution into (7), (6), (4) and (5) yields

$$\phi_1 = \nu\mu\psi_1 ,$$
$$\phi_2 = \psi_2\mu\nu,$$
$$\sigma_1 = \sigma_2\psi_2(\mu\nu)^{k+1}\mu,$$
$$\tau_2 = \mu(\nu\mu)^{k+1}\psi_1\tau_1 .$$

Note that the condition (3′) vanishes because ϕ_1 and ϕ_2 are necessarily nonempty.

Case 2.1.2. $|\phi_2| \leqslant |\psi_2|$ and $|\phi_1| \leqslant |\psi_1|$.

There exist κ, λ in $\mathbf{V}^*$ such that

$$\psi_2 = \phi_2\kappa,$$
$$\psi_1 = \lambda\phi_1 ,$$
$$\kappa\eta = \eta\lambda.$$

Again κ and λ are nonempty and we obtain in a way similar to that above that there exist μ, ν in $\mathbf{V}^*$ and $k \geqslant 0$ such that

$$\begin{aligned}
\psi_1 &= \nu\mu\phi_1\,,\\
\psi_2 &= \phi_2\mu\nu,\\
\sigma_1 &= \sigma_2\phi_2(\mu\nu)^k\mu,\\
\tau_2 &= \mu(\nu\mu)^k\phi_1\tau_1\,.
\end{aligned}$$

Again, the condition (3′) vanishes.

Case 2.2. $|\gamma| \leqslant |\phi_2|$ and $|\delta| \leqslant |\phi_1|$.

The relation (1′) now implies the existence of strings μ and θ such that

$$\phi_2 = \gamma\eta, \tag{6'}$$

$$\phi_1 = \theta\delta. \tag{7'}$$

Substitution into (1′) yields $\eta = \theta$.

According to (2′) it is sufficient to distinguish between the cases labeled 2.2.1 and 2.2.2.

Case 2.2.1. $|\gamma| \geqslant |\psi_2|$ and $|\delta| \geqslant |\psi_1|$.

Again there exist κ and λ in $\mathbf{V}^*$ such that

$$\begin{aligned}
\gamma &= \psi_2\kappa,\\
\delta &= \lambda\psi_1\,.
\end{aligned}$$

Substitution into (2′) yields $\kappa = \lambda$. Finally, (7′), (6′), (4), (5) and (3′) lead to

$$\begin{aligned}
\phi_1 &= \eta\lambda\psi_1\,,\\
\phi_2 &= \psi_2\lambda\eta,\\
\sigma_1 &= \sigma_2\psi_2\lambda,\\
\tau_2 &= \lambda\psi_1\tau_1\,,\\
\lambda\psi_1\psi_2 &\neq \epsilon.
\end{aligned}$$

Case 2.2.2. $|\gamma| \leqslant |\psi_2|$ and $|\delta| \leqslant |\psi_1|$.

Again there exist κ, λ in $\mathbf{V}^*$ such that

$$\begin{aligned}
\psi_2 &= \gamma\kappa,\\
\psi_1 &= \lambda\delta,\\
\kappa &= \lambda.
\end{aligned}$$

Together with (7′), (6′), (4), (5) and (3′) this leads to

$$\phi_1 = \eta\delta, \qquad \phi_2 = \gamma\eta,$$
$$\psi_1 = \lambda\delta, \qquad \psi_2 = \gamma\lambda,$$
$$\sigma_1 = \sigma_2\gamma,$$
$$\tau_2 = \delta\tau_1\,,$$
$$\gamma\delta \neq \epsilon.$$

This completes the proof of the theorem: the conditions (i)–(iv) correspond, respectively, with the cases 1, 2.2.2, 2.1.1 together with 2.2.1, and 2.1.2.

The theorem is illustrated by Fig. 2.

