This dissertation has been
microfilmed exactly as received

69-13,913

CHANCE, Lelia Downing, 1933-
GRAMMATICAL SETS IN HALF-RING MORPHOLOGIES.

The University of Oklahoma, Ph.D., 1969
Language and Literature, linguistics
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/2553

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# GRAMMATICAL SETS IN HALF-RING MORPHOLOGIES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

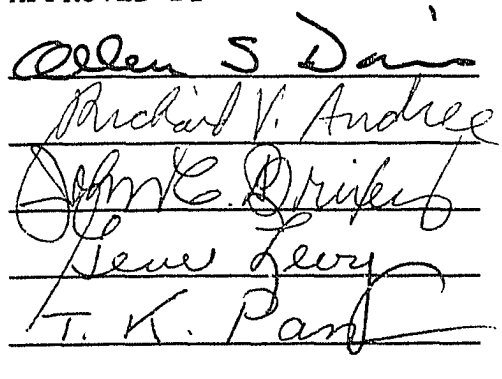
LELIA DOWNING CHANCE

Norman, Oklahoma

1968

GRAMMATICAL SETS IN HALF-RING MORPHOLOGIES

APPROVED BY



DISSERTATION COMMITTEE

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS. . . . . iv

Chapter

I. INTRODUCTION. . . . . 1

II. HALF-RING MORPHOLOGIES. . . . . 9

III. GRAMMATICAL SETS. . . . . 26

IV. LINGUISTIC SETS . . . . . 52

BIBLIOGRAPHY. . . . . 108

## ACKNOWLEDGEMENTS

I wish to express my gratitude to my thesis advisor, Professor A. S. Davis, for his guidance and encouragement.

The research in this dissertation (as well as the author) was supported under a National Aeronautics and Space Administration Traineeship at the University of Oklahoma.

# GRAMMATICAL SETS IN HALF-RING MORPHOLOGIES

## CHAPTER I

## INTRODUCTION

The current lively interest in structural linguistics among mathematicians is recent. Its history may be said to begin in 1960, when a mathematical model for the syntax of a language, called a phrase structure grammar, was defined by Noam Chomsky [2,3]. The assumption motivating the model was that sentences in language are generated by a sequence of rewriting rules which, beginning with the concept "sentence" itself, relate or subdivide general syntactical categories into classes more and more specific, and finally into the particular words (or morphemes) used in the sentence.

An example will help to clarify this idea. We assume these grammatical facts:

(1) A sentence may be composed of a subject followed by a predicate.

(2) A subject may be a noun phrase.

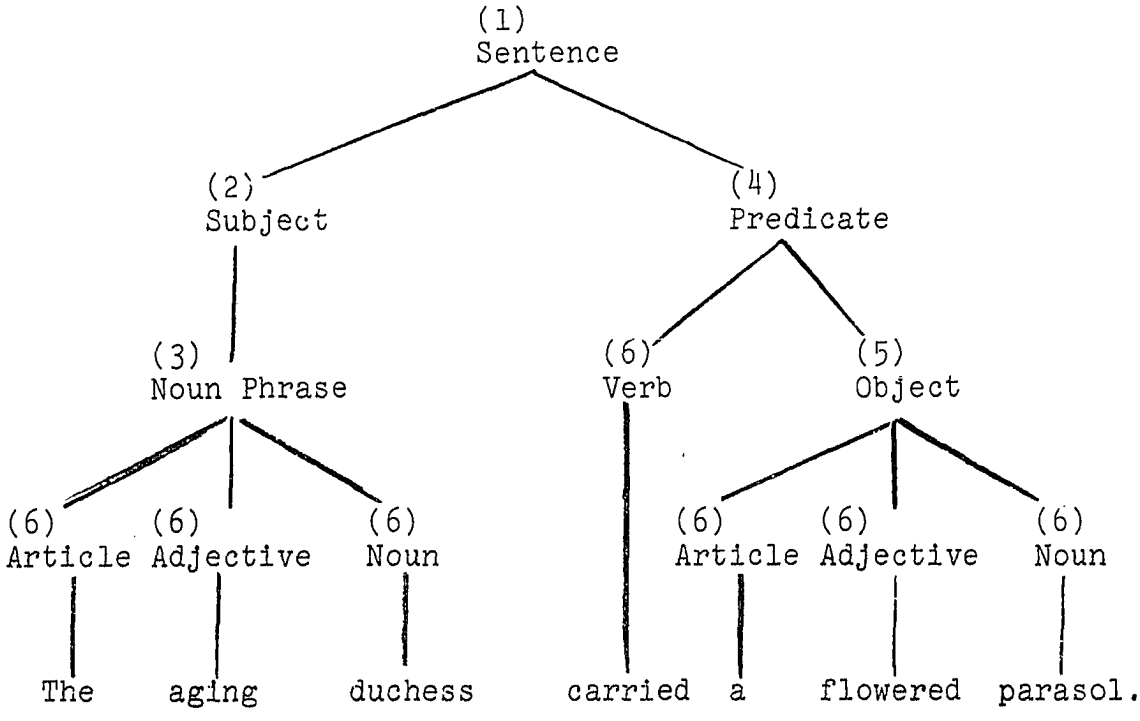
(3) A noun phrase may be a noun preceded by an article and an adjective.

(4) A predicate may be a verb followed by an object.

(5) An object may be a noun phrase.

(6) "The" is an article; "a" is an article; "aging" is an adjective; "flowered" is an adjective; "duchess" is a noun; "parasol" is a noun; "carried" is a verb.

Now, applying the rewriting rules inherent in statements 1 through 6, we may construct the sequence




We can also construct the sentence "A flowered duchess carried the aging parasol," or "A flowered parasol carried the againg duchess," which illustrates the fact that structure, not meaning, is what the grammar is intended to model.

Phrase structure grammars are classified into types according to the type of rewriting rules or productions allowed. They are, in increasing order of generality: right (or left) linear, context-free, context-sensitive, and arbitrary phrase structure grammars. There is now a large body of knowledge about these grammars, along with associated models of machines. The machines, with a finished sentence as input, perform a sequence of operations which result in the acceptance of a sentence which is well-formed according to a specified set of grammatical rules.

Since this paper is concerned with an extension of the notion of context-free grammar, some familiarity with

phrase structure grammars must be assumed. Virtually all the results used here, along with a thorough treatment of results in the area up to 1965, can be found in [7].

The interest in context-free grammars was fed by the discovery that they were equivalent to a format for the specification of programming languages called Backus normal form. Algol was specified in this form, and a class of languages called Algol-like--those whose syntax could be specified in Backus normal form--was found to be the same as the class of languages generated by context-free grammars [10]. However, because of some side restrictions on the form of Algol statements, it turned out that Algol was not in fact an Algol-like language [6]. This discovery motivated a search for a model slightly more general than the context-free grammar and its associated accepting machine, the pushdown acceptor (pda).

Linguists dealing with natural languages found objections to phrase structure grammars as a model. The class of context-free grammars was too small to mirror the complexities of natural language; the class of context-sensitive ones somewhat unwieldy. Chomsky himself resorted to the use of additional operations called transformations, which are applied to primitive sentence forms generated by context-free grammars. A remarkable number of new accepting machines have been defined, which (without corresponding generating rules) delimit a language as that collection of sentences accepted by the machine. A summary of most of these, along with a chart showing the known and conjectured relationships between them, appears in [9].

There has also been a bustling business in the generalization of context-free grammars. Notable among the new grammars are the programmed grammars of Rosenkrantz [13], which use context-free rules, whose eligibility for application depends on which production was applied last and on

the form of the intermediate string at the moment of application. The indexed grammars of Aho [1], utilize a new type of rule, called an indexed production, in addition to context-free rules.

Underlying the notion of context-free languages and the above generalizations is the fact that all words (which unfortunately is the term used for well-formed strings corresponding to the intuitive notion of sentences which we discussed earlier) in a language are assumed to be elements in the free semigroup generated by a finite collection of symbols, where the operation is juxtaposition. Davis [5] suggested that this simple juxtaposition is an oversimplification of the way grammatical elements are linked together to form syntactically correct strings. He proposed, as a substitute for the semigroup, an algebraic system called a half-ring morphology, with three operations, as a suitable model for the natural linkages of syntactical elements. We illustrate this with an example.

A transitive verb calls for both a subject and an object. It is natural to think of it as a two-place predicate, with two numbered blanks, one to be filled with a subject, the other with an object, as in

( 1 carried 2 ).

We form 2-tuples of the form (subject, object), where each of these is without blanks, although they may be composed of smaller elements containing blanks. Then the composition operation · in the half-ring morphology is so defined that ( 1 carried 2 ) · (the aging duchess, a flowered parasol) = The aging duchess carried a flowered parasol. That is, the first element of the pair is substituted for the blank numbered 1, the second for the blank number two. The second operation of the half-ring morphology, concatenation, represents the formation of n-tuples. In the grammatical rules we are then able to replace "followed by" with more complex types of linkage.

