SRI International

ON SOME FORMAL PROPERTIES OF METARULES

Technical Note 305R

October 1985

By: Hans Uszkoreit, Computer Scientist
Artificial Intelligence Center
Computer Science and Technology Division
and
Center for the Study of Language and Information

This research was supported by the National Science Foundation Grant IST-81035550.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

On Some Formal Properties of Metarules*

Hans Uszkoreit and Stanley Peters

1. Introduction

Grammars contain rules for generating sentences. Metarules are statements about these rules. They are metagrammatical devices that can be used to generate rules of the grammar or to encode certain relations among them such as redundancies in their form.

The linguistic framework of Generalized Phrase Structure Grammar (GPSG) (Gazdar 1982; Gazdar, Pullum, and Sag 1981; Gazdar and Pullum 1982) utilizes metarules in describing natural languages. The rules of a GPSG are context-free (CF) phrase-structure rules. A metarule is an ordered pair of rule templates $\langle A, B \rangle$ (often written $A \Rightarrow B$) that is to be interpreted as follows: if the grammar contains a rule of the form A, it also contains a corresponding rule of the form B. As this interpretation suggests, the set of grammar rules is closed under application of the metarules. It is therefore possible to give an inductive definition of the grammar by listing just a subset of the rules—called the basic rules—together with the list of metarules; the full set of rules is derived by applying the metarules to the basic rules and then, recursively, to the output of all such applications.

In GPSG, metarules are used to describe many of the linguistic phenomena for which transformations have previously been employed. Gazdar and other proponents of GPSG claim that GPSGs are powerful enough to capture all the generalizations about natural languages that were expressed by transformations and, at the same time, sufficiently constrained to generate

*This research was supported by the National Science Foundation Grant IST-8103550. A preliminary report on some of the results contained in this paper was presented at the 1982 LSA meeting in San Diego (Peters and Uszkoreit, 1982). We are grateful to William Marsh, Jane Robinson and Stuart Shieber for comments on an earlier draft of the paper. Due to the extended period of time during which the paper was in preparation, several relevant recent publications are not referred to. Among those is Gazdar et al. (1985) where an extensive account of a new version of GPSG is presented.

only context-free languages. The design of the framework is built on the conjecture that all natural languages are CF.

We neither aim to discuss the theory of GPSG in its entirety nor to restrict our attention to this individual framework. We do not concern ourselves here with most of the mechanisms used by GPSG: derived categories, feature cooccurrence restrictions, ID/LP notation, subcategorization by rule, etc. Instead we are interested in certain formal properties of all grammars that use metarules of the form and in the way described above to close a set of CF grammar rules.[1] Since our definitions are abstract enough to make our results applicable to several such theories, we term the grammars studied in this paper Metarule Phrase Structure or MPS grammars (see the Appendix for definitions).

Our results mainly concern the weak generative capacity of MPS grammars. We show that the power of the formalism is greater than was previously assumed. Unconstrained MPS grammars have Turing machine power, i.e., they are capable of generating the full family of recursively enumerable string sets. Some constrained versions of the MPS grammar formalism still exhibit a certain degree of incommensurate excessive generative power. Finding a strong enough constraint that is both linguistically motivated and descriptively adequate will be a difficult task.

2. The Role of Variables and Phantom Symbols

What provides such a degree of generative power to a formalism that enumerates CF phrase structure rules? Clearly, a finite list of (basic) CF phrase-structure rules will generate only a CF language. How can metarules that merely add more CF rules alter the picture? To find an answer to this question, we begin by taking a closer look at metarules.

We have not yet said anything about the form of the rule templates that occur in

[1]For an investigation of the use of metarules in the framework of Annotated Phrase Structure Grammar, see Konolige (1981).

metarules. A simple template might look exactly like a CF rule. It then matches just that one rule.[2] Most metarules proposed so far for grammars of natural language fragments are more general, in that they use templates that match a larger set of rules. This is achieved by employing variables in metarules. It is helpful to classify the variables that have been used into two categories. The first consists of abbreviatory, or inessential, variables, which range over finite sets of admissible values. Such variables may be useful in permitting the expression of linguistically important generalizations. However, abbreviatory variables can always be eliminated from a grammar, since each metarule containing them can be replaced by the finitely many metarules obtained when these variables are instantiated in all admissible ways. Thus, inessential variables do not affect either the set of strings that can be generated or the sets of tree structures that can be assigned to generated strings by MPS grammars.

The second kind of variables are nonabbreviatory, or essential, variables, which range over all strings of terminal and nonterminal symbols. Rule (1) is an example of a metarule that contains both abbreviatory and essential variables. Here X and Y are essential variables and A is an abbreviatory variable with the range $\{NP, PP\}$ and α is an abbreviatory variable ranging over agreement feature specifications. The rule is supposed to generate VP rules with subject-controlled reflexivized constituents; the variable α over agreement features also determines the appropriate reflexive pronoun.[3]

$$(1)\quad \underset{[\alpha]}{VP} \rightarrow X \; A \; Y \quad \Rightarrow \quad \underset{[\alpha]}{VP} \rightarrow X \; \underset{\substack{[\alpha] \\ [self]}}{A} \; Y$$

Clearly, the use of essential variables can not only create infinite rule sets, but, as has been noted, can also yield grammars that generate non-CF languages. (Gazdar, 1982; Thompson,

---

[2] If the rules that serve as the input for metarules contain complex symbols, and if the symbols of the metarule template are allowed to underspecify these complex symbols by leaving out syntactic features, there might then be several rules that match a simple template.

[3] A similar rule, equipped with the appropriate semantic translation schema, was proposed by Gazdar and Sag (1980).

1982). The designers of GPSG—intending to constrain their framework's potential for unwarranted power—have been concerned about this. But the formal properties of languages that can be generated with the aid of essential variables have not been known. Nor has it been clear how these properties are affected by the number of essential variables employed.

We understand that Aravind Joshi has made the following conjecture: using just one essential variable does not increase generative power beyond CF languages—even if infinitely many rules are derived. This conjecture was apparently accepted in GPSG as an established fact (Gazdar, 1982, footnote 28). A single essential variable came to be considered 'safe'. In this paper, we take a closer look at the impact of essential variables on MPS grammars and demonstrate that Joshi's conjecture is false. (We term MPS grammars with just one essential variable MPS/1 grammars.)