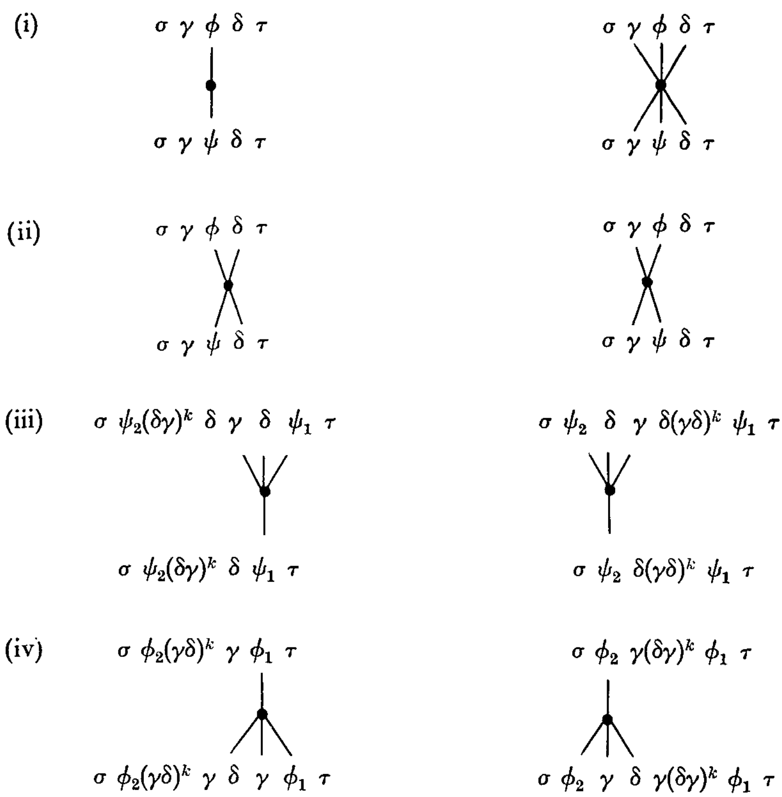

FIG. 2. These eight figures illustrate the four conditions (i)–(iv) of Theorem 1; for instance, the top two figures correspond with the condition (i) and illustrate that $F(\sigma\gamma, \delta\tau, \phi \rightarrow \psi) = F(\sigma, \tau, \gamma\phi\delta \rightarrow \gamma\psi\delta)$.

5.2. COROLLARY 1.[6] *A general phrase-structure grammar* (**V**, **T**, **R**, Z) *is resolvable if for all strings* γ, δ, ϕ, ψ *in* $\mathbf{V}^*$ *none of the following four conditions is satisfied*:

(i) $\gamma\delta \neq \epsilon$,
$\phi \to \psi$ *is in* **R**, *and* $\gamma\phi\delta \to \gamma\psi\delta$ *is in* **R**;

(ii) $\gamma\delta \neq \epsilon$,
$\phi\delta \to \psi\delta$ *is in* **R**, *and* $\gamma\phi \to \gamma\psi$ *is in* **R**;

(iii) $\gamma\delta\phi \to \phi$ *is in* **R** *and* $\psi\delta\gamma \to \psi$ *is in* **R**;

(iv) $\phi \to \delta\gamma\phi$ *is in* **R** *and* $\psi \to \psi\gamma\delta$ *is in* **R**.

The utility of this corollary stems from the unsolvability of the following problem: determine whether or not an arbitrary general phrase-structure grammar is resolvable. This is easily proved as follows: if $G = (\mathbf{V}, \mathbf{T}, \mathbf{R}, Z)$ is a general phrase-structure grammar consider

$$G' = (\mathbf{V} \cup \{Z', A', a'\}, \mathbf{T} \cup \{a'\}, \mathbf{R} \cup \{Z' \to a'A', a'A' \to a'Z, A' \to Z\});$$

G' is resolvable if and only if the language defined by G is empty; the latter problem is known to be unsolvable (see, e.g., [1, p. 230]).

