Economical Transformations of Phrase-Structure Grammars to Scattered Context Grammars

Alexander Meduna*

Abstract

This paper presents a transformation that converts any phrase-structure grammar, H, in Penttonen normal form to an equivalent scattered context grammar whose size differs from the size of H quite insignificantly; specifically, G has only five more nonterminals, four more context-dependent productions, and one more context-free production than H. An analogical result holds for Kuroda normal form, too.

1 Introduction

Transformations that convert grammars of one type to equivalent grammars of another type are central to the formal language theory. Initially, this theory designed these transformations regardless of the size of the output grammars. Because this size was usually enormously greater than the size of the input grammars, transformations of this kind were of no use in practice. Therefore, at present, the formal language theory modifies these transformations to produce the equivalent output grammars as small as possible (see [2], [3], [5], [6], [7], and Chapter 4 in [1], including references therein).

Following this line of the language theory, the present paper explains how to transform any phrase-structure grammar to an equivalent scattered context grammar whose size differs from the size of the input phrase-structure grammar quite insignificantly. More precisely, it converts any phrase-structure grammar, H, in Penttonen normal form to an equivalent scattered context grammar, G, so G has only five more nonterminals, four more context-dependent productions, and one more context-free production than H. Then, this paper states an analogical result in terms of Kuroda normal form.

2 Definitions

This paper assumes that the reader is familiar with the language theory (see Chapter 0 in [1]).

*Computing Center at Technical University of Brno, Udolni 19, Brno 60200, Czech Republic

Basic Notions

For a set, Q, *card*(Q) denotes the cardinality of Q. Set

$$\mathbf{N} = \{1, 2, \ldots\} \text{ and } \mathbf{I} = \{0, 1, 2, \ldots\}.$$

Let V be an alphabet. V^* represents the free monoid generated by V under the operation of concatenation. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$; algebraically, V^+ is the free semigroup generated by V under the operation of concatenaton.

For $w \in V^*, |w|$ denotes the lenght of w. Set

subword$(w) = \{x : x \in V^*$ and x is a subword of $w\}$;
prefix$(w) = \{x : x$ is a prefix $w\}$;
suffix$(w) = \{x : x$ is a suffix $w\}$;
alph$(w) =$ *subword*$(w) \cap V$.

For $a \in V$ and $w \in V^*$; *occur*(a, w) denotes the number of occurrences of a's in w.

Grammars

A *phrase-structure grammar* is a quadruple

$$G = (N, P, S, T)$$

where N and T are alphabets such that $N \cap T = \emptyset$. Symbols in N are referred to as *nonterminals* while symbols in T are *terminals*. N contains S - the *start symbol* of G. P is a finite set of productions of the form

$$x \longrightarrow y$$

where $x, y \in V^*$ so *alph*$(x) \cap N \neq \emptyset$. If $x \longrightarrow y \in P$ and $u_i \in (N \cup T)^*$ for $i = 1, 2$, then

$$u_1 x u_2 \Longrightarrow u_1 y u_2;$$

whenever this paper needs to specify the subword, x, rewritten during $u_1 x u_2 \Longrightarrow u_1 y u_2$, it underlines it as

$$u_1 \underline{x} u_2 \Longrightarrow u_1 y u_2.$$

Observe that $\Longrightarrow$ represents a relation on $(N \cup T)^*$. Let $\Longrightarrow^m$ denote the m-fold product of $\Longrightarrow$, where $m \in \mathbf{I}$. Furthermore, $\Longrightarrow^+$ and $\Longrightarrow^*$ denote the transitive closure of $\Longrightarrow$ and the transitive and reflexive closure of $\Longrightarrow$, respectively. The *language generated by* $G, L(G)$, is defined as

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

Let $G = (N, P, S, T)$ be a phrase-structure grammar. G is in *Penttonen normal form* if P has only these two kinds of productions

$$AB \longrightarrow AC \text{ where } A, B, C \in N, \text{ and}$$

$$A \longrightarrow x \text{ where } A \in N \text{ and } x \in NN \cup T \cup \{\varepsilon\}.$$

A *scattered context grammar* is a quadruple

$$G = (N, P, S, T)$$

where N, T, S have the same meaning as in a phrase-structure grammar, and P is a finite set of productions of the form

$$(A_1, A_2, \ldots, A_n) \longrightarrow (x_1, x_2, \ldots, x_n),$$

where $n \in \mathbf{N}$, and for all $i = 1, 2, \ldots, n, A_i \in N$ and $x_i \in (N \cup T)^*$.