Davis' suggestion is that context-free rules be used to generate meaningful strings of elements in a morphology along with operator symbols, and then, after the generation process is complete, to perform the indicated operations to obtain finished, filled-in sentences. That is what this paper attempts to do: to investigate the sets obtained in such a way.

The other immediate ancestor of this approach is a paper of Mezei and Wright [11]. Their generalization of languages generated by context-free rules in semigroups to languages generated by context-free rules in arbitrary algebraic systems is precisely what is needed to implement Davis' suggestion for half-ring morphologies. The alternative formulation of recognizable sets in Chapter 3 is an application of their approach. The term recognizable set is due to them; the term grammatical set also appears in their paper, attributed by them to David Muller. The special cases of their results which are used in this paper are summarized in Chapter 3. It is hard to overestimate their value in simplifying proofs and adding a taste of much-needed elegance.

The paper is organized as follows. Chapter II contains the definition and basic results for half-ring morphologies. The theorems which appear there are due to Davis and are stated without proof (sometimes in slightly altered form) in [5]. The proofs are mine, and are included for completeness. It will be useful to refer in later chapters to some of the constructions used in these proofs. Lemma 2.2 is proved in [4], where half-ring morphologies were defined for a different use.

In Chapter III, a half-ring grammar is defined. It is the special context-free grammar which will generate well-formed expressions involving morphology elements and morphology operation symbols. The equivalent formulation

of Mezei and Wright using finite congruences on a generic algebra is presented. The collection of strings whether generated by a half-ring grammar or representing a union of congruence classes is called a recognizable set. A grammatical set is then defined as the collection of morphology elements resulting from carrying out the operations represented by the strings in the recognizable set. We state a best form for a grammar, which is an outcome of the results in [11].

Various closure properties of grammatical sets are investigated. In the case of the usual semigroup languages, Ginsburg and Greibach have abstracted a collection of closure properties by which they define an Abstract Family of Languages (AFL) [8]. Theorems 3.3, 3.4, 3.6, and 3.7 provide what I feel are appropriate analogues of these properties in the half-ring case. Theorem 3.21 demonstrates a closure property related to the AFL requirement that languages be closed under intersection with a regular set. I am considerably less sure that this property is the proper analogue to the AFL one.

A number of examples of grammatical sets in linear morphologies are given, including sets which can not be generated by context-free rules in a semigroup. Theorem 3.10 gives the result that every grammatical set is the homomorphic image of a grammatical set in a free morphology, a fact which will be useful in Chapter 4.

Regular sets are particularly well-behaved subsets of a free semigroup, generated by rules of a particularly simple form (see [7]). In attempting to define an analogous class for grammatical sets in morphologies, we introduce the notion of A-regularity, for any recognizable set A. For each A, we obtain a class closed under union, intersection, and complementation with respect to the set generated by A. A particular recognizable set F, called

the set of factorizations because of its relationship to the factorizations of phrases in a morphology discussed in Chapter 2, is defined. In Chapter 4, we find that the F-regular grammatical sets in a free morphology are generated by rules of an attractive simplicity, and further, that all context-free languages (in the usual sense) are homomorphic images of such grammatical sets under a very simple homomorphism.

The concept of concatenative depth and the dimension and degree of a grammatical set are introduced in Chapter 3, since they will be needed in Chapter 4.

The last section of Chapter 3 deals with ambiguity. Two types of ambiguity are defined for grammatical sets: structural and morphological. The first type is a function of the generating rules; the second has to do with the particular morphology into which the recognizable set is mapped to produce a grammatical set. It turns out (Theorem 3.27) that there is no morphological ambiguity in free morphologies, and that there is no structural ambiguity in F-regular grammatical sets in any morphology (Theorem 3.28). The relationship between structural ambiguity and the inherent ambiguity of semigroup context-free languages is discussed, and an example is given in which a language known to be inherently ambiguous is generated without either morphological or structural ambiguity. In Chapter 4, we show that all semigroup context-free languages can be generated as grammatical sets without structural ambiguity.

Chapter 4 is concerned with special grammatical sets in linear morphologies which we call restricted linguistic sets, and which are shown to be appropriate for the linguistic model we have in mind. They are (Theorem 4.10) those grammatical sets containing only "completely filled in" expressions in the morphology (called formulas), which represent a single sentence (rather than a string of

sentences, or a paragraph, for example) and which can be generated by variables representing grammatical categories which yield n-tuples of a fixed length and a fixed distribution of blanks for the category. Some closure properties of linguistic sets and restricted linguistic sets are found.

A number of results having to do with the representation of the usual context-free languages as grammatical sets are presented.

It happens that, in a linear morphology, the formulas are in the form of strings of (juxtaposed) symbols in a set S, as are the words in context-free languages. If we consider these strings as elements of the free semigroup generated by S, we are able to examine some closure properties usually associated with context-free language. These we call substratum properties.

The chapter concludes with a method for extending the model to allow "erasures," or the elimination of unnecessary empty blanks in the formation of sentences.

# CHAPTER II

## HALF-RING MORPHOLOGIES

The definitions, notation, and terminology presented in this chapter follow Davis [5], with minor alterations.

Morphologies. By a half-ring we mean an algebraic system $(E,*,\cdot)$ with binary operations $*$ and $\cdot$ satisfying

(i) $x\cdot(y\cdot z) = (x\cdot y)\cdot z$

(ii) $x*(y*z) = (x*y)*z$

(iii) $x*y = x*z$ implies $y = z$

(iv) $(x*y)\cdot z = (x\cdot z)*(y\cdot z)$

for all x, y, z in E. (Notationally, $*$ takes precedence over $\cdot$, so that $x\cdot y*z$ is $(x\cdot y)*z$, not $x\cdot(y*z)$.) The operations $*$ and $\cdot$ will be called concatenation and composition, respectively. Following custom, we will denote a morphology $(E,*,\cdot)$ by E.

Consider the half-ring generated by a denumerable sequence of elements $\underline{1}$, $\underline{2}$, etc., subject to just these defining relations:

(a) $\underline{1}\cdot(\underline{m}*x) = \underline{m}$

(b) $\underline{n+1}\cdot(\underline{m}*x) = \underline{n}\cdot(x*\underline{m})$

for all m, n = 1, 2, etc., and all x. It is easy to verify that such a half-ring does exist. Any such half-ring will be called a blank-morphology. The theorem which follows shows that there is (up to isomorphism) only one blank-morphology.

Let $B = (B,*,\cdot)$ be the half-ring generated by the natural numbers 1, 2, etc., where $*$ is juxtaposition, and composition is defined by

$$(n_1 n_2 \ldots n_k)\cdot(m_1 m_2 \ldots m_r) = m_{\bar{n}_1}\, m_{\bar{n}_2} \ldots m_{\bar{n}_k},$$

where $\bar{n}_i \equiv n_i \pmod r$ and $1 \le \bar{n}_i \le r$, for each $i = 1,2,\ldots,k$. Then B is the collection of all finite strings or sequences of natural numbers. B is easily seen to be a blank-morphology.

The following lemmas follow immediately from the definitions.

Lemma 2.1: In a blank-morphology,

$$\underline{n}\cdot(\underline{m_1} * \ldots * \underline{m_r}) = \underline{m_{\bar{n}}},$$

for natural numbers $n$, $m_i$, $i = 1,2,\ldots,r$, where $1 \le \bar{n} \le r$ and $\bar{n} \equiv n \pmod r$.

Lemma 2.2: In a blank-morphology,

$$\underline{n_1} * \ldots * \underline{n_k} = \underline{m_1} * \ldots * \underline{m_r},$$

for numbers $n_i$, $i = 1,2,\ldots,k$, and $m_j$, $j = 1,2,\ldots,r$, if and only if $k = r$ and $n_i = m_i$ for $i = 1,\ldots,k$.

Theorem 2.3: Every blank-morphology is isomorphic to B (above).

