Hitotsubashi Journal of Arts and Sciences 28 (1987) 39–43. © The Hitotsubashi Academy

A FORMAL DEDUCTIVE SYSTEM FOR CFG

TAKASHI NAGASHIMA*

We will formulate context-free grammar as a form of logical calculus. As against the familiar generative (rewriting) system of context-free grammar, our system is deductive. That is to say, a word belongs to a language if this fact is deduced within the formal system.

In [H29] Hertz introduced the concept of sentence. A sentence is a formal expression of the form

$$v_1, \ldots, v_n \rightarrow w$$

where the components $v_1, \ldots, v_n$ and w are atomic "elements." In [G32] Gentzen investigated Hertz's theory of sentence systems and improved Hertz's calculus. The concept of sequent in *LJ* and *LK* in [G35] is a generalization of Hertz's sentence. The deductive system proposed here is a calculus of sequents and it is a generalization of Gentzen's calculus in [G32]. The concept of sequent in our system is a generalization of Hertz's sentence in respect of components. With regard to inference rules, we adopt the substitution rule in addition to Gentzen's rules in [G32].

Though in formal language theories a nonterminal is sometimes called a variable, we regard a nonterminal as a unary predicate rather than a variable. For example, a production rule such as

$$\text{Clause} \rightarrow \text{Noun Verb}$$

means "if α is a Noun and β is a Verb then $\alpha\beta$ is a Clause" or formally

$$\text{Noun}[\alpha], \text{Verb}[\beta] \rightarrow \text{Clause}[\alpha\beta].$$

Therefore it seems more natural to regard "Noun," "Verb," "Clause" as predicates than to regard them as variables.

Our system is related to the first order predicate logic. The set Σ^* consisting of all words over an arbitrary finite alphabet Σ is supposed to be the individual domain. A variable ranges over Σ^* but a constant denotes an element of Σ. Any element of the domain Σ^* can be denoted by a string of constants.

The symbols of the system are:

Predicates: $Q, R, \ldots$ (at least one);
Variables: $\alpha, \beta, \ldots$ (countably many);
Constants: $a, b, \ldots$ (at least one);
Auxiliary symbols: [,], $\rightarrow$.

The set of predicates and the finite set of constants may be arbitrarily fixed.

A *term* is a finite (possibly empty) string consisting of variables and constants. The empty term is denoted by ε. A *formula* is an expression of the form $Q[x]$ where Q is any

* The author was partially supported by Grant-in-Aid for Scientific Research (No. 62540145), Ministry of Education, Science and Culture.

predicate and x is any term. An expression is *closed* if it contains no variables. A finite (possibly empty) sequence of formulas is denoted by Γ, Δ or Θ. For any variable α and any term x, the expression obtained by substituting x for α in a formal expression E is denoted as $E(\alpha/x)$. A *sequent* is an expression of the form

$$\mathbf{A}_1, \ldots, \mathbf{A}_n \rightarrow \mathbf{B}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_n$, $\mathbf{B}$ are formulas. The formulas occurring in the left hand side of "$\rightarrow$" are called *antecedent* formulas, and the right hand side formula is called *succedent* formula. If Γ and Δ are equal as sets, two sequents $\Gamma \rightarrow \mathbf{B}$ and $\Delta \rightarrow \mathbf{B}$ are considered identical. A sequent is *trivial* if its succedent formula is equal to one of the antecedent formulas. A sequent is *linear* if it has only one antecedent formula. A sequent is *tautological* if it is trivial and linear. A sequent is *regular* if it has the form $Q[\alpha] \rightarrow R[x\alpha]$ or $\rightarrow R[x]$ where x is a closed term. An *axiom system* is a fixed set of sequents, and its members are called *axioms*. An axiom system is *regular* if all the axioms are regular.

The inference schemata are as follows.

Substitution:

$$\frac{\Gamma \rightarrow \mathbf{A}}{\Gamma(\alpha/x) \rightarrow \mathbf{A}(\alpha/x).}$$

Thinning:

$$\frac{\Gamma \rightarrow \mathbf{A}}{\Delta \rightarrow \mathbf{A},}$$

where every element of Γ occurs in Δ.

Cut:

$$\frac{\Gamma \rightarrow \mathbf{A} \quad \mathbf{A}, \Delta \rightarrow \mathbf{B}}{\Gamma, \Delta \rightarrow \mathbf{B}.}$$

A *proof* is a tree consisting of sequents. The one sequent at the root of the tree is called the *endsequent*, and the sequent at any leaf of the tree is called an *initial sequent*. Each sequent except the initial sequents is a lower sequent of an inference. Each sequent except the endsequent is (one of) the upper sequent(s) of an inference. If the endsequent is S and if every initial sequent is either a tautological sequent or an axiom of K, then it is called a *proof of S from K*. A sequent S is *provable from K* and denoted as $K \vdash S$ if there exists a proof of S from K.