Let $G = (N, P, S, T)$ be a scattered context grammar, and let $v, w \in (N \cup T)^*$. If for some $n \in \mathbf{N}$

A. $(A_1, A_2, \ldots, A_n) \longrightarrow (x_1, x_2, \ldots, x_n) \in P$, and

B. $v = u_1A_1u_2A_2 \ldots u_nA_nu_{n+1}$ and $w = u_1x_1u_2x_2 \ldots u_nx_nu_{n+1}$ with $u_i \in (N \cup T)^*$ for $i = 1, 2, \ldots, n+1$,

then v *directly derives* w in G, symbolically written as

$$v \Longrightarrow w.$$

Express $v \Longrightarrow w$ as $u_1A_1u_2A_2 \ldots u_nA_nu_{n+1} \Longrightarrow u_1x_1u_2x_2 \ldots u_nx_nu_{n+1}$.

Whenever this paper needs to specify the nonterminals, A_1 through A_n, rewritten during this direct derivation, it underlines them as

$$u_1\underline{A_1}u_2\underline{A_2} \ldots u_n\underline{A_n}u_{n+1} \Longrightarrow u_1x_1u_2x_2 \ldots u_nx_nu_{n+1}.$$

For $m \in \mathbf{I}, \Longrightarrow^m$ denotes the m-fold product of $\Longrightarrow$. Furthermore, $\Longrightarrow^+$ and $\Longrightarrow^*$ denote the transitive closure of $\Longrightarrow$ and the transitive and reflexive closure of $\Longrightarrow$, respectively. The *language generated by* $G, L(G)$, is defined as

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

Recall that phrase-structure grammars, phrase-structure grammars in Penttonen normal form, and scattered context grammars have the same generative power. Indeed, they all characterize the family of recursively enumerable languages (see [1], [4], and [7]).

Context-Dependent and Context-Free Productions

Let G be a grammar, and let P be $G's$ set of production. In this paper, we separate P into two disjoint subsets - the set of context-free productions, *ContextFree*(P), and the set of context-dependent productions, *ContextDependent*(P). A production, $p \in P$, belongs to *ContextFree*(P) if and only if the left-hand side of p consists of one nonterminal; otherwise, p belongs to *ContextDependent*(P).

Specifically, if $G = (N, P, S, T)$ is a phrase-structure grammar, then

$ContextFree(P) = \{A \longrightarrow x : A \longrightarrow x \in P \text{ and } A \in N\}$, and
$ContextDependent(P) = P - ContextFree(P)$.

If $G = (N, P, S, T)$ is a scattered context grammar, then

$ContextFree(P) = \{(A) \longrightarrow (x) : (A) \longrightarrow (x) \in P\}$, and
$ContextDependent(P) = P - ContextFree(P)$.

Equivalently, if $G = (N, P, S, T)$ is a scattered context grammar, then

$(A_1, \ldots, A_n) \longrightarrow (x_1, \ldots, x_n) \in ContextDependent(P)$ if and only if $n \geq 2$;
otherwise, $(A_1, \ldots, A_n) \longrightarrow (x_1, \ldots, x_n) \in ContextFree(P)$.

3 Results

This section demonstrates that for every phrase-structure grammar, $H = (N', P', S', T)$, in Penttonen normal form, there exists an equivalent scattered context grammar, $G = (N, P, S, T)$, that satisfies

A. $L(G) = L(H)$;

B. $card\,(N) \leq card(N') + 5$;

C. $card(ContextDependent\,(P)) = card(ContextDependent(P')) + 4$;

D. $card(ContextFree(P)) = card(ContextFree(P')) + 1$.

Theorem 1 *Let $H = (M, R, S, T)$ be a phrase-structure grammar in Penttonen normal form. Then, there exists a scattered context grammar, $G = (N, P, E, T)$, that satisfies*