Note that the corollary states the necessary and sufficient condition for the function G: $\mathbf{V}^* \times \mathbf{V}^* \times \mathbf{R} \to \mathbf{V}^* \times \mathbf{V}^*$, being an extension of the function F, to be injective.

6. THE CASE OF CONTEXT-SENSITIVE GRAMMARS

A *context-sensitive grammar* is a general phrase-structure grammar (**V**, **T**, **R**, Z) for which each rule is of the form $\phi A\psi \to \phi\omega\psi$ with A in $\mathbf{V} - \mathbf{T}$, ϕ, ψ in $\mathbf{V}^*$ and ω in $\mathbf{V}^* - \{\epsilon\}$.

THEOREM 2. *A context-sensitive grammar* (**V**, **T**, **R**, Z) *is resolvable if and only if none of the three conditions* (i), (ii) *and* (iv) *of Theorem* 1 *is satisfied.*

COROLLARY 2. *A context-sensitive grammar* (**V**, **T**, **R**, Z) *is resolvable if for all strings* γ, δ, ϕ, ψ *in* $\mathbf{V}^*$ *none of the three conditions* (i), (ii) *and* (iv) *of Corollary* 1 *is satisfied.*

Again, it is not decidable whether or not an arbitrarily given context-sensitive grammar is resolvable. The proof may be given in the same way as for a general phrase-structure grammar.

[6] This corollary is a slightly modified version of [10, Lemma 4.9.1].

7. The Case of Context-Free Grammars

A *context-free grammar* is a general phrase-structure grammar $(\mathbf{V}, \mathbf{T}, \mathbf{R}, Z)$ for which each rule $\phi \to \psi$ has ϕ in $\mathbf{V} - \mathbf{T}$.

Theorem 3. *A context-free grammar* $(\mathbf{V}, \mathbf{T}, \mathbf{R}, Z)$ *is resolvable if and only if none of the following two conditions is satisfied:*

(i) *there exist* σ, τ *in* $\mathbf{V}^*$ *and* A *in* $\mathbf{V} - \mathbf{T}$ *such that* $A \to \epsilon$ *is in* $\mathbf{R}$, $\sigma A A \tau$ *is a sentential form*;

(ii) *there exist* γ, δ, σ, τ *in* $\mathbf{V}^*$, A, B *in* $\mathbf{V} - \mathbf{T}$ *and* $k \geqslant 0$ *such that* $A \to A\gamma\delta$ *and* $B \to \delta\gamma B$ *are in* $\mathbf{R}$, $\sigma A(\gamma\delta)^{k+1}\gamma B\tau$ *is a sentential form.*

Proof. This theorem is a version of Theorem 1 for the grammar being context-free. In fact, the case (i) corresponds with the case (iii) of Theorem 1; it should thereby be noted that the condition "there exist σ, τ in $\mathbf{V}^*$ such that $\sigma A A \tau$ is a sentential form" is equivalent with "there exist σ, τ in $\mathbf{V}^*$ and $k \geqslant 0$ such that $\sigma A^{k+2}\tau$ is a sentential form"; moreover, $\sigma A \tau$ is a sentential form if $\sigma A A \tau$ is. Similarly, the case (ii) corresponds with the case (iv) of Theorem 1; a little thought indicates that the condition "there exists $k \geqslant 0$ such that $\sigma A(\gamma\delta)^{k+1}\gamma B\tau$ is a sentential form" is equivalent with "there exists $k \geqslant 0$ such that $\sigma A(\gamma\delta)^{k}\gamma B\tau$ and $\sigma A(\gamma\delta)^{k+1}\gamma B\tau$ are sentential forms."

This theorem is illustrated in Fig. 3.