Proof: Let $H = (E,*,\cdot)$ be a blank-morphology generated by $G = \{1, 2,\ldots\}$. Let $\theta$: B→H be defined as follows:

$$\theta(n_1 \ldots n_k) = \underline{n_1} * \underline{n_2} * \ldots * \underline{n_k}.$$

Suppose $n_1 \ldots n_k$ and $m_1 \ldots m_r$ are non-null elements in B. Then, using the notation introduced above,

$$\theta[(n_1 \ldots n_k)\cdot(m_1 \ldots m_r) = \theta(m_{\bar{n}_1} * \ldots * m_{\bar{n}_k})$$

$$= \underline{m_{\bar{n}_1}} * \ldots * \underline{m_{\bar{n}_k}}, \text{ and}$$

$$\theta(n_1 \ldots n_k)\,;\theta(m_1 \ldots m_r) = (\underline{n_1} * \ldots * \underline{n_k})\cdot(\underline{m_1} * \ldots * \underline{m_r})$$

$$= \underline{m_{\bar{n}_1}} * \ldots \underline{m_{\bar{n}_k}},$$

by Lemma 2.1. Hence $\theta$ is a homomorphism. The onto property of $\theta$ follows from the fact that it maps B onto a set

of generators for E; the one-one property follows from Lemma 2.2. Hence $\theta$ is an isomorphism.

From now on, we will call B, or any morphology isomorphic to it, the blank-morphology.

For any positive integer n, let $h_n$: B→B be the map defined by: $h(x) = n$, for all x in B. Such maps will be called constant maps.

Theorem 2.4: The only endomorphisms of the blank-morphology are the identity map and the constant maps. Hence B has no non-trivial automorphisms.

Proof: Let $\theta$ be a non-trivial endomorphism of B, and let n by any integer such that $\theta(n) = m$ and $n \neq m$. Suppose $m>n$. For any numbers $a_1,\ldots,a_{n-1}, a_{n+1},\ldots,a_m$, we have

$$n\cdot(a_1a_2\ldots a_{n-1}na_{n+1}\ldots a_m) = n.$$

Applying $\theta$,

$$\theta(n)\cdot[\theta(a_1)\theta(a_2)\ldots\theta(a_{n-1})\theta(n)\theta(a_{n+1})\ldots\theta(a_m)] = \theta(n)$$
$$= \quad m\cdot[\theta(a_1)\theta(a_2)\ldots\theta(a_{n-1}) \;\; m \;\; \theta(a_{n+1})\ldots\theta(a_m)] = \; m \; .$$

But by Lemma 2.1,

$$m\cdot[\theta(a_1)\theta(a_2)\ldots\theta(a_{n-1}) \;\; m \;\; \theta(a_{n+1})\ldots\theta(a_m)] = \theta(a_m).$$

Since $a_m$ was arbitrary, we have $\theta(s) = m$ for all natural numbers s; a similar argument for $m<n$ shows that $\theta$ must be the constant map $h_m$.

Morphologies in general are half-rings in which the blank-morphology is embedded in a manner to be made precise in what follows.

A morphology is a system $(E,*,\cdot,\pi,')$ consisting of a half-ring $(E,*,\cdot)$ whose elements are called expressions, among which $\pi$ is distinguished as a first blank, and a unary shift operation $'$ such that

(v) $(x\cdot y)' = x\cdot y'$

(vi) $\pi\cdot\pi = \pi$

(vii) $\pi\cdot(x*y) = \pi\cdot x$

(viii) $x'\cdot(\pi\cdot y*z) = x\cdot(z*\pi\cdot y)$, and $x'\cdot\pi = x\cdot\pi$,

for all x, y, z in E.

Consider the half-ring H generated by the single element $\pi$, where * is juxtaposition, and composition is defined by

$$x \cdot y = x \text{ for all } x, y \text{ in } H.$$

Enlarging H by defining the unary shift as the identity operator, $x' = x$, we see that H becomes trivially a morphology for which $\pi' = \pi$. To exclude this trivial case, we add the restriction

(ix) $\pi' \neq \pi$,

which guarantees that in any morphology, the submorphology generated by $\pi$, $\pi'$, $\pi''$, etc., called <u>blanks</u>, is the blank-morphology. Denote $\pi'$ by $\pi^{(1)}$, and for $n>1$, $\pi^{(n+1)} = \pi^{(n)'}$. Then note that $\pi' \neq \pi$ implies that $\pi^{(m)} \neq \pi^{(n)}$ for all m, n such that $m \neq n$. Henceforth, blanks will be denoted by 1, 2, 3, etc.

In a morphology, an expression x is <u>closed</u> if $x \cdot y = x$ for all y. The degree of a closed expression is zero; otherwise the degree of x is either infinite or is the least n such that $x \cdot (1*2*\ldots*n) = x$. The dimension of x, if not infinite, is the least (unique) $m>0$ such that $(1*2*\ldots*m) \cdot x = x$. Expressions of dimension one are <u>phrases</u>. Closed phrases are <u>formulas</u>. A minimal set of phrases which generates the morphology is a <u>vocabulary</u>, whose members are called <u>morphemes</u>.

We will consider here only <u>locally finite</u> morphologies, that is, those satisfying

(x) for each x there are non-negative integers m and n such that $(1*\ldots*m) \cdot x = x = x \cdot (1*\ldots*n)$, and in this paper, "morphology" will mean a locally finite morphology.

<u>Linear morphologies</u>. Let S by any set, called an <u>alphabet of symbols</u>, or simply an <u>alphabet</u>. Let $N = \{1,2,\ldots\}$ be a denumerable set of numerals, disjoint from S. Let W be the set of all non-null finite strings $s_1 s_2 \ldots s_k$,

where $s_i$ is in S $\cup$ N for i = 1,2,...,k. Let E be the set of all n-tuples of elements in W, for n = 1,2,... . For x in E, call n the dimension of x. Define, for x and y in E, of dimensions r and s respectively, the sequences x′, x*y, and x·y in E as follows:

(1) x′ = the result of replacing each numeral n in x by n+1.

(2) x*y = the (r+s)-tuple whose components are defined by

$$(x*y)_i = \begin{cases} x_i, & \text{if } 0<i\leq r \\ y_{i-r}, & \text{if } r<i\leq r+s \end{cases}$$

for i = 1,2,...,r+s, and

(3) x·y = the r-tuple whose components are defined by $(x\cdot y)_i$ = the result of substituting $y_k$ for k in $x_i$ modulo s, for each integer k, for i = 1,2,...,r.

Let $\pi$ be the 1-tuple (1). Then $(E,*,\cdot,\pi,')$ is called the total linear morphology over S. Note that the dimension here defined corresponds to the definition of dimension in a general morphology. Any submorphology of the total linear morphology is a linear morphology over S.

Lukasiewicz morphologies are those linear morphologies over a set S which are generated by a vocabulary V each of whose members is of the form (s) or (sl...n), for s in S, and such that if (s) and (tl...n) are in V for any n, then s ≠ t; and if (sl...n) and (tl...m) are in V for any m, n, then s ≠ t.

Factorization of Phrases. Given a set V of phrases in any morphology, the set of V-factorizations is defined recursively by

1) if x is a closed member of V or is a blank, then the one-tuple (x) is a V-factorization;

2) if x in V is of positive degree n and $F_1,\ldots,F_n$ are V-factorizations, then the tuple $(x, F_1,\ldots,F_n)$ is a V-factorization.

The product $\overline{F}$ of a V-factorization F is defined recursively by

3) if x is a closed member of V or is a blank, then $\overline{(x)} = x$;

4) if $(x,F_1,\ldots,F_n)$ is a V-factorization, then $\overline{(x,F_1,\ldots,F_n)} = x\cdot(\overline{F}_1*\ldots*\overline{F}_n)$.

If $\overline{F} = x$, then F is said to be a V-factorization of x. It is easy to see that if V is a vocabulary for a morphology, then every phrase has at least one V-factorization. If each phrase has just one V-factorization, call the vocabulary monotectonic. Otherwise the vocabulary is polytectonic. A morphology which has a monotectonic vocabulary is a monotectonic morphology; otherwise it is polytectonic.

If an expression x is such that, for some n sufficiently large, $x\cdot(1*\ldots*(i-1)*y*(i+1)*\ldots*n) = x$ for every phrase y, then x will be said to be free of the i-th blank. The number of blanks in an expression is n - k, where n is the degree and k is the number of blanks, among the first n, of which the expression is free. An expression is initialized if the number of blanks in it is the same as its degree.

The following useful facts about morphologies are easily established. We will denote the dimension of x by dim (x), and the degree of x by deg (x).

Lemma 2.5: 1) For all x, y, $\dim(x\cdot y) = \dim(x)$,
2) $\deg(x\cdot y) \leq \deg(y)$.
3) If x is closed, $x' = x$.
4) If $\deg x \leq n$, then $x\cdot(1*\ldots*n) = x$.

Lemma 2.6: For all x, y,
1) $\dim(x*y) = \dim(x) + \dim(y)$,
2) $\deg(x*y) = \max\{\deg(x), \deg(y)\}$,
3) $(x*y)' = x'*y'$, and

4) if dim (x) = k, and n ≠ k, then $(1*\ldots*n)\cdot x \neq x$.

Lemma 2.7: For all phrases $x_i$, $y_j$, for $1 \leq i \leq k$, $1 \leq j \leq m$, $x_1*\ldots*x_k = y_1*\ldots*y_m$ if and only if k = m and $x_i = y_i$ for i = 1,2,...,k.