A. $L(G) = L(H)$;

B. $card(M) = card(N) + 5$;

C. $card(ContextDependent(P)) = card(ContextDependent(R)) + 4$;

D. $card(ContextFree(P)) = card(ContextFree(R)) + 1$.

Proof.: Let

$$H = (M, R, S, T)$$

be a phrase-structure grammar in Penttonen normal form, where M denotes $H's$ alphabet of nonterminals, R denotes its set of productions, S is its start symbol, and T denotes its alphabet of terminals. Without loss of generality, assume that

$$\{E, F, [,], \$\} \cap M = \emptyset.$$

In the following, we describe how to construct a scattered context grammar, G, such that $L(G) = L(H)$ and G satisfies the conditions of Theorem 1.
Define the scattered context grammar

$$G = (N, P, E, T),$$

where
$N = \{E, F, [,], \$\} \cup M$
and

$$\begin{array}{rcl} P & = & \{E \longrightarrow F[]F[F]S[], \\ & & (F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F), \\ & & (F, F, \$, F) \longrightarrow (F, F, \varepsilon, F), \\ & & (F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[), \\ & & (F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)\} \\ & \cup & \{(A, B) \longrightarrow (A[,]C) : AB \longrightarrow AC \in R \text{ with } A, B, C \in M\} \\ & \cup & \{(A) \longrightarrow (x) : A \longrightarrow x \in R, A \in M, x \in MM\} \\ & \cup & \{(A) \longrightarrow (\$a) : A \longrightarrow x \in R, A \in M, a \in T \cup \{\varepsilon\}\} \end{array}$$

Observe that

$card(M) = card(N) + 5;$
$card(ContextDependent(P)) = card(ContextDependent(R)) + 4;$
$card(ContextFree(P)) = card(ContextFree\ (R)) + 1.$

The proof that $L(G) = L(H)$ is based on the following.
By productions in

$$\begin{array}{rl} & \{(A, B) \longrightarrow (A[,]C) : AB \longrightarrow AC \in R \text{ with } A, B, C \in M\} \\ \cup & \{(A) \longrightarrow (x) : A \longrightarrow x \in R, A \in M, x \in MM\} \\ \cup & \{(A) \longrightarrow (\$a) : A \longrightarrow x \in R, A \in M, a \in T \cup \{\varepsilon\}\}, \end{array}$$

G simulates $H's$ derivations. By productions in

$$\begin{array}{l} \{E \longrightarrow F[]F[F]S[], \\ (F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F), \\ (F, F, \$, F) \longrightarrow (F, F, \varepsilon, F), \\ (F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[), \\ (F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)\}, \end{array}$$

G verifies that the simulation was performed properly.
Next, this proof establishes several claims to demonstrate $L(G) = L(H)$ in a rigorous way.

Claim 1 Let

$$E \Longrightarrow^{+} u \Longrightarrow^{+} t$$

in G, where $u \in (N \cup T)^*$ and $t \in T^*$. Then

$$u = wFx,$$

where $w \in T^*$ and $x \in ((N - \{E\}) \cup T)^*$ with $occur(F, v) = 2$.

Proof. of Claim 1: Consider any derivation,

$$E \Longrightarrow^{+} u \Longrightarrow^{+} t$$

in G, where $u \in (N \cup T)^*$ and $t \in T^*$. Examine P to see that

$$occur(E, u) = 0 \text{ and } occur(F, u) = 3.$$

Express u as

$$u = wFv,$$

where $w, v \in (N \cup T)^*$ and $occur(F, v) = 2$. Notice that

$$alph(w) \cap \{E, F\} = \emptyset.$$

By contradiction, prove that

$$alph(w) \cap (N - \{E, F\}) = \emptyset.$$

Assume that $alph(w) \cap (N - \{E, F\}) \neq \emptyset$. Consider

$(F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F)$,
$(F, F, \$, F) \longrightarrow (F, F, \varepsilon, F)$,
$(F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[)$, and
$(F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)$.

As $occur(F, v) = 2$, none of these productions can rewrite any symbol in w. Because $alph(w) \cap (N - \{E, F\}) \neq \emptyset$, for every y such that $wFv \Longrightarrow^{+} y$,

$$alph(y) \cap (N - \{E, F\}) \neq \emptyset,$$

which contradicts $wFv \Longrightarrow^{+} t$ with $t \in T^*$. Thus,

$$alph(w) \cap (N - \{E, F\}) = \emptyset.$$

Consequently,

$$u = wFv,$$

where $w \in T^*$ and $v \in ((N - \{E\}) \cup T)^*$ with $occur(F, v) = 2$.
Therefore, Claim 1 holds.

□

Define the morphism, α, from $((N - \{E\}) \cup T)^*$ to $(\{[,], F, \#\} \cup T)^*$ as

$\alpha(Y) = Y$ for all $Y \in (\{[,], F\} \cup T)$, and
$\alpha(X) = \#$ for all $X \in (\{\$\} \cup M)$.