We shall see further that there is an interesting interaction between essential variables and the kind of nonterminals called 'useless symbols' in formal language theory. These are the symbols α for which no derivation $S \overset{*}{\mapsto} \phi\alpha\psi \overset{*}{\mapsto} w$ exists in a grammar for any strings ϕ, ψ and w, with $w \in V_T^*$. The theorem that every nonempty CF language is generated by a CF grammar with no such symbols explains their name (Hopcroft and Ullman, 1979, pp. 88-90). The term 'useless symbol' is misleading when applied to MPS grammars, however, for such symbols are not eliminable in this type of grammar.

To see why this is so, consider any rule-derivation of the form $R_1 \overset{*}{\Rightarrow} R_2 \overset{*}{\Rightarrow} R_3$ where R_1 is a basic rule, R_2 a rule containing a 'useless symbol' α, and R_3 a rule with no 'useless symbols'. Obviously it is conceivable that α may actually not be useless in this case, as it plays a role in the derivation of a useful rule. Indeed, R_3 may be derivable only through intermediate steps, which, like R_2, contain a 'useless symbol'.

Within the framework of Generalized Phrase Structure Grammar, symbols of this kind have been utilized in grammars for natural-language fragments under the name 'phantom categories'. The first phantom category proposed was TVP, for transitive verb phrase. Gazdar

and Kay (1980) offer the following basic rule for introducing verbs like *tell* in *She told him that it would rain.*

(2) $TVP \rightarrow V \ \overline{S}$

The category TVP is not mentioned on the right-hand side of any rule. Therefore, the nonterminal cannot be used in the derivation of any string. Rule (2) contributes to the grammar by serving as the input to metarules. One of these metarules is (3), which derives active VP rules from TVP rules. The same set of TVP rules also matches the input template of a metarule that generates passive VP rules.

(3) $TVP \rightarrow V \ X \quad \Rightarrow \quad VP \rightarrow V \quad NP \quad X$

Thus, 'useless symbols' are not necessarily useless, nor do they have to be nonterminals, as the term 'phantom category' suggests. From now on we shall simply call them 'phantom symbols' and hope that this will be a reasonable compromise between accuracy and mnemonic value. We shall see that these symbols are indeed not always eliminable from MPS grammars (v. Theorem 5).

3. The Weak Generative Capacity of MPS grammars

Before formulating and proving some theorems about MPS grammars, we briefly summarize our findings. They reveal that:

(i) MPS/1 grammars generate all recursively enumerable (r.e.) languages.

Since MPS grammars generate only r.e. languages, it follows that MPS/1 grammars generate exactly the class of r.e. languages.

(ii) MPS grammars without phantom symbols generate languages that, if infinite, are 'arithmetically dense'.

(iii) MPS/1 grammars without phantom symbols generate some nonrecursive languages.

Since many context-sensitive languages, indeed, many indexed languages are not 'arithmetically dense', MPS grammars without phantom symbols generate a class of languages that cannot be compared with either the class of context-sensitive languages or with the class of recursive languages. Of course,

(iv) MPS/1 grammars without phantom symbols generate all context-free languages.

The most elegant counterexample we know to Joshi's conjecture is due to Chris Culy (1982). Because of its simplicity we present this grammar instead of our earlier example. Consider the following MPS/1 grammar, which consists of one basic rule and three metarules:

Basic rule:

$$S \rightarrow abc$$

Metarules:

$$S \rightarrow aX \Rightarrow S \rightarrow Xaa$$
$$S \rightarrow bX \Rightarrow S \rightarrow Xbb$$
$$S \rightarrow cX \Rightarrow S \rightarrow Xcc$$

X is the only essential variable. The sentences of the language are exactly the right-hand sides of the rules in the grammar. Recursive application of the metarules yields the language $\{abc, bcaa, caabb, aabbcc, abbccaa, bbccaaaa, bccaaaabb, ...\}$. The fact that this language is not CF can easily be verified by intersecting it with the regular set $a^*b^*c^*$. The class of CF languages is closed under intersection with regular sets. The result of this intersection, however, is $\{a^{2^n}b^{2^n}c^{2^n} \mid n \geq 0\}$, which, like any other infinite subset of $\{a^nb^nc^n \mid n \geq 0\}$, is not CF.

The classic example, $a^nb^nc^n$ itself, can also be generated by an MPS/1 grammar. Here is one that does the job:

Basic rule:

$$S \rightarrow AabcA$$

Metarules:

$$\begin{aligned}
S \rightarrow AaaX &\Rightarrow S \rightarrow AaXa \\
S \rightarrow AabX &\Rightarrow S \rightarrow AbXaa \\
S \rightarrow AbbX &\Rightarrow S \rightarrow AbXb \\
S \rightarrow AbcX &\Rightarrow S \rightarrow AcXbb \\
S \rightarrow AccX &\Rightarrow S \rightarrow AcXc \\
S \rightarrow AcAX &\Rightarrow S \rightarrow AXccA \\
S \rightarrow AXA &\Rightarrow S \rightarrow X
\end{aligned}$$

This time two nonterminals are used. Since one of them, A, never occurs on the left-hand side of any rule, it is therefore a phantom symbol. The question whether there is an MPS/1 grammar without phantom symbols that generates the same language is currently open.

Before we show that MPS/1 grammars generate all r.e. languages, we should mention the fact that they cannot generate anything outside the class of r.e. sets. (See the Appendix for definitions of terms pertaining to these grammars.)

Remark 1. Every language generated by an MPS grammar is an r.e. set of strings.