Let $G=(N, \Sigma, P, A)$ be a context-free grammar. The language generated by G is denoted as $L(G)$. Let the set of constants be Σ, and suppose that for any nonterminal $B \in N$ there corresponds a predicate. For the sake of simplicity, we do not distinguish a nonterminal and its corresponding predicate. Any production $p \in P$ can be expressed as

$$p=(B \rightarrow x_0 C_1 x_1 C_2 x_2 \ldots C_n x_n),$$

where $n \geqq 0$, $B \in N$, $C_i \in N$ $(i=1, 2, \ldots, n)$, $x_i \in \Sigma^*$ $(i=0, 1, \ldots, n)$. Let $\alpha_1, \ldots, \alpha_n$ be distinct variables and define p^* as the sequent

$$C_1[\alpha_1], \ldots, C_n[\alpha_n] \rightarrow B[x_0 \alpha_1 x_1 \ldots \alpha_n x_n].$$

The axiom system G^* is defined as the set

$$G^* = \{p^* \mid p \in P\}.$$

If the grammar G is regular (right linear) then the axiom system G^* is regular.

Now we will generalize Gentzen's concept of normal proof and prove a normal form theorem. A proof consisting of exactly one tautological sequent is *normal.* A proof consisting of exactly one thinning with a tautological upper sequent is *normal.* A proof of a nontrivial sequent is *normal* if it satisfies the following conditions:

(1) No trivial sequent occurs.

(2) Any cut occurs above neither a substitution nor the left upper sequent of a cut.

(3) If any thinning occurs, it is the lowest inference.

The following is a generalization of Theorem III in [G32]. Though Gentzen proved it in another way, any proof can be transformed into a normal proof by a finitary procedure (cf. the remark after Theorem III).

THEOREM 1. *For any axiom system K and any sequent S, if $K \vdash S$ then there exists a normal proof of S from K.*

PROOF. It is clear that any trivial sequent has a normal proof, It is easy to show that a proof of a nontrivial sequent can be transformed into a normal proof by permuting inferences successively. For instance, if a part of given proof runs:

$$\cfrac{\cfrac{\Gamma \to A \quad A, \Delta \to B}{\Gamma, \Delta \to B} \qquad B, \Theta \to C}{\Gamma, \Delta, \Theta \to C,}$$

then this part is transformed into:

$$\cfrac{\Gamma \to A \qquad \cfrac{A, \Delta \to B \quad B, \Theta \to C}{A, \Delta, \Theta \to C}}{\Gamma, \Delta, \Theta \to C.}$$

□

The *principal portion* of a normal proof consists of the endsequent, the upper sequent of every thinning and the right upper sequent of every cut.

LEMMA. *Let $G = (N, \Sigma, P, A)$ be a context-free grammar, $B \in N$ and $x \in (N \cup \Sigma)^*$. If $B \Longrightarrow {}^*x$ and x is decomposed as*

$$x = x_0 C_1 x_1 C_2 x_2 \ldots C_n x_n \quad (C_i \in N,\ x_i \in \Sigma^*),$$

then

$$G^* \vdash C_1[\alpha_1], \ldots, C_n[\alpha_n] \to B[x_0 \alpha_1 x_1 \ldots \alpha_n x_n]$$

for distinct variables $\alpha_1, \ldots, \alpha_n$.

PROOF. By induction on the length of a derivation $B \Longrightarrow {}^*x$.

Basis: $x = B$, $G^* \vdash B[\alpha] \to B[\alpha]$.

Inductive step: Let the last step of derivation be $uCw \Longrightarrow uvw$, where $(C \to v)$ is a production, $x = uvw$ and

$$u = u_0 D_1 u_1 \ldots,$$
$$v = v_0 E_1 v_1 \ldots,$$
$$w = w_0 F_1 w_1 \ldots$$