Lemma 2.8: For an expression x of degree n, let $j_1, j_2, \ldots, j_n$ denote those blanks of which x is not free. Then for any expressions y and z such that

$$j_r \cdot y = j_r \cdot z \text{ for } r = 1,2,\ldots,k,$$

we have that

$$x \cdot y = x \cdot z.$$

Lemma 2.9: For any expression x, deg (x) = n if and only if (i) for m>n, x is free of the m-th blank, and (ii) x is not free of the n-th blank.

Lemma 2.10: In a linear morphology, if x is initialized, of degree n>0, then for any expression y,

1) deg $(x \cdot y)$ = deg $[(1*\ldots*n)\cdot y]$.

2) deg $(x \cdot y)$ = max {deg $(n \cdot y)$ | x is not free of the n-th blank}.

Theorem 2.11: For every expression x there exist elements y and z of the blank-morphology such that $x \cdot y$ is initialized and $(x \cdot y) \cdot z = x$. Hence each vocabulary for a morphology may be replaced in a one-one fashion by a vocabulary whose members are initialized.

Proof: Let x be any expression of degree n. If x is initialized, the theorem is trivially satisfied by y = z = $1*\ldots*n$. If x is not initialized, then suppose the number of blanks in x is n-k. Then x is free of k blanks which we denote by the $i_1$-th,...,$i_k$th.

Let p be any permutation of the integers 1,2,...,n, such that $n-k+1 \leq p(i_j) \leq n$ for j = 1,2,...,k. Denote p(i) by $p_i$ and $p^{-1}(i)$ by $p'_i$, for i = 1,2,...,n. Let $y = p_1*\ldots*p_n$ and let $z = p'_i*\ldots*p'_n$. Then y and z each belong to the blank-morphology, and $(x \cdot y) \cdot z = x \cdot (y \cdot z) = x \cdot (1*\ldots*n) = x$, as required.

It remains to show that $x \cdot y$ is initialized. We will establish that $x \cdot y$ is free of the i-th blank for $i \neq n-k$ and not free of the i-th blank for $i = n-k$. Then by Lemma 2.9 we may conclude that $x \cdot y$ is initialized, of degree n-k.

Suppose that $i>n$. By Lemmas 2.5 and 2.6, $\deg(x \cdot y) \leq \deg(y) = n$; hence by Lemma 2.9, $x \cdot y$ is free of the i-th blank.

Now consider, for any phrase w, the expression

$$(x \cdot y) \cdot (1 * \ldots * i-1 * w * i+1 * \ldots * n)$$
$$= x \cdot (p_1 * \ldots * p_n) \cdot (1 * \ldots * i-1 * w * i+1 * \ldots * n)$$
$$= x \cdot (q_1 * \ldots * q_n),$$

where
$$q_j = \begin{cases} p_j, & \text{for } p_j \neq i \\ w, & \text{if } p_j = i. \end{cases}$$

By the construction, if $i$ n-k, then $p_j = i$ implies that x is free of the j-th blank. Hence we have

$$x \cdot (q_1 * \ldots * q_n) = x \cdot (p_1 * \ldots * p_n)$$
$$= x \cdot y,$$

and $x \cdot y$ is free of the i-th blank.

If $1 \leq i \leq n-k$, and $x \cdot y$ is free of the i-th blank, then by the construction, x is not free of the $p_i'$-th blank. Choose a phrase w such that $x \cdot (1 * \ldots * p_i'-1 * w * p_i'+1 * \ldots * n) \neq x$. Define y' and z' as follows:

$$y' = \begin{cases} y, & \text{if } \deg(w) \leq n \\ y * n+1 * \ldots * \deg(w), & \text{if } \deg(w) > n \end{cases}$$

$$z = \begin{cases} z, & \text{if } \deg(w) \leq n \\ z * n+1 * \ldots * \deg(w), & \text{if } \deg(w) > n. \end{cases}$$

Then $x = x \cdot y \cdot z = x \cdot y \cdot z'$
$= x \cdot y \cdot (1 * \ldots * i-1 * w \cdot y' * i+1 * \ldots * n) \cdot z'$
$= x \cdot y' \cdot (1 * \ldots * i-1 * w \cdot y' * i+1 * \ldots * n) \cdot z'$
$= x \cdot (q_1 * \ldots * q_n),$

where

$$q_j = \begin{cases} j, \text{ if } p_j \neq i \\ w \cdot y' \cdot z', \text{ if } p_j = i. \end{cases}$$

But $y' \cdot z' = (1 * \ldots * m)$, where $m = \deg(w)$, so $w \cdot y' \cdot z' = w$ and $x \cdot (1 * \ldots * p_i' - 1 * w * p_i' + 1 * \ldots * n) = x$, a contradiction. Hence $x \cdot y$ is free of precisely those blanks $m > n-k$, and is initialized of degree $n-k$.

Theorem 2.12: Every member of a monotectonic vocabulary is already initialized.

Proof: Suppose x is any expression of degree n in a monotectonic vocabulary V, and x is free of the i-th blank. Then $x = x \cdot (1 * \ldots * n)$, and $(x, (1), \ldots, (n))$ is a factorization of x. But $x = x \cdot (1 * \ldots * i-1 * x * i+1 * \ldots * n)$, hence $(x, (1), \ldots, (i-1), (x), (i+1), \ldots, (n))$ is a second factorization of x, a contradiction.

From now on, by vocabulary we will mean initialized vocabulary.

An element $(j_1 * \ldots * j_n)$ of the blank-morphology is called a permutation if $j_i = p(i)$, $i = 1,2,\ldots,n$, where p is some permutation of the integers $1,2,\ldots,n$.

Theorem 2.13: Given two initialized vocabularies W' and W for a monotectonic morphology, for each morpheme W' and W there is a unique morpheme w' in W' and a permutation p such that $w' = w \cdot p$. Thus a monotectonic morphology has essentially one vocabulary, and all vocabularies in a monotectonic morphology are monotectonic.

Proof: Let V be a monotectonic vocabulary for the morphology. We will establish the result when W' = V, from which the theorem follows immediately. Suppose v, of degree n, is in V. Then v has a W-factorization $F = (w, F_1, \ldots, F_n)$. Denote $\overline{F}_1 * \ldots * \overline{F}_n$ by $\hat{F}$. Then $v = w \cdot \hat{F}$. Similarly, for a V-factorization $G = (v', G_1, \ldots, G_k)$, $w = v' \cdot \hat{G}$, where $\deg(v') = k$. Hence we have

$$v = v' \cdot \hat{G} \cdot \hat{F}$$
$$= v' \cdot (1 * \ldots * k) \cdot \hat{G} \cdot \hat{F}.$$

For i = 1,2,...,k, let $H_i$ be the (unique) V-factorization of $i\cdot(\hat{G}\cdot\hat{F})$. Then h = (v', $H_1$,...,Hk) is a V-factorization of v; and since (v, (1),...,(n)) is a V-factorization of v, we have n = k, v = v', and $\overline{H}_i$ = i for i = 1,2,...,n. Since $i\cdot\hat{G}\cdot\hat{F} = H_i$, we have $\hat{G}\cdot\hat{F}$ = (1*...*n).

Now in a monotectonic morphology, if, for some expression x,y, and some integer m, x·y = m, then x = n for some integer n, and n·y = m. For, suppose F = (v",$F_1$,...,$F_k$) is a factorization of x, where v" is a morpheme. Then

$$\begin{aligned} x\cdot y &= v''\cdot(\overline{F}_1*\ldots*\overline{F}_k)\cdot y \\ &= v''\cdot(\overline{F}_1\cdot y*\ldots*\overline{F}_k\cdot y). \end{aligned}$$

Let $R_i$ be the factorization of $\overline{F}_i\cdot y$, for $1\leq i\leq k$. Then (v", $R_1$,...,$R_k$) is a factorization of m, as in (m). Hence x is either closed or a blank; since x·y = m, x is not closed, so x is a blank.

Now it follows readily that G and F and permutations $p^{-1}$ and p respectively and v = w·p.

To establish the uniqueness of w, suppose $v = w'\cdot p'$ for some permutation p' and some w' in W. Then

$$\begin{aligned} w\cdot p &= w'\cdot p' \\ w\cdot p\cdot p^{-1} &= w'\cdot p'\cdot p^{-1} \\ w &= w'\cdot(p'\cdot p^{-1}), \end{aligned}$$

and w = w', by the minimality of a vocabulary.

Theorem 2.14: If a morphology with vocabulary V is monotectonic, then it is isomorphic to a Lukasiewicz morphology over V as a set of symbols. Conversely every morphology isomorphic to a Lukasiewicz morphology is monotectonic.