Proof: It is easy to show how the set $\overrightarrow{\ }_G$ of all rules of G is recursively enumerated, starting from the finite set of basic rules and then iteratively applying the finitely many metarules of G. Since $\overrightarrow{\ }_G$ is a r.e. relation, it can easily be demonstrated that $\overrightarrow{\ }_G$ is likewise. But then $\overset{*}{\overrightarrow{\ }}_G$ is also r.e., since r.e. relations are closed under the operation of taking ancestrals. But then $\{\phi \mid S \overset{*}{\overrightarrow{\ }}_G \phi\}$ is an r.e. set of strings, since it is definable by existential quantification from an r.e. relation (viz., $\exists \psi[\psi = S \wedge \psi \overset{*}{\overrightarrow{\ }}_G \phi]$). Hence $L(G) = V_T^* \cap \{\phi \mid S \overset{*}{\overrightarrow{\ }}_G \phi\}$ is r.e., since V_T^* is and the class of r.e. sets is closed under intersection. ∎

The following theorem is the main step in proving that every r.e. set is generated by an MPS/1 grammar. It states that, for every r.e. language L over an alphabet Σ, there exists an MPS/1 grammar such that L is the intersection of Σ^* and the set of all strings generable from S by rules of the MPS/1 grammar. The proof is based on a procedure that converts any unrestricted rewriting system G', generating a language L over Σ into an MPS/1 grammar, that generates the union of L with a subset of the complement of Σ^*.

Theorem 2. There is a regular language $R\ (= \Sigma^*)$ such that, for any recursively enumerable language $L \subseteq \Sigma^*$, there is an MPS/1 grammar $G = (V_T, V_N, S, \rightarrow, V_V, \Rightarrow)$ such that $L = R \cap \{\omega \mid S \overset{*}{\underset{G}{\mapsto}} \omega\}$.

Proof: Given Σ and an r.e. language $L \subseteq \Sigma^*$, let $G' = (V'_T, V'_N, S', \rightarrow')$ be an unrestricted rewriting system such that $V'_T = \Sigma$ and $L = \{x \in (V'_T)^* \mid S' \overset{*}{\underset{G'}{\mapsto}} x\}$. Construct G as follows. Let S and A be two new symbols not in $V'_T \cup V'_N$. Give G the single basic rule $S \underset{G}{\rightarrow} AS'A$ and the following finite set of metarules, which make use of the single string variable X (where $V_V = \{X\}$).

(i) $S \rightarrow A\phi X \Rightarrow S \rightarrow A\psi X$ whenever $\phi \rightarrow' \psi$ is a rule of G'

(ii) $S \rightarrow A\alpha X \Rightarrow S \rightarrow AX\alpha$ whenever $\alpha \in V'_T \cup V'_N \cup \{A\}$

(iii) $S \rightarrow AXA \Rightarrow S \rightarrow X$

Note that S is the only symbol of G appearing on the left-hand side of any basic or derived rule. Furthermore, S does not appear on the right-hand side of any rule. Therefore, for any string ω other than S itself, $S \overset{*}{\underset{G}{\mapsto}} \omega$ iff $S \underset{G}{\rightarrow} \omega$ is a rule of G. To establish that $L = \Sigma^* \cap \{\omega \mid S \overset{*}{\underset{G}{\mapsto}} \omega\}$, it suffices to show that, for any $x \in \Sigma^*$, $S \underset{G}{\rightarrow} x$ is a rule of G if and only if $S' \overset{*}{\underset{G'}{\mapsto}} x$. To this end, we now prove a more general statement from which this one follows directly, viz.: $S \underset{G}{\rightarrow} A\omega A$ is a rule of G if and only if $S' \overset{*}{\underset{G'}{\mapsto}} \omega$ for any $\omega \in (V'_T \cup V'_N)^*$. (The final metarule of G and the fact that all other metarules preserve the property, possessed by

G's basic rule, of there being exactly two As on the right-hand side establishes the required connection between this statement and the preceding one.)

If direction: Suppose $S' \overset{*}{\underset{G'}{\Rightarrow}} \omega$. Let $\chi_1, \chi_1', \ldots, \chi_n, \chi_n' \; \phi_1, \psi_1, \ldots, \phi_n, \psi_n$ be such that $S' = \chi_1\phi_1\chi_1' \underset{G'}{\Rightarrow} \chi_1\psi_1\chi_1' = \chi_2\phi_2\chi_2' \underset{G'}{\Rightarrow} \cdots \chi_n\psi_n\chi_n' = \omega$.

Because a rule $S \underset{G}{\rightarrow} A\sigma\tau$ is derivable from the rule $S \underset{G}{\rightarrow} A\tau\sigma$ in G whenever $\sigma, \tau \in (V_T' \cup V_N' \cup \{A\})^*$, $S \underset{G}{\rightarrow} AS'A = S \underset{G}{\rightarrow} A\chi_1\phi_1\chi_1'A \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} A\phi_1\chi_1'A\chi_1 \Rightarrow S \underset{G}{\rightarrow} A\psi_1\chi_1'A\chi_1 \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} A\chi_1\psi_1\chi_1'A = S \underset{G}{\rightarrow} A\chi_2\phi_2\chi_2'A \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} A\phi_2\chi_2'A\chi_2 \Rightarrow \cdots S \underset{G}{\rightarrow} A\psi_n\chi_n'A\chi_n \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} A\chi_n\psi_n\chi_n'A = S \underset{G}{\rightarrow} A\omega A$ is a derivation via the metarules in G of the rule $S \underset{G}{\rightarrow} A\omega A$ from the basic rule $S \underset{G}{\rightarrow} AS'A$.

Only-if direction: Suppose $S \underset{G}{\rightarrow} A\omega A$ is a rule of G, and let $S \underset{G}{\rightarrow} AS'A \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} \omega_1 \Rightarrow S \underset{G}{\rightarrow} \omega_2 \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} \omega_3 \Rightarrow \cdots S \underset{G}{\rightarrow} \omega_{2n} \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} A\omega A$ be a derivation of this rule in G such that the metarule applied at each step $S \underset{G}{\rightarrow} \omega_{2i+1} \Rightarrow S \underset{G}{\rightarrow} \omega_{2(i+1)}$ is $S \rightarrow A\phi X \Rightarrow S \rightarrow A\psi X$ for some rule $\phi \rightarrow' \psi$ of G', and, in addition, where the only metarules needed to generate the relations $S \underset{G}{\rightarrow} \omega_{2i} \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} \omega_{2i+1}$ (including $S \underset{G}{\rightarrow} AS'A \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} \omega_1$ and $S \underset{G}{\rightarrow} \omega_{2n} \overset{*}{\Rightarrow} S \underset{G}{\rightarrow} A\omega A$) are of the form $S \rightarrow A\alpha X \Rightarrow S \rightarrow AX\alpha$, for $\alpha \in V_T' \cup V_N' \cup \{A\}$. Each ω_j must have A as its first symbol and contain exactly one other occurrence of A (no metarule adds or deletes A, since it is not a symbol of G'). Now let $\chi_0, \chi_0', \ldots, \chi_{n-1}, \chi_{n-1}', \phi_0, \psi_0, \ldots, \phi_{n-1}, \psi_{n-1}$ be such that $\omega_{2i+1} = A\phi_i\chi_i'A\chi_i$, $\omega_{2(i+1)} = A\psi_i\chi_i'A\chi_i$ and $\phi_i \rightarrow' \psi_i$ is a rule of G'. Clearly, $\chi_0\phi_0\chi_0' \underset{G'}{\Rightarrow} \chi_0\psi_0\chi_0' = \chi_1\phi_1\chi_1' \underset{G'}{\Rightarrow} \cdots \chi_{n-1}\psi_{n-1}\chi_{n-1}'$ is a derivation in G'. Note that $\chi_0\phi_0\chi_0' = S'$ and $\chi_{n-1}\psi_{n-1}\chi_{n-1}' = \omega$ and therefore $S' \overset{*}{\underset{G'}{\Rightarrow}} \omega$. ∎