Claim 2 Let

$$E \Longrightarrow^+ wFvFxFy \Longrightarrow^+ u$$

in G, where $w, u \in T^*$, and $v, x, y \in ((N - \{E, F\}) \cup T)^*$. Then,

$$v \in \{[,]\}^*.$$

Proof of Claim 2: Let

$$E \Longrightarrow^+ wFvFxFy \Longrightarrow^+ u$$

in G, where $w, u \in T^*$, and $v, x, y \in ((N - \{E\}) \cup T)^*$. By contradiction we next prove that

$$\{\#\}^+ \cap \mathit{subword}(\alpha(v)) = \emptyset.$$

Assume that

$$\{\#\}^+ \cap \mathit{subword}(\alpha(v)) \neq \emptyset.$$

Examine

$(F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F)$,
$(F, F, \$, F) \longrightarrow (F, F, \varepsilon, F)$,
$(F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[)$, and
$(F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)$.

The form of these productions and $\{\#\}^+ \cap$ $\mathit{subword}(\alpha(v)) \neq \emptyset$ imply that every word that G derives from $wFvFxFy$ contains $\#$. Specifically,

$$\{\#\}^+ \cap \mathit{subword}(\alpha(u)) \neq \emptyset,$$

which contradicts $u \in T^*$. Hence,

$$\{\#\}^+ \cap \mathit{subword}(\alpha(v)) = \emptyset.$$

Thus,

$$\$ \notin \mathit{alph}(v).$$

Consider

$$\{(A) \longrightarrow (\$a) : A \longrightarrow x \in R, A \in M, a \in T \cup \{\varepsilon\}\}.$$

Observe that this set includes all productions containing symbols from $T \cup \{\$\}$. Therefore, as $\$ \notin \mathit{alph}(v)$,

$$T \cap \mathit{alph}(v) = \emptyset.$$

Consequently,

$$\alpha(v) \in \{[,]\}^*.$$

□

Claim 3 Let

$$E \Longrightarrow^{+} y$$

in G, where $y \in ((N - \{E\}) \cup T)^{*}$. Then,

$$][\notin subword(y).$$

Proof of Claim 3: Let

$$E \Longrightarrow^{+} y$$

in G, where $y \in ((N - \{E\}) \cup T)^{*}$. All productions containing [or] are included in this set

$$\begin{array}{ll} & \{E \longrightarrow F[]F[F]S[], \\ & (F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F), \\ & (F, F, \$, F) \longrightarrow (F, F, \varepsilon, F), \\ & (F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[), \\ & (F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)\} \\ \cup & \{(A, B) \longrightarrow (A[,]C) : AB \longrightarrow AC \in R \text{ with } A, B, C \in M\}. \end{array}$$

By $E \longrightarrow F[]F[F]S[], G$ generates $F[]F[F]S[]$. Notice that

$$][\notin subword(F[]F[F]S[]).$$

By using, the productions in

$$\{(A, B) \longrightarrow (A[,]C) : AB \longrightarrow AC \in R \text{ with } A, B, C \in M\},$$

G can generate no word containing][. Finally, consider the other four productions:

$$\begin{array}{l} (F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F), \\ (F, F, \$, F) \longrightarrow (F, F, \varepsilon, F), \\ (F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[), \text{ and} \\ (F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)\}. \end{array}$$

If the current sentential form does not contain][, then every word directly derived from this sentential form by using any of these productions does not contain][either. Thus,

$$][\notin subword(y).$$

A formal version of this proof is left to the reader.

□

Claim 4 Let

$$E \Longrightarrow^{+} wFvFxFy \Longrightarrow^{+} u$$

in G, where $w, u \in T^{*}; v \in \{[,]\}^{*}$; and $x, y \in ((N - \{E, F\}) \cup T)^{*}$. Then,

$$v \in \{[\}^{*}\{]\}^{*}.$$

Proof of Claim 4: This claim follows from Claims 2 and 3.

$\square$

Claim 5 Let

$$E \Longrightarrow^{+} wFvFxFy \Longrightarrow^{+} u$$

in G, where $w, u \in T^*; v \in \{[\}^*\{]\}^*$; and $x, y \in ((N - \{E, F\}) \cup T)^*$. Then,

$$v \in \{[^i]^i : i \in \mathbf{N}\}.$$

Proof of Claim 5: Consider

$$E \Longrightarrow^{+} wFvFxFy \Longrightarrow^{+} u$$

in G, where $w, u \in T^*; v \in \{[\}^*\{]\}^*$; and $x, y \in ((N - \{E, F\}) \cup T)^*$.