Proof: Let M be monotectonic with vocabulary V. Let θ be the map which makes correspond to each morpheme v in V of degree k the element v1...k of the linear morphology over the set of symbols $S = \{v \mid v \in V\}$. Let M' be the Lukasiewicz submorphology generated by the set θ(V). Extend θ to M as follows: for a factorization F = (x, $F_1$,...,$F_k$), where

$\overline{F} = y$, define $\theta(y)$ to be $\theta(x)\cdot(\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_k))$. Since V is monotectonic, $\theta$ is well-defined, and is easily seen to be an isomorphism onto M'.

To show the converse, first we show a special property of phrases in a Lukasiewicz morphology. Define a partial order on the phrases in M', a Lukasiewicz morphology over the set of symbols S: $x \leq y$ if $x = x_1 \ldots x_k$, $y = y_1 \ldots y_r$ for $x_i$ and $y_j$ in S $\cup$ N, $1 \leq i \leq k$, $1 \leq j \leq r$, and for $1 \leq i \leq k$, $x_i = y_i$. (This makes $r \geq k$ a necessary condition.)

The property is this: $x \leq y$ if and only if $x = y$. We prove the nontrivial part of this assertion by induction on the length r of y. If $r = 1$, then $y = y_1 = x_1 = x$. Suppose the hypothesis is true for y of length no greater than r, and suppose the length of y is r+1.

Let $F = (v, F_1, \ldots, F_n)$ be a V-factorization of y, and let $G = (w, G_1, \ldots, G_m)$ be a V-factorization of x. Then $v = s1 \ldots n$ and $w = t1 \ldots m$ for some s, t in S, and some non-negative integers n, m. Hence, by the rules of composition, $y_1 = s$ and $x_1 = t$; since $x \leq y$ and the length of x is at least one, we have $s = t$; since v and w are both in V, we must have $v = w$, and $n = m$.

If $n = m = 0$, then $y = v = w = x$. If $n \geq 1$, then $y = (s1 \ldots n)\cdot(\overline{F}_1 * \ldots * \overline{F}_n) = s\overline{F}_1 \ldots \overline{F}_n$, and $x = (s1 \ldots n)\cdot(\overline{G}_1 * \ldots * \overline{G}_n) = s\overline{G}_1 \ldots \overline{G}_n$, and either $\overline{F}_1 \leq \overline{G}_1$ or $\overline{G}_1 \leq \overline{F}_1$. In either case, by the induction hypothesis, $\overline{F}_1 = \overline{G}_1$, since the length of each is less than r+1. Now suppose that for $i \leq j$, $\overline{F}_i = \overline{G}_i$. Then either $\overline{F}_{j+1} \leq \overline{G}_{j+1}$ or $\overline{G}_{j+1} \leq \overline{F}_{j+1}$. In either case, $\overline{F}_{j+1} = \overline{G}_{j+1}$, since the length of each is less than r+1. So for all j, $1 \leq j \leq n$, $\overline{F}_j = \overline{G}_j$; hence $y = x$. This completes the proof of the property as claimed.

Now let x be a phrase in M' with factorizations $F = (v_1, F_1, \ldots, F_r)$ and $G = (v_2, G_1, \ldots, G_m)$. By an argument in the proof above, we have $v_1 = v_2 = (s1 \ldots r)$ for some s

in S, $r \geq 0$, and $r = m$. Since $x = S\overline{F}_1 \ldots \overline{F}_r = S\overline{G}_1 \ldots \overline{G}_r$, either $\overline{F}_1 \leq \overline{G}_1$ or $\overline{G}_1 \leq \overline{F}_1$; since $\overline{F}_1$ and $\overline{G}_1$ are phrases, $\overline{F}_1 = \overline{G}_1$ by the property established above. Suppose, for $i \leq j$, $F_i = G_i$; then $\overline{F_{j+1}}$ and $\overline{G_{j+1}}$ are comparable, hence equal. Then for all j, $1 \leq j \leq r$, $\overline{F}_j = \overline{G}_j$. We now complete the proof by induction on the depth of a factorization, defined as follows:

(1) If F = (v) for v in V, or F = (n) for a blank n, then F has depth zero.

(2) If $F = (v, F_1, \ldots, F_n)$, then depth (F) = $\max_{1 \leq j \leq n} \{\text{depth } (F_j)\} + 1$.

Suppose that max {depth (F), depth (G)} = 0. Then (1) G = (s) for some s in $V \cap S$ or (2) G = (n) for a blank n. In case (1), x = s, hence F = (s) = G, since s clearly has only one factorization; in case (2), again blanks have only one factorization, so F = (n) = G.

Suppose that for max {depth (F), depth (G)} $\leq n$, F = G, and consider the case when max {depth (F), depth (G)} = n+1. Then $G = ((sl \ldots r), G_1, \ldots, G_r)$, where depth $(G_i) \leq n$, $1 \leq i \leq r$, and $F = ((sl \ldots r), F_1, \ldots, F_r)$, where depth $(F_i) \leq n$, $1 \leq i \leq r$, and $F_i = G_i$, $1 \leq i \leq r$. Since for $1 \leq i \leq r$, max {depth $(F_i)$, depth $(G_i)$} = n, and $F_i$, $G_i$ are two factorizations of $\overline{F}_i = \overline{G}_i$, then by the induction hypothesis, $F_i = G_i$. Hence G = F and the proof is complete.

By an interpretation of a morphology A in a morphology B we mean a homomorphism of A into B, i.e., a mapping $\theta$: $A \to B$: $x \to x^\theta$ which preserves operations: $\pi^\theta = \pi$, $x'^\theta = x^{\theta'}$, $(x*y)^\theta = x^\theta * y^\theta$, and $(x \cdot y)^\theta = x^\theta \cdot y^\theta$, for all x, y in A. We shall refer to the image of A in B also as the "interpretation of A" (under $\theta$) and will call A a formulation of its image. Thus to say that one morphology can be formulated in another is to say that there is an interpretation mapping the latter onto the former.

Theorem 2.15: Under any interpretation of one morphology in another, the blank-morphology of the first maps isomorphically onto that of the other.

Proof: By Theorems 2.4 and 2.5, and the requirement that $\pi^{\theta} = \pi$ (i.e., $\theta(1) = 1$), any interpretation is either an isomorphism on the blank-morphology, or the constant map $h_n$, for some natural number n. However, in the latter case, we have in the image morphology

$$\pi = \pi^{\theta} = \pi'^{\theta} = \pi',$$

which contradicts the requirement that in a morphology, $\pi' \neq \pi$.

It is easily shown that the dimension of an expression is always preserved under an interpretation, and the degree is never increased. However, ...ee may decrease, as shown by the example whic... ...em 2.16.

A mapping of a sub... ...ogy into another is conservative if it p... ... and does not increase degree. A mor... ...generated by a vocabulary if every cons... ... of that vocabulary can be extended to an ...ion of the whole morphology. A morphology is free if it possesses a vocabulary by which it is freely generated.

Theorem 2.16: The free morphologies are precisely those which are isomorphic to Lukasiewicz morphologies. Hence a morphology is free if and only if it is monotectonic.

Proof: Let $M = (M, *, \cdot, \pi, ')$ be freely generated by V. Let M' be the Lukasiewicz morphology generated by the set $\theta(V)$ constructed in the proof of Theorem 2.14. Note that $\theta$ is conservative. Since M is free, $\theta$ can be extended to a homomorphism $\theta: M \to M'$. We will show that $\theta$ is an isomorphism. Clearly $\theta$ is onto, since $\theta(V)$ is a vocabulary for M'. Suppose $\theta(x) = \theta(y)$ for some phrases x,y in M, $x \neq y$. Let n be the least non-negative integer such that there are x, y in M, $x \neq y$, $\theta(x) = \theta(y)$, and there is a factorization

Theorem 2.15: Under any interpretation of one morphology in another, the blank-morphology of the first maps isomorphically onto that of the other.

Proof: By Theorems 2.4 and 2.5, and the requirement that $\pi^{\Theta} = \pi$ (i.e., $\Theta(1) = 1$), any interpretation is either an isomorphism on the blank-morphology, or the constant map $h_n$, for some natural number n. However, in the latter case, we have in the image morphology

$$\pi = \pi^{\Theta} = \pi'^{\Theta} = \pi',$$

which contradicts the requirement that in a morphology, $\pi' \neq \pi$.

It is easily shown that the dimension of an expression is always preserved under an interpretation, and the degree is never increased. However, the degree may decrease, as shown by the example which follows Theorem 2.16.

A mapping of a subset of one morphology into another is conservative if it preserves dimension and does not increase degree. A morphology is freely generated by a vocabulary if every conservative mapping of that vocabulary can be extended to an interpretation of the whole morphology. A morphology is free if it possesses a vocabulary by which it is freely generated.