We obtain two corollaries by specifying in appropriate ways both the terminal and nonterminal vocabularies V_T and V_N of the grammar we have been constructing.

Corollary 3. Every r.e. set is generated by an MPS/1 grammar with at most one phantom symbol.

Proof: The MPS grammar constructed in the proof of Theorem 2 generates the r.e. set L if we choose its terminal vocabulary V_T to be $\Sigma(= V'_T)$ and its nonterminal vocabulary V_N to be $\{S, A\} \cup V'_N$. In the grammar, all members of $\{A\} \cup V'_N$ are phantom symbols, as are those members of V'_T that do not occur in any string in L. To complete the proof, we need only show how to encode this grammar so that it uses just one phantom symbol.

One way to do this is as follows. Let $V_T \subseteq V'_T$ be the smallest set such that $L \subseteq V_T^*$. Now number the symbols of $\{A\} \cup V'_N \cup (V'_T - V_T)$ consecutively as $\alpha_0, \ldots, \alpha_n$. Choose some $c \in V_T$. Let $f : \{S, A\} \cup V'_N \cup V'_T \rightarrow \{S\} \cup V_T$ be defined by $f(\beta) = \beta$ if $\beta \in \{S\} \cup V_T$ and $f(\beta) = Ac^iA$ if β is the ith symbol in $\{A\} \cup V'_N \cup (V'_T - V_T)$. Extend f to $(\{S, A\} \cup V'_N \cup V'_T)^*$ by putting $f(\gamma\delta) = f(\gamma)f(\delta)$ for all strings γ and δ over the vocabulary. Now f is the encoding we need because it is one-to-one, i.e., if $f(\gamma) = f(\delta)$ then $\gamma = \delta$. Thus, the grammar we desire has the basic rule $f(S) \rightarrow f(AS'A)$ and the following metarules:

(i) $f(S) \rightarrow f(A\phi)X \Rightarrow f(S) \rightarrow f(A\psi)X$ whenever $\phi \rightarrow' \psi$ is a rule of G'

(ii) $f(S) \rightarrow f(A\alpha)X \Rightarrow f(S) \rightarrow f(A)Xf(\alpha)$ whenever $\alpha \in V'_T \cup V'_N \cup \{A\}$

(iii) $f(S) \rightarrow f(A)Xf(A) \Rightarrow f(S) \rightarrow X$

The rule derivations of our new grammar are precisely the encoding of rule derivations in our old grammar. ∎

Corollary 4. Some nonrecursive language is generated by an MPS/1 grammar without phantom symbols.

Proof: Let L be an r.e. but not recursive language, and let $G = (V_T, V_N, S, \rightarrow, V_V, \Rightarrow)$ be the MPS/1 grammar whose basic rule $\rightarrow$ and metarules $\Rightarrow$ are constructed as in the proof of Theorem 2 and where $V_T = \Sigma \cup V'_N \cup \{A\}$, $V_N = \{S\}$ and $V_V = \{X\}$. Then $L = L(G) \cap$

Σ^* whence it follows that $L(G)$ is not recursive—as Σ^* is, L is not, and the recursive sets are closed under intersection. ∎

We now proceed to show that not all r.e. languages are generated by MPS/1 grammars without phantom symbols. Theorem 5 states that all languages generated by MPS grammars without phantom symbols, however many essential variables they employ, have the property of being arithmetically dense.[4] Thus, these languages constitute a proper subset of the class of r.e. languages. In fact, they omit some indexed, and therefore context-sensitive and *a fortiori* recursively decidable, languages such as $\{a^{n^2} \mid n \geq 0\}$. It is useful in formulating and proving the theorem to employ the following notation.

For any string ϕ, we let $|\phi|$ denote the length of ϕ, $[\phi]_\alpha$ denote the number of occurrences of the symbol α in ϕ, and $[\phi]_G$ denote the total number of occurrences in ϕ of all symbols other than nonterminals that cannot be rewritten in G to a nonempty terminal string. That is, when $N = \{A \in V_N \mid \text{there is no } y \in V_T^+ \text{ such that } A \overset{*}{\underset{G}{\Rightarrow}} y\}$, *then* $[\phi]_G = \Sigma_{\alpha \in V_T \cup V_N - N}[\phi]_\alpha$. When context makes clear what grammar G is intended, we suppress the subscript and write $[\phi]$ instead of $[\phi]_G$.

Theorem 5. If G is an MPS grammar with no phantom symbols and if $L(G)$ is infinite, then there exists a constant c such that to any natural number n corresponds some $x_n \in L(G)$ satisfying $cn \leq |x_n| < c(n+1)$.

Proof: Let $G = \langle V_T, V_N, S, \rightarrow, V_V, \Rightarrow \rangle$ satisfy the hypotheses of the theorem. Our proof divides into two cases according to whether or not there is a bound on $[\phi]$ for all derived rules $A \rightarrow \phi$ of G. More precisely, we consider whether or not a number l exists such that every rule derivation $A_1 \rightarrow \phi_1 \Rightarrow A_2 \rightarrow \phi_2 \Rightarrow \ldots A_j \rightarrow \phi_j$ has $[\phi_j] \leq l$ if $A_1 \rightarrow \phi_1$ is a basic rule of G. Our strategy is to show that if such a bound exists, then $L(G)$ is a context-free language and

[4]This is identical with the constant-growth property, which is discussed in connection with Tree Adjoining Grammars in Joshi, 1983.

thus has the desired density, whereas if there are derived rules $A_j \to \phi_j$ with arbitrarily large $[\phi_j]$, we can use these rules to generate an infinite, dense sublanguage of $L(G)$, whence $L(G)$ itself must be dense.