By contradiction, we next demonstrate

$$|v| \geq 1.$$

Assume that

$$|v| = 0.$$

Examine P to see that at this point, $wFvFxFy$ derives no word over T, which contradicts $wFvFxFy \Longrightarrow^{+} u$ with $u \in T^*$. Thus, $|v| \geq 1$.

Let $i, j \in \mathbf{I}$ such that $i < j$ and $|v| = i + j$. By contradiction, we now prove

$$[^i]^j \neq \alpha(v).$$

Assume that

$$[^i]^j = \alpha(v).$$

Examine P to see that under this assumption,

$$] \in alph(u).$$

Thus,

$$u \notin T^*,$$

which contradicts $u \in T^*$. Hence,

$$[^i]^j \neq \alpha(v).$$

Analogously, prove $[^i]^j \neq \alpha(v)$, where $i, j \in \mathbf{I}$ so iFj and $|v| = i + j$.

Thus, for any $i, j \in \mathbf{I}$ such that $i \neq j$,

$$[^j]^i \neq \alpha(v).$$

Therefore, Claim 5 holds.

$\square$

Claim 6 Let

$$E \Longrightarrow^+ wFvFxFy \Longrightarrow^+ u$$

in G, where $w, u \in T^*; v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}$; and $x, y \in ((N - \{E, F\}) \cup T)^*$. Then,

$$x \in (\{\$\} \cup T \cup M)^*.$$

Proof of Claim 6: In G, consider any derivation that has this form

$$E \Longrightarrow^+ wFvFxFy \Longrightarrow^+ u$$

in G, where $w, u \in T^*; v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}$; and $x, y \in ((N - \{E, F\}) \cup T)^*$. By contradiction, demonstrate

$$\{[,]\} \cap alph(x) = \emptyset.$$

Assume that

$$\{[,]\} \cap alph(x) \neq \emptyset.$$

Examine P; specifically,

$(F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[)$, and
$(F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)$.

In this case,

$$\{[,]\} \cap alph(u) \neq \emptyset,$$

which contradicts $u \in T^*$. Hence

$$\{[,]\} \cap alph(x) = \emptyset.$$

Therefore, $x \in ((N - \{E, F, [,]\}) \cup T)^*$, so

$$x \in (\{\$\} \cup T \cup M)^*.$$

Claim 7 Let

$$E \Longrightarrow^+ wFvFxFy \Longrightarrow^+ u$$

in G, where $w, u \in T^*$, and $v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}, x \in (\{\$\} \cup T \cup M)^*, y \in ((N - \{E, F\}) \cup T)^*$. Then,

$$y \in (K \cup K\{[^i]^i : i \in \mathbf{N}\}K)^*\{[]\}$$

with

$$K = ((N - \{E, F,], [\}) \cup T).$$

Proof of Claim 7: Examine P; specifically,

$(F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F)$,
$(F, F, \$, F) \longrightarrow (F, F, \varepsilon, F)$,
$(F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[)$, and

$(F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)\}.$

Based on this examination, observe that this claim follows from Claims 5 and 6 (a formal verification of this observation is left to the reader).

□

Define the morphism, β, from $((N - \{E\}) \cup T)^*$ to $(N - \{[,], \$, E, F\}) \cup T)^*$ as

$\beta(Y) = \varepsilon$ for all $Y \in \{[,], \$, F\}$, and
$\beta(X) = X$ for all $X \in (N - \{[,], \$, F\}) \cup T$.

Claim 8 Let

$$E \Longrightarrow^m w \Longrightarrow^* z$$

in G, where $m \in \{\mathbf{N}, z \in T^*, w \in (N \cup T)^*$. Then,

$$S \Longrightarrow^* \beta(w)$$

in H.

Proof of Claim 8: This claim is established by induction on $m = 0, \ldots$.

Basis:
Let $m = 1$. That is,

$$E \Longrightarrow F[]F[F]S[] \Longrightarrow^* v$$

in G. Notice that $\beta([]F[F]S[]) = S$. As

$$S \Longrightarrow^0 S$$

in H, the basis holds.

Induction Hypothesis:
Suppose that there exists $j \in \mathbf{N}$ such that Claim 8 holds for every $m \leq j$.