Theorem 2.16: The free morphologies are precisely those which are isomorphic to Lukasiewicz morphologies. Hence a morphology is free if and only if it is monotectonic.

Proof: Let $M = (M, *, \cdot, \pi, ')$ be freely generated by V. Let M' be the Lukasiewicz morphology generated by the set $\Theta(V)$ constructed in the proof of Theorem 2.14. Note that $\Theta$ is conservative. Since M is free, $\Theta$ can be extended to a homomorphism $\Theta$: $M \to M'$. We will show that $\Theta$ is an isomorphism. Clearly $\Theta$ is onto, since $\Theta(V)$ is a vocabulary for M'. Suppose $\Theta(x) = \Theta(y)$ for some phrases x,y in M, $x \neq y$. Let n be the least non-negative integer such that there are x, y in M, $x \neq y$, $\Theta(x) = \Theta(y)$, and there is a factorization

$F = (v_1, F_1, \ldots, F_m)$ of x and a factorization $G = (v_2, G_1, \ldots, G_r)$ of y such that max {depth (F), depth (G)} = n. Suppose n = 0. Then depth (F) = depth (G) = 0; we have four cases:

1) $F = (v_1)$, $G = (v_2)$ for some $v_1, v_2 \in V$

2) $F = (v_1)$, $G = (n)$ for $v_1 \in V$, $n \in N$

3) $F = (n)$, $G = (v_2)$ for $v_2 \in V$, $n \in N$

4) $F = (n)$, $G = (m)$ for $n, m \in N$.

In case 1, $x = v$, $y = v_2$; hence $\theta(v_1) = \theta(v_2)$; but by the construction of $\theta(v_1)$, this implies $v_1 = v_2$, a contradiction. Cases 2 and 3 are symmetric. In case 2, $x = v_1$, $y = n$; hence $\theta(v_1) = \theta(n) = n$, by Theorem 2.15, again a contradiction of the construction. In case 4, Theorem 2.15 gives $x = \theta(x) = n$, $\theta(y) = m = y$, again a contradiction. So $n \neq 0$.

Suppose $n>0$. Then $\theta(x) = \theta(v_1)\cdot(\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_m)) = \theta(v_2)\cdot(\theta(\overline{G}_1)*\ldots*\theta(\overline{G}_r))$. For $1 \leq i \leq m$, let $F_i'$ be a $\theta(V)$-factorization of $\theta(\overline{F}_i)$, and for $1 \leq j \leq r$, let $G_j'$ be a $\theta(V)$-factorization of $\theta(\overline{G}_j)$. Then $F' = (\theta(v_1), F_1', \ldots, F_m')$ and $G' = (\theta(v_2), G_1', \ldots, G_r')$ are two factorizations of $\theta(x)$. Since M' is monotectonic, $\theta(v_1) = \theta(v_2)$, $m = r$, and for $1 \leq i \leq m$, $F_i' = G_i'$ and $\overline{F}_i' = \theta(\overline{F}_i) = \theta(\overline{G}_i) = G_i'$. Suppose $\overline{F}_i \neq \overline{G}_i$; but max {depth $(F_i)$, depth $(G_i)$} $\leq n-1$, contradicting the minimality of n. So $\overline{F}_i = \overline{G}_i$. Also, by the construction of $\theta(V)$, we see that $v_1 = v_2$. Hence $x = \overline{F} = \overline{G} = y$, another contradiction. So there are no phrases x,y in M, $x \neq y$, such that $\theta(x) = \theta(y)$. Now by applying Lemma 2.7, we see that $\theta$ is 1-1, and hence an isomorphism.

If M is a Lukasiewicz morphology, then it is monotectonic, by Theorem 2.14. Let M' be any morphology, $\theta$ any conservative map on V such that $\theta(V)$ M'. For blanks n in M, let $\theta(n) = n$. For non-blank phrases x in M, extend $\theta$ as follows: let $(v, F_1, \ldots, F_n)$ be the unique

factorization of x. Then $\theta(x) = \theta(v)\cdot(\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_n))$.

For arbitrary expressions $x = x_1*\ldots*x_n$ in M, where each $x_i$ is a phrase, let $\theta(x) = \theta(x_1)*\theta(x_2)*\ldots*\theta(x_n)$. This extension of $\theta$ is well-defined. From the construction of $\theta$, we have immediately that for all x, y in M, $\theta(x*y) = \theta(x)*\theta(y)$, and $\theta(1) = 1$.

If $x = x_1*\ldots*x_n$, where the $x_i$ are phrases, then

$\theta(x') = \theta(x_1'*\ldots*x_n')$, by Lemma 2.1,

$= \theta(x_1')*\ldots*\theta(x_n')$, by the construction,

$[\theta(x)]' = \theta(x_1)'*\ldots*\theta(x_n)'$; hence $\theta(x') = \theta(x)'$ for all x in M if and only if $\theta(y') = \theta(y)'$ for all phrases y in M. Suppose there is a phrase y in M such that $\theta(y') \neq [\theta(y)]'$. Let n be the least integer such that there is a y, $\theta(y') \neq \theta(y)'$ and the factorization $F = (v, F_1,\ldots,F_r)$ of y has depth n. If $n = 0$, then (i) $F = (v)$, $y = v$, where v is closed or (ii) $F = (n)$ for some blank n. In case (i), $v' = (v\cdot 1)' = v\cdot 1' = v$. Hence $\theta(v) = \theta(v')$. Since $\theta$ does not increase degree, deg $(\theta(v)) = 0$. Hence $\theta(v)' = \theta(v)$, giving a contradiction. In case (ii), $y' = n+1$, $\theta(y) = n$, $\theta(y) = n+1 = \theta(y')$, another contradiction. Suppose $n>0$. Then suppose deg $(y) = s$.

$$y = v\cdot(\overline{F}_1*\ldots*\overline{F}_r)$$
$$= v\cdot(\overline{F}_1*\ldots*\overline{F}_r)\cdot(1*\ldots*s);$$
$$y = v\cdot(\overline{F}_1*\ldots*\overline{F}_r).$$
$$\theta(y) = \theta(v)\cdot(\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_r))$$
$$\theta(y)' = [\theta(v)\cdot(\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_r))]'$$
$$= \theta(v)\cdot(\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_r))'$$
$$= \theta(v)\cdot(\theta(\overline{F}_1)'*\ldots*\theta(\overline{F}_r)')$$
$$= \theta(v)\cdot(\theta(\overline{F}_1')*\ldots*\theta(\overline{F}_r')), \text{ by the minimality of n.}$$
$$y' = v\cdot(\overline{F}_1*\ldots*\overline{F}_r)'$$
$$= v\cdot(\overline{F}_1'*\ldots*\overline{F}_r').$$

Let $G_i$ be a factorization of $\overline{F}_i'$, $1\leq i\leq r$; then $(v, G_1,\ldots,G_r)$ is a (hence the unique) factorization of y', and

$\theta(y') = \theta(v)\cdot(\theta(\overline{F}_1')*\ldots*\theta(\overline{F}_r'))$

$= \theta(y)'$, a contradiction. Hence

$\theta(x') = \theta(x)'$, for all x in M.

To show that $\theta(x\cdot y) = \theta(x)\cdot\theta(y)$ for all x,y in M, it will suffice to restrict x to phrases. If equality fails, let n be the least integer such that there is a phrase x and an element y in M, $\theta(x\cdot y) \neq \theta(x)\cdot\theta(y)$, and the factorization $F = (v, F_1,\ldots,F_r)$ of x has depth n. If n = 0, (i) F = (v), $v \in V$, v closed, or (ii) F = (m), $m \in N$. In case (i), $(x\cdot y) = x$, hence $\theta(x\cdot y) = \theta(x)$. Since $\theta$ is conservative on V, deg $(\theta(x)) = 0$, hence $\theta(x)\cdot\theta(y) = \theta(x) = \theta(x\cdot y)$, a contradiction. In case (ii), suppose $y = y_1*\ldots*y_k$, for some integer $k>0$, where $y_i$ are phrases, $1\leq i\leq k$. Then $x\cdot y = y_{\underline{m}}$, where $\overline{m} = m(\mathrm{mod}\ k)$. $\theta(x\cdot y) = \theta(y_{\underline{m}})$.

$$\theta(x)\cdot\theta(y) = m\cdot[\theta(y)] = m\cdot[\theta(y_1)*\ldots*\theta(y_k)]$$
$$= \theta(y_{\underline{m}})$$
$$= \theta(x\cdot y), \text{ a contradiction.}$$

If $n>0$, $\theta(x\cdot y) = \theta[v\cdot(\overline{F}_1*\ldots*\overline{F}_r)\cdot y]$

$= \theta[v\cdot(\overline{F}_1\cdot y*\ldots*\overline{F}_r\cdot y)]$;

let $G_i$ be the factorization of $\overline{F}_i\cdot y$, $1\leq i\leq r$. Then $(v, G_1,\ldots,G_r)$ is a (hence the unique) factorization of $x\cdot y$, and

$\theta(x\cdot y) = \theta(v)\cdot[\theta(\overline{G}_1)*\ldots*\theta(\overline{G}_r)]$

$= \theta(v)\cdot[\theta(\overline{F}_1\cdot y)*\ldots*(\overline{F}_r\cdot y)]$

$= \theta(v)\cdot[\theta(\overline{F}_1)\cdot\theta(y)*\ldots*\theta(\overline{F}_r)\cdot\theta(y)]$

by the minimality of n, $= \theta(v)\cdot[\theta(\overline{F}_1)*\ldots*\theta(\overline{F}_r)]\cdot\theta(y)$

$= \theta(x)\cdot\theta(y)$, a contradiction.