For the first case, we proceed by the following lemma, which does not exploit the lack of phantom symbols.

Lemma 6. If G is an MPS grammar for which there is a number l such that $[\phi] \leq l$ for every derived rule $A \to \phi$ of G, then $L(G)$ is a context-free language.

Proof: Let $G = \langle V_T, V_N, S, \to, V_V, \Rightarrow \rangle$ satisfy the hypotheses of the lemma, and let $N = \{A \in V_N \mid \forall y \in V_T^* \, (y = \epsilon \text{ if } A \overset{*}{\underset{G}{\Rightarrow}} y)\}$. To obtain a finite set of context-free rules generating $L(G)$, we simply delete from the right-hand side of every basic and derived rule of G each occurrence of all nonterminals belonging to N. Let $\to'$ be the set of rules obtained in this way, and set $G' = \langle V_T, V_N, S, \to' \rangle$. We show that $L(G') = L(G)$ and that G' is a context-free grammar, viz., $\to'$ is finite.

To see that $L(G') \supseteq L(G)$, note that to each derivation from S to a terminal string x in G there corresponds a derivation from S to x in G', which differs only in that nonterminals A belonging to N are never introduced and, therefore, the steps in which they are rewritten, ultimately to ϵ, are omitted. For the other direction of inclusion, when given a derivation from S to a terminal string x in G', we can construct a derivation from S to x in G by stepwise (a) applying the same rule at the same place as applied in the given derivation, if that rule belongs to G as well as to G', or (b), if such is not the case, applying a rule of G that reduces to the rule of G' upon deletion of all occurrences of all $A \in N$, and then following this by steps that rewrite to ϵ each of the nonterminals $A \in N$, just introduced. Thus $L(G') = L(G)$.

To see that $\to'$ is a finite set, we observe that there is an upper bound on the length of the right-hand side of any rule in $\to'$. Finiteness follows because each rule is a string of bounded length in the finite vocabulary $V_T \cup (V_N - N)$—paired with a string of length 1 in the finite vocabulary V_N. The desired upper bound is simply the smallest number l such that

$[\phi] \leq l$ for every derived rule $A \rightarrow \phi$ of G; this number must exist by the hypotheses of the lemma. ∎

Returning to the proof of Theorem 5, recall that we were considering the case of any MPS grammar G that satisfies the hypotheses both of the theorem and of Lemma 6. In this case the conclusion of the theorem follows by virtue of the standard 'pumping lemma' for context-free languages. As $L(G)$ is a context-free language, there is a number p such that to every $z \in L(G)$ with $|z| > p$ there correspond u, v, w, x and y such that $|v| + |x| > 0$ and moreover, for all $j \geq 0$, $uv^j wx^j y$ belongs to $L(G)$. (We could have concluded the theorem with this stronger 'intercalation' statement as far as the present case is concerned, but in our second case our argument yields only the weaker conclusion we are about to deduce.) Since by hypothesis, $L(G)$ is infinite, we can certainly find a $z_1 \in L(G)$ such that $|z_1| > p$. Put $c = |z_1|$ and, letting u_1, v_1, w_1, x_1 and y_1 be the strings associated with z_1, put $z_j = u_1 v_1^j w_1 x_1^j y_1$ for each $j \geq 0$. Since $|z_j| = j(|v_1| + |x_1|) + (|u_1| + |w_1| + |y_1|)$ and both $|v_1| + |x_1|$ and $|u_1| + |w_1| + |y_1|$ are at most c, the different strings z_j occur with ample density to assure that at least one will have length lying between cn and $c(n+1)$ for any natural number n that may be given. Thus, $L(G)$ is dense and the theorem is proved in the first case.

Turning now to the second case, we assume our given MPS grammar satisfies the hypotheses of the theorem and furthermore that for any number l there is a derivation $A_1 \rightarrow \phi_1 \Rightarrow A_2 \rightarrow \phi_2 \Rightarrow \ldots A_j \rightarrow \phi_j$ from a basic rule by metarules of G such that $[\phi_j] > l$.

We choose, for each $A \in V_N$, terminal strings x_A, y_A, and z_A so that $S \overset{*}{\underset{G}{\Rightarrow}} x_A A z_A$ and $A \overset{*}{\underset{G}{\Rightarrow}} y_A$, where $y_A \neg = \epsilon$ if possible. The absence of phantom symbols in G guarantees that we can find such terminal strings. We also adopt the notation $\overline{\phi}$ for the terminal string which results from replacing all occurrences of every $A \in V_N$ in ϕ by y_A, where ϕ is any string in $(V_T \cup V_N)^*$.

To construct an infinite sequence $w_0, w_1, w_2, \ldots$ of strings in $L(G)$ and a constant c so that $cn \leq |w_n| < c(n+1)$ for every $n \geq 0$, we proceed as follows. Choose w_0 to be y_S. We

will take c to be $|y_S|$ + some positive number, assuring $|w_0| < c$. Now suppose $w_0, w_1, \ldots, w_m$ to be given, satisfying $cn \leq |w_n| < c(n+1)$ for every $n = 0, \ldots, m$ and for the constant c that we will construct. To obtain w_{m+1} meeting the condition $c(m+1) \leq |w_{m+1}| < c(m+2)$, note that there is a derivation $A_1 \to \phi_1 \Rightarrow A_2 \to \phi_2 \Rightarrow \ldots A_j \to \phi_j$ from a basic rule by metarules of G such that $[\phi_j] \geq c(m+1)$. We choose w_{m+1} from among the members $u_1, u_2, \ldots, u_j$ of $L(G)$, where $u_i = x_{A_i}\overline{\phi}_i z_{A_i}$, to be the one with the smallest subscript i' such that $|u_{i'}| \geq c(m+1)$. This is possible since $|u_j| = |x_{A_j}| + |\overline{\phi}_j| + |z_{A_j}| \geq [\phi_j] \geq c(m+1)$; each u_i is in $L(G)$ because $S \overset{*}{\underset{G}{\Rightarrow}} x_{A_i}A_i z_{A_i}, A_i \to \phi_i$ is a rule of G and $\phi_i \overset{*}{\underset{G}{\Rightarrow}} \overline{\phi}_i$. To show that $|w_{m+1}| < c(m+2)$, as required, we observe that $|u_{i'-1}| < c(m+1)$ by the choice of i' and will prove that $|u_{i'}| \leq |u_{i'-1}| + c$ for our constant c. (We guarantee that $i' > 1$ by choosing c greater than $[\phi]$ for any of G's finitely many basic rules $A \to \phi$.)