Induction Step:
Let

$$E \Longrightarrow^{j+1} w \Longrightarrow^* z$$

in G, where $z \in T^*$ and $w \in (N \cup T)^*$. Based on Claims 1 through 7, express this derivation as

$$S \Longrightarrow^j tFvFxFy \Longrightarrow w \Longrightarrow^+ z$$

in G, where $z, t \in T^*$, and $v \in \{[]\}^i\{[]\}^i$ for some $i \in \mathbf{N}, x \in (\{\$\} \cup T \cup M)^*$, $y \in (K \cup K\{([^i]^i : i \in \mathbf{N}\}K)^*\{[]\}$ with $K = ((N - \{E, F,], [\}) \cup T)$. Let p be the production that G uses to make $tFvFxFy \Longrightarrow w$. By the induction hypothesis,

$$S \Longrightarrow^* \beta(tFvFxFy)$$

in H. Next, this proof considers all possible forms of p. Before this consideration notice that p surely differs from $E \longrightarrow F[]F[F]S[]$ because E does not appear in $tFvFxFy$.

1. Assume that p has this form

$$(A, B) \longrightarrow (A[,]C),$$

where $A, B, C \in N$. Because $z, t \in T^*, v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}, x \in (\{\$\} \cup T \cup M)^*$, and $y \in (K \cup K\{[^i]^i : i \in \mathbf{N}\}K)^*\{[]\}$ with $K = ((N - \{E, F,], [\}) \cup T)$, the previous claims imply that

$$u = tFvFxFy'\underline{A}y''\underline{B}y''' \text{ and } w = tFvFxFy'A[y'']Cy''',$$

for some $y' \in \mathit{prefix}(y), y''' \in \mathit{suffix}(y)$, and

$$y'' \in (\{[^i]^i : i \in \mathbf{I}\}.$$

Thus,

$$\beta(y'') = \varepsilon.$$

As $(A, B) \longrightarrow (A[,]C) \in P$,

$$AB \longrightarrow AC \in R.$$

Notice that

$$\beta(tFvFxFy')\underline{A}\beta(y'')\underline{B}\beta(y''') \Longrightarrow \beta(tFvFxFy')A\beta(y'')C\beta(y''')$$

in H. Because $\beta(tFvFxFy')A\beta(y'')C\beta(y''') = \beta(w)$,

$$S \Longrightarrow^* \beta(w)$$

in H.

2. Assume that p has this form

$$(A) \longrightarrow (x).$$

$A \in M, x \in MM$. At this point,

$$u = tFvFxFy'\underline{A}y'' \text{ or } u = tFvFx'\underline{A}x''Fy.$$

Assume that $u = tFvFxFy'\underline{A}y''$. At this point,

$$w = tFvFxFy'xy''.$$

As $(A) \longrightarrow (x) \in P$,

$$A \longrightarrow x \in R.$$

Notice that H makes

$$\beta(tFvFxFy')\underline{A}\beta(y'') \Longrightarrow \beta(tFvFxFy')x\beta(y'')$$

by using $A \longrightarrow x$. Because $\beta(tFvFxFy')x\beta(y'') = \beta(w)$,

$$S \Longrightarrow^* \beta(w)$$

in H. Analogically, prove that $S \Longrightarrow^* \beta(w)$ in the case when $u = tFvFx'\underline{A}x''Fy$.

3. Assume that p has this form

$$(A) \longrightarrow (\$a),$$

where $A \in M$ and $a \in T \cup \{\varepsilon\}$. Observe that

either $x = x'Ax''$ so $u = tFvFx'\underline{A}x''Fy$
where $x' \in prefix(x)$ and $x'' \in suffix(x)$

or $y = y'\underline{A}y''$ so $u = tFvFxFy'\underline{A}y''$
where $y' \in prefix(y)$ and $y'' \in suffix(y)$.

Assume that $u = tFvFx'\underline{A}x''Fy$. At this point,

$$w = tFvFx'\$ax''Fy.$$

As $(A) \longrightarrow (\$a) \in P$,

$$A \longrightarrow a \in R.$$

Notice that H makes

$$\beta(tFvFx'\$)\underline{A}\beta(x''Fy) \Longrightarrow \beta(tFvFx'\$)a\beta(x''Fy).$$

Because $\beta(tFvFx'\$)a\beta(x''Fy) = \beta(w)$,

$$S \Longrightarrow^* \beta(w)$$

in H.
Analogously, establish $S \Longrightarrow^* \beta(w)$ under the assumption that $u = tFvFxFy'\underline{A}y''$.