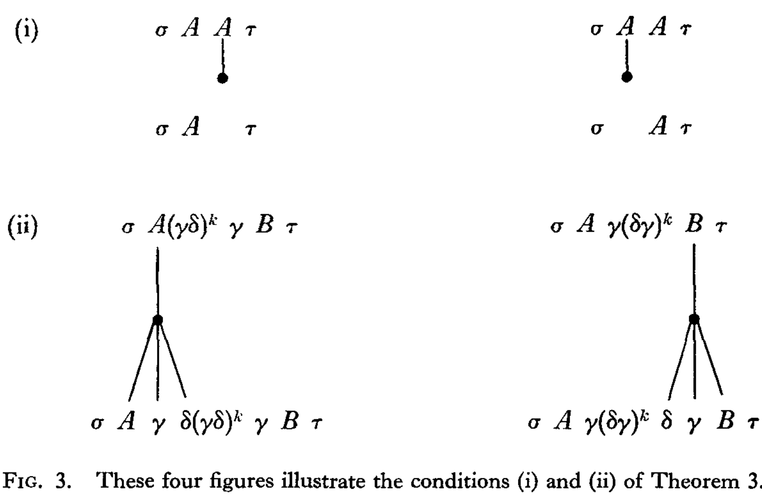

Fig. 3. These four figures illustrate the conditions (i) and (ii) of Theorem 3.

The following problem is easily shown to be solvable: determine whether or not an arbitrary context-free grammar is resolvable. In fact, consider first the case (ii) of Theorem 3 or, more precisely, consider the condition "there exist σ, τ in $\mathbf{V}^*$ and

$k \geqslant 0$ such that $\sigma A(\gamma\delta)^{k+1}\gamma B\tau$ is a sentential form." Call $\mathbf{L}$ the set of all sentential forms of the context-free grammar and put

$$\mathbf{M} = \mathbf{V}^* \cdot \{A\} \cdot \{(\gamma\delta)^{k+1} \mid k \geqslant 0\} \cdot \{\gamma B\} \cdot \mathbf{V}^*. \text{ (see Footnote 7.)}$$

Clearly, the condition is satisfied if and only if the set $\mathbf{L} \cap \mathbf{M}$ is not empty. Now, it is easily seen that $\mathbf{L}$ and $\mathbf{M}$ are a context-free and a regular language, respectively. As the intersection of a context-free and a regular language is context-free (see, e.g., [1, p. 132]), and as moreover the emptiness problem for a context-free language is solvable (see, e.g., [1, p. 230]), it is possible to determine whether or not the condition indicated holds. A similar argument may be given for the case (i) of Theorem 3. As a consequence, there exists a procedure determining whether or not an arbitrary context-free grammar is resolvable.

8. The Case of Leftmost Derivations

A derivation of a sentence of a general phrase-structure grammar, say $\omega_1, \omega_2, \ldots, \omega_n$ is *leftmost* if in addition to the conditions (1) to (5) of Section 2 the conditions

$$|\sigma_i| < |\sigma_{i+1}\phi_{i+1}| \qquad (1 \leqslant i \leqslant n-2) \tag{6}$$

are satisfied. Intuitively these supplementary conditions express that the symbols rewritten during the $(i+1)$-th derivation step are not completely to the left of those rewritten during the i-th derivation step [3, 4, 6, 8]. Note that for context-free grammars the conditions (6) are equivalent with

$$\sigma_{i+1} \text{ is in } \mathbf{T}^* \qquad (1 \leqslant i \leqslant n-2). \tag{6'}$$

The notion of resolvability of a grammar has now to be replaced by a more restricted one. More precisely, if F and $\mathbf{S}$ are those defined in Section 3, let $\mathbf{S}_1 \subseteq \mathbf{S}$ be defined by

$\mathbf{S}_1 = \{(\sigma, \tau, \phi \to \psi) \mid \sigma\phi\tau$ and $\sigma\psi\tau$ are sentential forms which appear as successive strings in a leftmost derivation$\}$

and let $F_1 : \mathbf{S}_1 \to \mathbf{V}^* \times \mathbf{V}^*$ be a restriction of the function F. A general phrase-structure grammar is *leftmost resolvable* when its associated function F_1 is injective. See Fig. 4 for a grammar which is leftmost resolvable without being resolvable.