Note now that, because of the way $\overline{\phi}_{i'}$ and $\overline{\phi}_{i'-1}$ are constructed, $|u_{i'}| - |u_{i'-1}| = |x_{A_{i'}}| + |z_{A_{i'}}| - (|x_{A_{i'-1}}| + |z_{A_{i'-1}}|) + \Sigma_{\alpha \in V_T \cup V_N} |y_\alpha|([\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha)$, if we take $y_\alpha = \alpha$ when $\alpha \in V_T$. Putting $k_0 = \max_{A \in V_N}(|x_A| + |z_A|)$, we see that $|x_{A_{i'}}| + |z_{A_{i'}}| - (|x_{A_{i'-1}}| + |z_{A_{i'-1}}|) \leq k_0$. Now let $k_1 = \max_{\alpha \in V_T \cup V_N} |y_\alpha|$ and $k_2 =$ the maximum size of any G's finitely many metarules. We will show that $[\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha \leq k_2$ for every $\alpha \in V_T \cup V_N$, whence it follows that $|y_\alpha|([\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha) \leq k_1 k_2$. Putting $k_3 = |V_T \cup V_N|$ then gives $\Sigma_{\alpha \in V_T \cup V_N} |y_\alpha|([\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha) \leq k_1 k_2 k_3$, guaranteeing that $|u_{i'}| - |u_{i'-1}| \leq k_0 + k_1 k_2 k_3$. Our desired constant c can thus be chosen as $1 + \max(|y_S|, k_4, k_0 + k_1 k_2 k_3)$, where k_4 is the maximum of $[\phi]$ for any basic rule $A \to \phi$. To see that $[\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha \leq k_2$ for each $\alpha \in V_T \cup V_N$, observe that the rule $A_{i'} \to \phi_{i'}$ is directly derived from $A_{i'-1} \to \phi_{i'-1}$ by application of a metarule of G. So there are a number p, a permutation $q_1, \ldots, q_p$ of the numbers $1, \ldots, p$, variables $\xi_1, \ldots, \xi_p \in V_V$ and strings $\sigma_1, \ldots, \sigma_{p+1}, \tau_1, \ldots, \tau_{p+1}, \omega_1, \ldots, \omega_p \in (V_T \cup V_N)^*$ such that G has the metarule $A_{i'-1} \to \sigma_1 \xi_1 \ldots \sigma_p \xi_p \sigma_{p+1} \Rightarrow A_{i'} \to \tau_1 \xi_{q_1} \ldots \tau_p \xi_{q_p} \tau_{p+1}$, and moreover $\phi_{i'-1} = \sigma_1 \omega_1 \ldots \sigma_p \omega_p \sigma_{p+1}$ and $\phi_{i'} = \tau_1 \omega_{q_1} \ldots \tau_p \omega_{q_p} \tau_{p+1}$. Thus for each symbol α, $[\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha = t_1 + t_2$ where $t_1 = ([\tau_1]_\alpha + \ldots + [\tau_{p+1}]_\alpha) - ([\sigma_1]_\alpha + \ldots + [\sigma_{p+1}]_\alpha)$ and $t_2 = ([\omega_{q_1}]_\alpha + \ldots + [\omega_{q_p}]_\alpha) - ([\omega_1]_\alpha + \ldots + [\omega_p]_\alpha)$. But $t_2 = 0$ since $\omega_{q_1}, \ldots, \omega_{q_p}$ is a permutation of $\omega_1, \ldots, \omega_p$. And

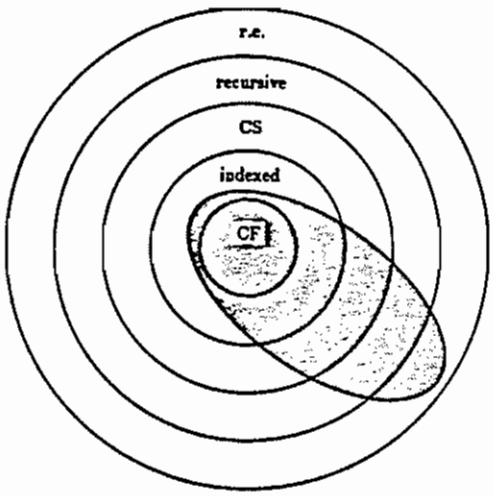
$t_1 \leq k_2$; indeed we might as well choose the maximum value of t_1 for any of G's metarules as k_2. So $[\phi_{i'}]_\alpha - [\phi_{i'-1}]_\alpha \leq k_2$ for each α, as was to be shown.

Clearly our constant c depends only on G and the fixed string y_S, not otherwise on a string w_n chosen at any stage. Thus repeating the procedure indefinitely yields the sequence $w_0, w_1, w_2, \ldots$ desired. ∎

The density theorem states one property exhibited by all languages generated by MPS grammars without phantom symbols. We conjecture that this density property is not a sufficient condition for an r.e. language to be generated by an MPS grammar without phantom symbols. Another open question is whether the commutative image [Parikh 1966] of any language generated by an MPS grammar without phantom symbols is a semi-linear set.

4. Discussion and Conclusion

The diagram below shows how MPS languages without phantom symbols fit into the schema known in formal language theory as the Chomsky hierarchy.



Our results strongly suggest that a grammatical framework for the description of natural language should not use metarules in the way they are defined here for MPS grammars. With or without phantom categories, MPS grammars are able to generate nonrecursive languages. And no linguist who takes into appropriate account the admittedly little knowledge we have about parsability and learnability constraints would want a theory of that power.