4. Assume that p is a production from

$\{(F, [,], F, F) \longrightarrow (F, \varepsilon, \varepsilon, F, F),$
$(F, F, \$, F) \longrightarrow (F, F, \varepsilon, F),$
$(F, [,], F, F, [,], [) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, F, [,]F, F[),$
$(F, [,], F, F, [,]) \longrightarrow (\varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \varepsilon)\}.$

Then,

$$\beta(tFvFxFy) = \beta(w)$$

By the induction hypothesis,

$$S \Longrightarrow^* \beta(w)$$

in H.
Thus, Claim 8 holds.

□

Claim 9

$$L(G) \subseteq L(H).$$

Proof of Claim 9: By Claim 6, if

$$E \Longrightarrow^{+} v$$

in G with $v \in T^{*}$, then

$$S \Longrightarrow^{*} v$$

in H. Therefore, Claim 9 holds.

□

Claim 10 Let

$$S \Longrightarrow^{j} u \Longrightarrow^{*} z$$

in H, where $j \in \mathbf{I}, z \in T^{*}$, and $u \in (M \cup T)^{*}$. Then,

$$E \Longrightarrow^{+} wFvFxFy$$

in G, where $w \in T^{*}, v \in \{[]\}^{i}\{]\}^{i}$ for some $i \in \mathbf{N}, x \in (\{\$\} \cup T)^{*}, y \in ((N - \{E, F\}) \cup T)^{*}$, so that

$$u \in \beta(wFvFxFy).$$

Proof of Claim 10: This claim is established by induction on $j = 0, \ldots$.

Basis:
Let $j = 0$; that is,

$$S \Longrightarrow^{0} S \Longrightarrow^{*} z$$

in H. Notice that G makes

$$E \Longrightarrow F[]F[F]S[]$$

by using $E \longrightarrow F[]F[F]S[]$, and $S \in \beta(F[]F[F]S[])$. Thus, the basis holds.

Induction Hypothesis:
Suppose that

$$S \Longrightarrow^{i} u \Longrightarrow^{*} z$$

in H, where $z \in T^{*}$, $u \in (M \cup T)^{*}$, for all $i = 0, \ldots, j$, for some $j \in \mathbf{I}$.

Induction Step:
Let

$$S \Longrightarrow^{j+1} u \Longrightarrow^{*} z$$

in H, where $z \in T^{*}$ and $u \in (M \cup T)^{*}$. Express this derivation as

$$S \Longrightarrow^{j} u \Longrightarrow t \Longrightarrow^{*} z$$

in H, where $t \in (M \cup T)^{*}$ and $u \Longrightarrow t$ is made according to $p \in R$. By the induction hypothesis,

$$E \Longrightarrow^{+} wFvFxFy$$

in G, where $w \in T^{*}, v \in \{[]\}^{i}\{]\}^{i}$ for some $i \in \mathbf{N}, x \in (\{\$\} \cup T)^{*}, y \in ((N - \{E, F\}) \cup T)^{*}$, so $u \in \beta(wFvFxFy)$. Let H make $u \Longrightarrow t$ by using a production, $p \in R$. Next, this proof considers all possible forms of p.

1. Assume that p has this form

$$AB \longrightarrow AC.$$

Express $u \Longrightarrow t$ in H as

$$u'\underline{AB}u'' \Longrightarrow u'ACu'',$$

where $u'ABu'' = u$ and $u'ACu'' = t$. As $w \in T^*, v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}, x \in (\{\$\} \cup T)^*, y \in ((N - \{E,F\}) \cup T)^*, u \in \beta(wFvFxFy)$, and

$$AB \in subword(y).$$

Assume that

$$y = y'A[^k]^k By'',$$

where $u' \in \beta(wFvFxFy'), k \in \mathbf{I}, u'' \in \beta(y'')$.
As $AB \longrightarrow AC \in R$,

$$(A, B) \longrightarrow (A[,]C) \in P.$$

Thus,

$$wFvFxFy'A[^k]^k By'' \Longrightarrow wFvFxFy'A[^{k+1}]^{k+1}Cy''$$

in G. Therefore,

$$E \Longrightarrow^+ wFvFxFy'A[^{k+1}]^{k+1}Cy''$$

in G so

$$u \in \beta(wFvFxFy'A[^{k+1}]^{k+1}Cy'').$$

2. Assume that p has this form

$$A \longrightarrow BC.$$

Because $w \in T^*, v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}, x \in (\{\$\}\cup T)^*, y \in ((N-\{E,F\})\cup T)^*$, and $u \in \beta(wFvFxFy)$,