[7] $\mathbf{S} \cdot \mathbf{T}$ denotes the set product of $\mathbf{S}$ and $\mathbf{T}$, i.e., $\mathbf{S} \cdot \mathbf{T} = \{xy \mid x \text{ is in } \mathbf{S} \text{ and } y \text{ is in } \mathbf{T}\}$.

(a) $G = (\{A, B, C, b, c\},$
$\{b, c\},$
$\{A \to BC, Bc \to bc, B \to b, C \to c\},$
$A)$

(b) A
BC
Bc
bc

(c)

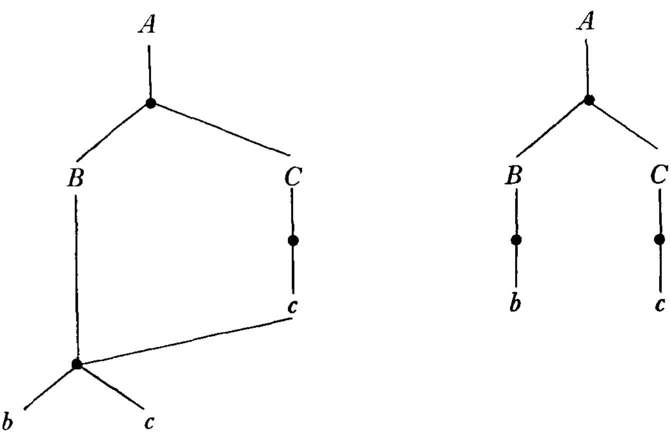

FIG. 4. (a) is a general phrase-structure (more precisely: context-sensitive) grammar; (b) is a derivation (even a leftmost derivation) of the sentence bc; (c) shows two syntactical structures of the sentence bc. Note that the grammar is leftmost resolvable without being resolvable,—as may be illustrated as follows:

$F(\epsilon, \epsilon, Bc \to bc) = F(\epsilon, c, B \to b) = (Bc, bc)$,
$F_1(\epsilon, \epsilon, Bc \to bc) = (Bc, bc)$ but $F_1(\epsilon, c, B \to b)$ is undefined because in a leftmost derivation the rule $B \to b$ must be applied before the rule $C \to c$.

Hence, when the derivation (b) is leftmost it univocally defines the first syntactical structure of (c).

It is easily seen that each context-free grammar is leftmost resolvable: if

$$F_1(\sigma, \tau, \phi \to \psi) = (\omega_1, \omega_2),$$

the arguments σ, τ and $\phi \to \psi$ may be univocally deduced from the strings ω_1 and ω_2 by making use of the following conditions:

$$\sigma \text{ is in } \mathbf{T}^*, \qquad \text{(see (6'))}$$
$$\phi \text{ is in } \mathbf{V} - \mathbf{T},$$
$$\omega_1 = \sigma\phi\tau$$

and

$$\omega_2 = \sigma\psi\tau.$$

This property does not hold for context-sensitive and, *a fortiori*, for general phrase-structure grammars—as is illustrated in Figs. 5 and 6. It is relatively easy to put forward sufficient conditions for F_1 to be injective which properly include the context-

(a) $G = (\{A, B, b, c\},$
$\{b, c\},$
$\{A \to Bc,\ B \to b,\ Bc \to bc\},$
$A)$

(b) A
Bc
bc

(c)