Of course, rather than abandon metarules, one can look for alternative ways of defining them. Two alternatives to the way we have defined metarules for MPS grammars have been proposed within GPSG. The first one, suggested by Gazdar (1982), redefines the form of metarules. Although Gazdar considered metarules with just one essential variable 'safe' with respect to the class of languages generated, he wanted nevertheless to rule out infinite rule sets. His remedy was to sacrifice essential variables. All his variables, string variables included, are supposed to be abbreviatory (although he does not specify the sets they range over).

Note, however, that the generalization stated by a metarule is weakened by converting essential into abbreviatory variables. Consider, for instance, the following metarule, which he proposed for providing VSO languages with the category VP. The metarule generates flat VSO sentence rules from VP rules:

(4) $VP \rightarrow V\ U \Rightarrow S \rightarrow V\ NP\ U$

Now we need to specify the range of the abbreviatory variable U. Let us imagine that the VSO language in question has the following small set of VP rules:

(5) $VP \rightarrow V$
$VP \rightarrow V\ NP$
$VP \rightarrow V\ \overline{S}$
$VP \rightarrow V\ \overline{VP}$
$VP \rightarrow V\ NP\ \overline{VP}$

The range of U has to be $\{\epsilon, NP, \overline{S}, \overline{VP}, NP\,\overline{VP}\}$. If the VP rules under (5) are the only rules that satisfy the left-hand side of (4), then (4) generates exactly the same rules as a corresponding rule would in which U had been replaced with an essential variable. But now let us imagine that our language acquired a new subcategorization frame for verbs, e.g. has added a verb that takes an NP and an $\overline{S}$ as complements. We have to add this VP rule:

(6) $VP \rightarrow V \quad NP \quad \overline{S}$

Our metarule (4) would predict that VPs headed by this verb do not have a corresponding flat SVO sentence rule. However, this situation is very unlikely to occur. The range of the abbreviatory variable U would have to be changed to extend the range of the metarule. It does not have to be a change in the basic rules that necessitates a change in the range of a variable; it could also be a change in a metarule that, either directly or indirectly, feeds the metarule containing the variable in question. The difference in meaning between a metarule with an abbreviatory variables and the corresponding metarule with essential variables can be viewed as an extension—intension contrast. Metarule (4) in our example is supposed to express the fact (among others) that all VPs can occur in SVO sentences as discontinuous constituents. If U is an abbreviatory variable, the phrase 'all VPs' refers only to a specific list of VPs known to the grammar writer at the time the grammar is written. If U is an essential variable, the metarule encodes a stronger statement—i.e., that actually **all** VPs in the language, independently of our knowledge or the present coverage of the grammar, participate in the syntactic phenomenon.

A second strategy for constraining metarules is to leave their form as we have defined them but modify their role within the grammar. Thompson (1982), following a suggestion by Martin Kay, devised an alternative to simply closing the PS rules of the grammar under the metarules. Under his proposal the rules of the grammar are in the 'Finite Closure' of the metarules.

According to the definition of Finite Closure, a metarule $M = A \Rightarrow B$ generates a rule

of the kind B from any rule that matches the template A and is either itself a base rule or is derived from a base rule without any application of M in its derivational history. Obviously, no infinite rule sets can be generated. The Finite Closure solution has now been generally adopted for GPSG.

Although the Finite Closure concept makes metarules 'safe', it is not a linguistically grounded solution. Many other stipulations would keep the rule set finite. As an example, let us consider a fictitious proposal to put a small upper bound on the length of the strings that essential variables can range over. This stipulation would be as ad hoc as the Finite Closure solution, but would not require a redefinition of metarule functioning; moreover, it would simplify the computation of the PS rules from the set of basic rules. Nevertheless, rescue measures such as the adoption of Finite Closure might constitute an appropriate research strategy as an intermediate step.

However, what is really needed is a linguistically motivated, well-defined concept of metarules. So far it is not even clear which classes of phenomena should be handled by metarules. Within GPSG, opinions regarding this question have been in flux. Two examples might illustrate this fluctuation. Word order phenomena, which at one point were handled by metarules, (Stucky, 1981) are now done with the ID/LP format, a mechanism that separates immediate dominance and linear precedence from each other in the grammar (Gazdar and Pullum, 1981). At first, long-distance dependencies were treated with the so-called slash categories and a special schema for deriving rules that contain them (Gazdar, 1981). Later metarules were used to derive the rules that eliminate and percolate gaps (Sag, 1981). In the newest version of the theory, metarules are less involved in gap handling; 'feature instantiation principles' have taken over most the task of gap percolation (Gazdar and Pullum, 1982).

A lack of consensus about the class of phenomena to be handled by metarules is not the only hurdle on the way to a sensible redefinition of metarules. If one considers the interaction between metarules and all the recently introduced mechanisms in GPSG, the task of redefining

metarules becomes even more complicated. We have shown that the interaction of metarules and phantom symbols has an influence on the generative power of the framework. Uszkoreit (1982) has noted a similar interaction between metarules and the ID/LP format. Assessing the influence of metarules on the generative power of an MPS grammar framework that also employs other formalisms can lead to unexpected results.

Any attempt at constraining or redefining metarules within a given grammatical framework has to be based on a thorough survey of the full variety of metarules that have been proposed, on a decision about the class of phenomena that should be captured by them, and on a detailed study of the interaction between metarules and other formalisms of the theory.[5]

[5]Since we finished the research for this paper, Shieber et al. (1983) have discussed a number of possible constraints on metarules (including several that have been actually proposed) and have arrived at largely negative conclusions as to their usefulness.

References

Culy, C.: 1982, 'On the Generative Power of Metarules', unpublished manuscript, Stanford University.

Gazdar, G.: 1982, 'Phrase Structure Grammar', in P. Jacobson and G.K. Pullum (eds.), **The Nature of Syntactic Representation**, Reidel, Dordrecht.

Gazdar, G.: 1981, 'Unbounded Dependencies and Coordinate Structure,' **Linguistic Inquiry** 12, 155-184.

Gazdar, G. and G.K. Pullum: 1981, 'Subcategorization, Constituent Order and the Notion "Head",' in M. Moortgat, H. v.d.Hulst and T. Hoekstra, eds., **The Scope of Lexical Rules**, 107-123, Foris, Dordrecht.