(u_i, v_i, $w_i \in \Sigma^*$, D_i, E_i, $F_i \in N$). By induction hypothesis, the sequent

$$D_1[\alpha_1], \ldots, C[\beta], F_1[\gamma_1], \ldots \to B[u_0\alpha_1 \ldots \beta w_0\gamma_1 \ldots]$$

is provable from G^*. The sequent

$$E_1[\delta_1], \ldots \to C[v_0\delta_1 \ldots]$$

is an axiom. From these two sequents, the sequent

$$D_1[\alpha_1], \ldots, E_1[\delta_1], \ldots, F_1[\gamma_1], \ldots \to B[u_0\alpha_1 \ldots v_0\delta_1 \ldots w_0\gamma_1 \ldots]$$

is deducible by an application of substitution and an application of cut. □

THEOREM 2. *Let $G=(N, \Sigma, P, A)$ be a context-free grammar. If $x \in L(G)$ then $G^* \vdash \to A[x]$.*

PROOF. It immediately follows from Lemma that for ant $B \in N$ and any $x \in \Sigma^*$, if $B \Longrightarrow {}^*x$ then $G^* \vdash \to B[x]$. □

THEOREM 3. *Let $G=(N, \Sigma, P, A)$ be a context-free grammar. If x is closed and $G^* \vdash \to A[x]$ then $x \in L(G)$.*

PROOF. There exists a normal proof of $\to A[x]$ from G^*. Any sequent S in the principal portion is closed and its succedent is $A[x]$. If S runs as

$$B_1[y_1], B_2[y_2], \ldots, B_n[y_n] \to A[x],$$

then each y_i is a subterm of x, and the subterms $y_1, y_2, \ldots, y_n$ do not overlap in x. We define $\psi(S) \in (N \cup \Sigma)^*$ to be the word obtained by replacing each occurrence of y_i by B_i for every $i=1, 2, \ldots, n$ in x. If the antecedent of S is empty then $\psi(S)=x$.

If S_1 and S_2 occur in the principal portion and if S_2 occurs immediately below S_1, then $\psi(S_1) \Longrightarrow \psi(S_2)$ in G. Hence by induction, $A \Longrightarrow {}^*x$ in G. □

As an application, we prove a well-known fact.

THEOREM 4. *A language is acceptable by a nondeterministic finite automaton if and only if it is regular.*

PROOF. Let $M=(Q, \Sigma, \delta, A, F)$ ($F \subset Q$, $\delta \subset Q \times \Sigma \times Q$, $A \in Q$) be a nondeterministic finite automaton. Suppose that for any state $B \in Q$ there corresponds a predicate. For the sake of simplicity, we denote a state and its corresponding predicate by the same letter. The axiom system K is defined as

$$K = \{(C[\alpha] \to B[a\alpha] \mid (B, a, C) \in \delta\} \cup \{\to D[\varepsilon] \mid D \in F\}.$$

There exists a regular grammar $G=(Q, \Sigma, P, A)$ satisfying $K=G^*$. Now we will prove that $x \in L(M)$ if and only if $K \vdash \to A[x]$. By induction on the length of x, it can be shown that if $(B, x, C) \in \delta^*$ then $K \vdash C[\alpha] \to B[x\alpha]$. For any $x \in L(M)$ there exists a $C \in F$ such that $(A,x,C) \in \delta^*$, hence there exists a proof of $C[\alpha] \to A[x\alpha]$ from K. By adding the figure

$$\dfrac{\to C[\varepsilon] \quad \dfrac{C[\alpha] \to A[x\alpha]}{C[\varepsilon] \to A[x]}}{\to A[x]}$$

to this proof, we obtain a proof of $\to A[x]$ from K.

For the proof of the converse, suppose $K\vdash\rightarrow A[x]$ and consider the principal portion of a normal proof. Any sequent in the principal portion except the endsequent has the form $B[y]\rightarrow A[xy]$ and it can be shown that $(A,x,B)\in\delta^*$. The lowermost part of the proof has the form

$$\frac{\rightarrow C[\varepsilon] \quad C[\varepsilon]\rightarrow A[x]}{\rightarrow A[x]}$$

for some $C\in F$. It follows that $x\in L(M)$. Hence $L(M)$ is regular.

Conversely, for any regular language L there exists a regular axiom system K, and an automaton M such that $L(M)=L$ can be constructed from K. □

HITOTSUBASHI UNIVERSITY

REFERENCES

[G32] G. Gentzen: Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsystemen. Math. Ann. **107** (1932), 329–350.

[G35] G. Gentzen: Untersuchungen über das logische Schliessen. Math. Z. **39** (1935), 176–210, 405–431.

[G69] G. Gentzen (ed. M. E. Szabo): The collected papers of Gerhard Gentzen. North-Holland, 1969.

[H29] P. Hertz: Über Axiomensysteme für beliebige Satzsysteme. Math. Ann. **101** (1929), 457–514.