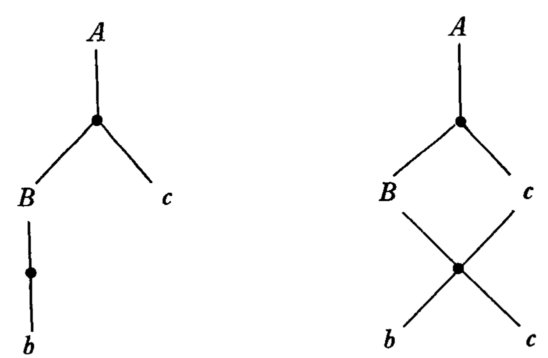

FIG. 5. (a) is a general phrase-structure (more precisely: a context-sensitive) grammar; (b) is a derivation (even a leftmost derivation) and (c) represents two syntactical structures of the sentence *bc*. The grammar is not leftmost resolvable as

$$F_1(\epsilon, c, B \to b) = F_1(\epsilon, \epsilon, Bc \to bc) = (Bc, bc)$$

and even when the derivation sub (b) is known to be leftmost it defines the two syntactical structures sub (c).

free grammar case. Unfortunately, these conditions are too strong to be of real interest and a deeper study of leftmost resolvability immediately leads to uncomputable predicates.

(a) $G = (\{A, a\},$
$\{a\},$
$\{A \to AA,\ AAA \to a\},$
$A)$

(b) A
AA
AAA
a

(c)

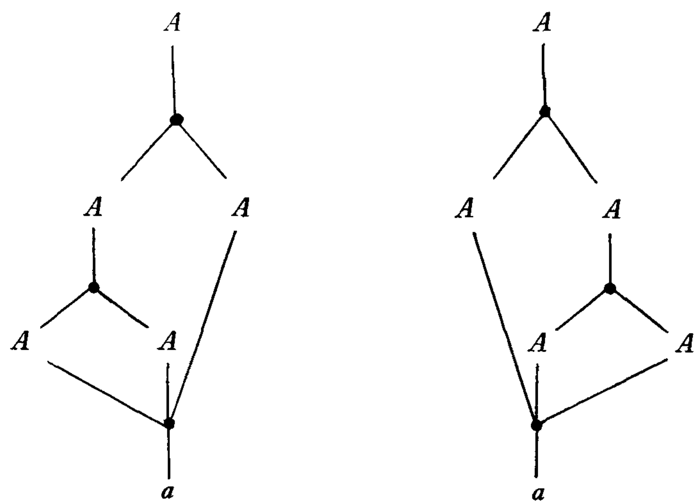

FIG. 6. The interpretation of this figure is the same as that of Fig. 5—except for the fact that the grammar is not context-sensitive.

REFERENCES

1. J. E. HOPCROFT AND J. D. ULLMAN, "Formal Languages and their Relation to Automata," Addison-Wesley, Reading, MA, 1969.
2. J. EICKEL, M. PAUL, F. L. BAUER, AND K. SAMELSON, A syntax controlled generator of formal language processors, *Comm. ACM* **6** (1963), 451–455.
3. T. V. GRIFFITHS, Some remarks on derivations in general rewriting systems, *Information and Control* **12** (1968), 27–54.
4. J. LOECKX, The parsing for general phrase-structure grammars, *Information and Control* **16** (1970), 443–464.
5. H. LANGMAACK AND J. EICKEL, Präzisierung der Begriffe Phrasenstruktur und strukturelle Mehrdeutigkeit in Chomsky-Sprachen, *in* "3. Colloquium über Automatentheorie" (W. Händler *et al.*, Eds.), pp. 263–287, Birkhäuser Verlag, Basel, 1967.
6. D. B. BENSON, The algebra of derivations and a semi-Thue parser, *Proc. 24th ACM Nat. Conf.* (1969), 1–9.

7. W. A. WOODS, Context-sensitive parsing, *Comm. ACM* **13** (1970), 437–445.
8. D. A. WALTERS, Deterministic context-sensitive languages, *Information and Control* **17** (1970), 14–61.
9. K. CULIK, Basic problems in the mathematical theory of languages. *in* "Basic Problems in the Mathematical Theory of Languages" (K. Culik, Ed.), North-Holland Publ. Co, Amsterdam, to appear.
10. J. EICKEL, Vereinfachung der Struktur kombinatorischer Systeme, Habilitationsschrift, Technische Hochschule München, Juni 1968.