Gazdar, G. and G.K. Pullum: 1982, 'Generalized Phrase Structure Grammar: A Theoretical Synopsis,' Indiana University Linguistics Club, Bloomington, Indiana.

Gazdar, G., G.K. Pullum, and I.A. Sag: 1981, 'Auxiliaries and related phenomena in a restrictive theory of grammar,' **Language** 58, 591-638.

Gazdar, G. and I.A. Sag: 1980, 'Passives and Reflexives in Phrase Structure Grammar,' in J. Groenendijk, T. Janssen, and M. Stokhof (eds.) **Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium,** Mathematical Centre Tracts 135, Amsterdam.

Joshi, A.K.: 1983, 'How Much Context-Sensitivity Is Required to Provide Reasonable Structural Descriptions: Tree Adjoining Grammars,' to appear in D. Dowty, L. Karttunen, and A. Zwicky, **Natural Language Processing: Psycholinguistic, Computational, and Theoretical Perspectives,** Cambridge University Press, Cambridge.

Konolige, K.: 1980, 'Capturing Linguistic Generalizations with Metarules in an Annotated Phrase-Structure Grammar,' in **Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics,** Philadelphia, Pennsylvania.

Peters, S. and H. Uszkoreit: 1982, 'Essential Variables in Metarules,' paper presented at the 1982 Annual Meeting of the Linguistic Society of America , San Diego, California.

Shieber,S.M., S.U. Stucky, H. Uszkoreit, J.J. Robinson: 1983, 'Formal Constraints on Metarules', in **Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics,** Cambridge, Mass.

Stucky, S.: 1981, 'Word Order Variation in Makua', unpublished Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Thompson, H.: 1982, 'Handling Metarules in a Parser for GPSG,' **Edinburgh DAI Research Paper No. 175,** J. Horecky, ed., **Proceedings of the Ninth International Conference on Computational Linguistics,** North Holland, Dordrecht.

Uszkoreit, H.: 1982, 'German Word Order in GPSG,' in D. Flickinger, M. Macken, and N. Wiegand (eds.), **Proceedings of the First West Coast Conference on Formal Linguistics,** Stanford University, Stanford, California.

Appendix: Definition of MPS grammars

Definition 7. MPS grammars:
An MPS grammar is a sextuple $(V_T, V_N, S, \rightarrow, V_V, \Rightarrow)$ such that V_T and V_N are finite, disjoint sets of terminal and nonterminal symbols respectively, $S \in V_N$, V_V is a finite set of essential variables disjoint from $V_T \cup V_N$, $\rightarrow$ is a finite subset of $V_N \times (V_T \cup V_N)^*$, and $\Rightarrow$ is a finite subset of $(V_N \times (V_T \cup V_N \cup V_V)^*) \times (V_N \times (V_T \cup V_N \cup V_V)^*)$ such tbat, if $\langle\langle A, \phi\rangle, \langle B, \psi\rangle\rangle \in \Rightarrow$, then each member of V_V occurs an equal number of times in ϕ and ψ and has at most one occurrence in each.

Let $G = (V_T, V_N, S, \rightarrow, V_V, \Rightarrow)$ be an MPS grammar.

Definition 8. directly yields:
A pair $\langle A, \phi\rangle$ directly yields a pair $\langle B, \psi\rangle$ by a metarule of G(in symbols$\langle A, \phi\rangle \Rightarrow \langle B, \psi\rangle$) if, for some pair $\langle\langle A, \chi\rangle, \langle B, \omega\rangle\rangle \in \Rightarrow$, there are a positive integer n and strings $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n-1}$, $\gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_{n-1} \in (V_T \cup V_N)^*$ and symbols $\varsigma_1, \ldots, \varsigma_{n-1}, \eta_1, \ldots, \eta_{n-1} \in V_V$ such that

(i) $\beta_i = \delta_j$ when $\varsigma_i = \eta_j$ $(1 \leq i, j < n)$

(ii) $\chi = \alpha_1 \varsigma_1 \ldots \alpha_{n-1} \varsigma_{n-1} \alpha_n$

(iii) $\phi = \alpha_1 \beta_1 \ldots \alpha_{n-1} \beta_{n-1} \alpha_n$

(iv) $\omega = \gamma_1 \eta_1 \ldots \gamma_{n-1} \eta_{n-1} \gamma_n$

(v) $\psi = \gamma_1 \delta_1 \ldots \gamma_{n-1} \delta_{n-1} \gamma_n$.

Definition 9. rule of G:
A pair $\langle A, \phi\rangle$ is a rule of G if there is a sequence $\langle B_1, \psi_1\rangle, \ldots, \langle B_n, \psi_n\rangle$ of pairs such that $\langle B_1, \psi_1\rangle \in \rightarrow, \langle B_n, \phi_n\rangle = \langle A, \phi\rangle$, and $\langle B_i, \psi_i\rangle$ directly yields $\langle B_{i+1}, \psi_{i+1}\rangle$ by a metarule of G for $1 \leq i < n$, i.e., $\langle A, \phi\rangle$ is a rule of G if there is a basic rule $\langle B, \psi\rangle$ of G such that $\langle B, \psi\rangle \stackrel{*}{\Rightarrow} \langle A, \phi\rangle$. We let $\stackrel{\rightarrow}{G}$ denote the set of all rules of G.

Definition 10. directly derives:
The relation $\stackrel{\rightarrow}{G}$ on $(V_T \cup V_N)^*$ holds of a pair $\langle \phi, \psi\rangle$ if and only if there are $\alpha, \beta, \omega \in (V_T \cup V_N)^*$ and $A \in V_N$ such that $\phi = \alpha A \beta, \psi = \alpha \omega \beta$, and $\langle A, \omega\rangle \in \stackrel{\rightarrow}{G}$.

Definition 11. derives:
The relation $\stackrel{*}{\stackrel{\rightarrow}{G}}$ is the ancestral (reflexive, transitive closure) of $\stackrel{\rightarrow}{G}$.

Definition 12. language generated by G:
The language generated hy G is $\{x \in V_T^* \mid S \stackrel{*}{\stackrel{\rightarrow}{G}} x\}$.

Definition 13. MPS/1 grammars:
An MPS/1 grammar is an MPS grammar $(V_T, V_N, S, \rightarrow, V_V, \Rightarrow)$ with a singleton V_V.