Hence for all x,y in M, $\theta(x)\cdot\theta(y) = \theta(x\cdot y)$, and $\theta$ is a homomorphism as required, and M is free.

Corollary 2.17: Every morphology is the interpretation of some free morphology. Thus every morphology has a monotectonic formulation.

Proof: Given a morphology M, with vocabulary V, let $S = \{\bar{v} | v \in V\}$ be a set of distinct symbols. Let $W = \{\bar{v}1...n | \deg(v) = n\}$, and let M' be the Lukasiewicz morphology generated by W. The correspondence $\theta(\bar{v}1...n) = v$ gives a conservative map on W, and the theorem above extends $\theta$ to the desired homomorphism.

For a morphology M, we will call the morphology M' of Corollary 2.17 the free morphology associated with M.

Example 2.18: Now we can easily construct an example of a homomorphism which decreases degree. Let M and N be the Lukasiewicz morphologies generated by V = {a, b1, c12}, and W = {a, b, c1} respectively. Let $\theta(a) = a$, $\theta(b1) = b$, $\theta(c12) = c1$. Since M is free, $\theta$ can be extended to a homomorphism which decreases the degree of b1 and c12.

Example 2.19: It is worth pointing out that it is necessary to make the restriction on the vocabulary of Lukasiewicz morphologies that if a1...n and b1...r are in V for $n,r \geq 0$, then $a \neq b$. For consider the linear morphology M generated by V = {s12, s1, a1, a}. V is not monotectonic, since the expression "saa" has factorizations $F_1$ = (s12, (a), (a)) and $F_2$ = (s1, $G_2$) where $G_2$ = (a1, (a)). Since V is reduced, then by Theorem 2.13, if M is monotectonic, V must be; hence M is not monotectonic, not free, not Lukasiewicz.

Example 2.20: This example shows that not every submorphology of a free morphology is free. Let M be the free (Lukasiewicz) morphology generated by V = {s12, a1, b}. Let A be the submorphology generated by W = {s12, salb, ab, b}. Note that W is a vocabulary, each of whose elements is in M. But W is not monotectonic, for:

$$\begin{aligned} sabb &= s12 \cdot (ab * b) \\ &= salb \cdot b \end{aligned}$$

Hence sabb has two factorizations, and M is not free.

# CHAPTER III

## GRAMMATICAL SETS

Half-ring grammars. Let C, K, and S be symbols called composition, concatenation, and shift, respectively. Let A be a finite alphabet of symbols distinct from C, K, and S. A contains a subset W of terminals; the other elements are called variables. For any set B, we define T(B), the set of terms over B, as the least set T such that

(i) $B \subset T$.

(ii) If $t \in T$, $St \in T$.

(iii) If $t, u \in T$, $Ctu \in T$.

(iv) If $t, u \in T$, $Ktu \in T$.

(Juxtaposition here denotes juxtaposition.)

We are interested in subsets of T(A), generated in a way we explain next.

Some familiarity with context-free languages [7] will be assumed. However, for completeness a definition is included. The notation differs slightly from that in [7].

A context-free grammar is a 4-tuple $G = (V,\Sigma,P,\sigma)$, where

(i) V is a finite alphabet of variables.

(ii) $\Sigma$ is a finite alphabet of terminals.

(iii) $\sigma \in \Sigma$.

(iv) P is a finite collection of ordered pairs called rewriting rules (also called productions) of the form $\alpha \to \beta$, where $\alpha \in V$ and $\beta \in (V \cup \Sigma)^*$.

[Definition: For any set of symbols B, the Kleene closure of B, denoted by $B^*$, is defined as follows: $B^0 = \{\varepsilon\}$,

where $\varepsilon$ denotes the empty string of symbols; $B^1 = B$; for $n>1$, $B^n = B \cdot B^{n-1} = \{xy \mid x \in B,\ y \in B^{n-1}\}$. Then $B^* = \bigcup_{n\geq 0} B^n$. When we wish to exclude the empty string, we write $B^+ = \bigcup_{n\geq 1} B^n$.]

We define the relations $\rightarrow$ and $\Rightarrow$ for $x,y$ in $(V \cup \Sigma)^*$ as follows:

(1) $x \rightarrow y$ if $x = u\alpha v$, $y = u\beta v$, and $\alpha \rightarrow \beta \in P$, for some $u,v \in (V \cup \Sigma)^*$.

(2) $x \Rightarrow y$ if there is a finite (possibly empty) sequence $x = x_0, x_1, \ldots, x_k = y$ such that for $0 \leq i \leq k-1$, $x_i \rightarrow x_{i+1}$.

Then the context-free language generated by $G$ is defined as the collection of strings $L(G) = \{x \text{ in } \Sigma^* \mid \sigma \Rightarrow x\}$.

If, for any strings of symbols $x$ and $y$ of variables in $V$ and terminals in $\Sigma$, we have $x \Rightarrow y$, then we say that $x$ yields $y$. Any sequence $x = x_0, x_1, \ldots, x_k = y$ satisfying (2) is called a derivation of $y$ from $x$. If $x_0 = \sigma$, then we will often call the sequence simply a derivation of $y$. The integer $k$ is the length of the derivation. A leftmost derivation is a sequence satisfying (2), with the added property that $x_i \rightarrow x_{i+1}$ by the application of a production to the leftmost variable appearing in $x_i$, for $1 \leq i \leq k-1$. It is well-known that, in any context-free language, if $x$ yields $y$, then $x$ yields $y$ by a leftmost derivation. Hence proofs will often consider only leftmost derivations.

Suppose that

$$(*) \qquad x = x_0 \xrightarrow[p_1]{} x_1 \xrightarrow[p_2]{} \cdots \xrightarrow[p_k]{} x_k = y$$

is a leftmost derivation, where the $p_j$ represent the productions applied at each step. Then suppose that for some $x_i$, some $x_{i+j}$, where $j \geq 0$, $x_i = u\alpha v$, where $\alpha$ is the leftmost terminal in $x_i$ and

$$x_i = u\alpha v \xrightarrow[p_{i+1}]{} uz_1 v \xrightarrow[p_{i+2}]{} uz_2 v \rightarrow \cdots \xrightarrow[p_{i+j}]{} uz_j v = x_{i+j}.$$

Then we will call the derivation

(**) $$\alpha \xrightarrow[p_{i+1}]{} z_1 \xrightarrow[p_{i+2}]{} z_2 \to \cdots \xrightarrow[p_{i+j}]{} z_j$$

a subderivation of (*). We remark that (**) is also a leftmost derivation.

A <u>half-ring grammar</u> G is a context-free grammar satisfying, for some finite alphabet A, where $W \subset A$,

(i) $\Sigma = W \cup \{C,K,S\}$.

(ii) $V = A \setminus W$.

(iii) for each production $\alpha \to \beta$ in P, $\beta \in T(A)$.

From now on, <u>grammar</u> will mean half-ring grammar, and since the symbols C,K,S always appear in $\Sigma$, we will denote G by the 4-tuple $(V,W,P,\sigma)$, where it is understood that $W \cup \{C,K,S\} = \Sigma$. We will call $L(G)$, where G is a half-ring grammar, a <u>recognizable set</u>. A <u>string</u> in G will be any finite sequence (represented by juxtaposition) of elements in $V \cup W \cup \{C,K,S\}$. A <u>terminal string</u> will consist only of elements in $W \cup \{C,K,S\}$.