$$A \in subword(y).$$

Express y as

$$y = y'Ay''.$$

As $A \longrightarrow BC \in R$,

$$(A) \longrightarrow (BC) \in P.$$

Thus,

$$wFvFxFy'Ay'' \Longrightarrow wFvFxFy'BCy''$$

in G. Therefore,

$$E \Longrightarrow^+ wFvFxFy'BCy''$$

in G so

$$u \in \beta(wFvFxFy'BCy'').$$

3. Assume that p has this form

$$A \longrightarrow a.$$

As $w \in T^*, v \in \{[\}^i\{]\}^i$ for some $i \in \mathbf{N}, a \in (\{\varepsilon\} \cup T), x \in (\{\$\} \cup T)^*, y \in ((N - \{E, F\}) \cup T)^*$, and $u \in \beta(wFvFxFy)$, we have

$$A \in subword(y).$$

Express y as

$$y = y'Ay''.$$

As $A \longrightarrow a \in R$,

$$(A) \longrightarrow (\$a) \in P.$$

Thus,

$$wFvFxFy'Ay'' \Longrightarrow wFvFxFy'\$ay''$$

in G. Therefore,

$$E \Longrightarrow^+ wFvFxFy'\$ay''$$

in G so

$$u \in \beta(wFvFxFy'\$ay'').$$

Therefore, Claim 10 holds.

□

Claim 11

$$L(H) \subseteq L(G).$$

Proof of Claim 11:
By Claim 10, if

$$S \Longrightarrow^* v$$

in H, where $v \in T^*$, then

$$S \Longrightarrow^* v$$

in G. Therefore, Claim 11 holds.

□

By Claims 9 and 11,

$$L(G) = L(H).$$

To summarize the proof,

A. $L(G) = L(H)$;

B. $card(M) = card(N) + 5$;

C. $card(ContextDependent(P)) = card(ContextDependent(R)) + 4$;

D. $card(ContextFree(P)) = card(ContextFree(R)) + 1$.

Thus, Theorem 1 holds.

□

In its conclusion, this paper points out that the previous theorem also holds for phrase-structure grammars in Kuroda normal form (see [2]). Recall that a phrase-structure grammar, $G = (N, P, S, T)$, is in *Kuroda normal form* if P has only these two kinds of productions

$$AB \longrightarrow CD \text{ where } A, B, C, D \in N, \text{ and}$$

$$A \longrightarrow x \in R, \text{ where } A \in N \text{ and } x \in NN \cup T \cup \{\varepsilon\}$$

Theorem 2 Let $H = (N', P', S', T)$ be a phrase-structure grammar in Kuroda normal form. Then, there exists a scattered context grammar, $G = (N, P, S, T)$, that satisfies

A. $L(G) = L(H)$;

B. $card(N) \leq card(N') + 5$;

C. $card(ContextDependent(P)) \leq card(ContextDependent(P')) + 4$;

D. $card(ContextFree(P)) \leq card(ContextFree(P')) + 1$.

Proof.: Prove this theorem by analogy with the proof of Theorem 1.

□

Acknowledgement: The author thanks the anonymous referee for useful comments concerning the first version of this paper.

References

[1] Dassow, J. and Paun, G.: Regulated Rewriting in Formal Language Theory. Springer, New York, 1989.

[2] Kuroda, S. Y. : "Classes of Languages and Linear Bounded Automata," Inform. Control 7, 207 - 223, 1964.

[3] Meduna, A. : "Syntactic Complexity of Scattered Context Grammars " Acta Informatica 32, 285 - 298, 1995.

[4] Meduna, A. : "A Trivial Method of Characterizing the Family of Recursively Enumerable Languages by Scattered Context Grammars," EATCS Bulletin 56 (1995), 104 -106.

[5] Meduna, A. : "Four-Nonterminal Scattered Context Grammars Characterize the Family of Recursively Enumerable Languages," International Journal of Computer Mathematics 63, 67-83, 1997.

[6] Paun, Gh. : "Six Nonterminals are Enough for Generating each R.E. Language by a Matrix Grammar," International Journal of Computer Mathematics 15, 23-37, 1993.

[7] Penttonen, M. : "One-Sided and Two-Sided Context in Formal Grammars," Inform. Control 25, 371 - 392, 1974.

Received March, 1997