<u>The generic algebra $\mathcal{J}_n$</u>. For any positive integer n, let $W_n = \{w_1,\ldots,w_n\}$ be a collection of distinct symbols. Let $J_n = T(W_n)$. Then $\mathcal{J}_n = (J_n,\overline{C},\overline{K},\overline{S})$ is the generic algebra on n symbols, where C,K are binary operations and S is a unary operation, defined by

$$\text{for } t_1, t_2 \in J_n,\ \overline{C}(t_1,t_2) = Ct_1t_2$$
$$\overline{K}(t_1,t_2) = Kt_1t_2$$
$$\overline{S}(t_1) = St_1.$$

This is the algebra, unique up to isomorphism, of which every algebra of the same species and generated by a copy of $J_n$ is a homomorphic image. Where no confusion will result, we will not differentiate between the symbols for the operations $\overline{C},\overline{K},\overline{S}$ and the symbols C,K,S. Note that if $W_n$ is the collection of terminals for a grammar G, then $L(G) \subset J_n$.

Grammatical sets. Now let M be a morphology, and let $B = \{b_1,\ldots,b_n\}$, $n>0$, be an ordered collection of phrases in M. Let $\hat{\eta}: W_n \to B$ be the one to one correspondence between $W_n$ and B, such that $\hat{\eta}(w_i) = b_i$, for $1 \leq i \leq n$. Let $\eta: J_n \to M$ be the (unique) homomorphic extension of $\hat{\eta}$ such that

(i) $\eta(w_i) = \hat{\eta}(w_i)$, $1 \leq i \leq n$.

(ii) $\eta(Ct_1t_2) = \eta(t_1)\cdot\eta(t_2)$, for all $t_1,t_2 \in J_n$.

(iii) $\eta(Kt_1t_2) = \eta(t_1)*\eta(t_2)$, for all $t_1,t_2 \in J_n$.

(iv) $\eta(St) = \eta(t)'$, for all $t \in J_n$.

Then given any grammar G, the image of L(G) under $\eta$ (denoted $\eta L(G)$) will be called the grammatical set (g-set) generated by G in the pair (M,B).

An alternative formulation. The use of the term "recognizable set" is motivated by a paper by Mezei and Wright [1967]. They define a recognizable set in $J_n$ as the union of congruence classes of some finite congruence R on $J_n$. As a special case of their main result, we have the important fact that the sets L(G), where G is a morphology grammar whose set of terminals W has cardinality n, are precisely these recognizable subsets of $J_n$. We will use this fact repeatedly.

It will often be convenient to use, rather than a congruence relation R itself, a collection $R = \{C_i\}_{i=1}^{r}$ of sets which are the congruence classes determined by R. We will call the partition $R = \{C_i\}_{i=1}^{r}$ itself a (finite) congruence on $J_n$ if

(1) $J_n = \bigcup_{1 \leq i \leq r} C_i$.

(2) $C_i \cap C_j = \phi$, for $1 \leq i < j \leq r$.

(3) for all i, there is a j, such that for all x in $C_i$, $Sx \in C_j$.

(4) for all pairs (i,j), $1 \leq i \leq r$, $1 \leq j \leq r$, there is a k such that for all $x \in C_i$, for all $y \in C_j$, $Cxy \in C_k$.

(5) for all pairs $(i,j)$, $1 \leq i \leq r$, $1 \leq j \leq r$, there is a $k$ such that for all $x \in C_i$, for all $y \in C_j$, $Kxy \in C_k$.

The congruence relation associated with R is, of course, defined by: $xRy$ if and only if there is an $i$, $1 \leq i \leq r$, such that $x \in C_i$ and $y \in C_i$.

As an immediate consequence of this equivalence, we know that our recognizable sets in $J_n$ are closed under finite intersection, finite union, and complementation with respect to $J_n$.

Again as a special case of Mezei and Wright's results, every non-empty g-set can be generated by a grammar $G = (V,W_n,P,\sigma)$ satisfying:

(1) If $\alpha \to \beta$ is in P, $\alpha \neq \sigma$, then $\beta$ has the form
   (i) $w_j$, $1 \leq j \leq n$
   or (ii) $C\gamma\delta$, $\gamma, \delta \in V$
   or (iii) $K\gamma\delta$, $\gamma, \delta \in V$
   or (iv) $S\gamma$, $\gamma \in V$.

(2) If $\alpha \in V$, there is an $x \in T(W_n)$ such that $\alpha \Rightarrow x$. A grammar with this property is called reduced.

(3) Suppose $L(G) = \bigcup_{1 \leq i \leq k} C_i$, for some $k \leq r$, where $R = \{C_1,\ldots,C_r\}$ is a congruence on $J_n$. Then $V = \{a_1,a_2,\ldots,a_n,\sigma\}$, where, for $1 \leq i \leq r$, $C_i = \{x \text{ in } T(W) \mid \alpha_i \Rightarrow x\}$, and $\sigma$ appears in precisely the productions $\sigma \to \alpha_i$, $1 \leq i \leq k$.

Such a grammar will be said to be in best form.

Notational conventions. We fix some notation, in order to avoid repeated qualification. $\mathcal{A}$ will denote a g-set in a pair $(M,B)$. Without explicit mention, we will associate with $\mathcal{A}$ a recognizable set $L(G)$, where $G = (V,W_n,P,\sigma)$, as well as a congruence $R = \{C_1,\ldots,C_r\}$ such that $\mathcal{A} = \eta(L(G)) = \eta(\bigcup_{1 \leq i \leq k} C_i)$.

All symbols will be subscripted and superscripted as necessary, for example $R_1 = \{C_1^1,\ldots,C_n^1\}$ and $G_1 = (V_1,W_n,P_1,\sigma_1)$.

If $A = \{A_1,\ldots,A_r\}$ and $B = \{B_1,\ldots,B_n\}$ are collections of sets, then we denote by $A \wedge B$ the collection $\{A_i \cap B_j \mid 1 \leq i \leq r, 1 \leq j \leq n\}$.

If A is a finite set, $|A|$ denotes the cardinality of A.

In a morphology $(M,*,\cdot,',\pi)$, we will denote $\pi$ by 1, $\pi'$ by 2, etc.

Lemma 3.1: If $\delta$ is a g-set in (M,A), and B is an ordered set containing precisely the elements of A, then $\delta$ is a g-set in (M,B).

Proof: Let $\phi: W_n \to W_n$ be the one-to-one correspondence such that for all $w_i$ in $W_n$, $\phi(w_i)$ is that element $w_j$ such that $a_i = b_j$. Let $\phi: J_n \to J_n$ be the unique homomorphism determined by $\phi$. In fact, $\phi$ is an isomorphism. Let $R = \{C_1,\ldots,C_r\}$ be the congruence associated with $\delta$. Then define the partition $R' = \{D_1,\ldots,D_r\}$ by: $x \in D_i$ if and only if $x = \phi(y)$ and $y \in C_i$. Then R' is a finite congruence on $J_n$, and if $\delta = n(\bigcup_{1 \leq i \leq k} C_i)$, then $\delta = n(\bigcup_{1 \leq i \leq k} D_i)$.

We will henceforth, with this lemma as justification, assume any convenient ordering of a set A over which a grammar is generated. The next lemma allows us the additional liberty of embedding A in a larger set.

Lemma 3.2: If $\delta$ is a g-set in (M,B) and $B \subset D$, where D is a finite set of phrases in M, then $\delta$ is a g-set in (M,D).

Proof: Let $\delta = n(\bigcup_{1 \leq i \leq k} C_i)$, where $R = \{C_1,\ldots,C_r\}$ is a congruence on $J_n$, and $B = \{b_1,b_2,\ldots,b_n\}$. Suppose D has m elements. By Lemma 3.1, we may assume without loss of generality that $D = \{b_1,b_2,\ldots,b_n,d_{n+1},\ldots,d_m\}$. Define a partition $R' = \{C_1,\ldots,C_r,C_{r+1}\}$ of $J_m$, where $C_{r+1} = J_m \setminus \{\bigcup_{1 \leq i \leq r} C_i\}$. It is easy to see that R' is a congruence on $J_m$, when we notice that $C_{r+1}$ consists precisely of those terms which contain at least one symbol $w_i$, for $i>n$. Then since $\delta = n(\bigcup_{1 \leq i \leq k} C_i)$, $\delta$ is a g- set in (M,D).

Theorem 3.3: If $\mathcal{S}_1$ is a g-set in (M,C) then $\mathcal{S}_1 \cup \mathcal{S}_2$ is a g-set in (M,B $\cup$ C).

Proof: By Lemma 3.2, both $\mathcal{S}_1$ and $\mathcal{S}_2$ are g-sets in (M,B $\cup$ C). Suppose $|B \cup C| = n$, and $\mathcal{S}_1 = \eta(\bigcup_{1 \le i \le k_1} C_i^1)$, where $R_1 = \{C_1^1, \ldots, C_{r_1}^1\}$, $\mathcal{S}_2 = \eta(\bigcup_{1 \le j \le k_2} C_j^2)$, where $R_2 = \{C_1^2, \ldots, C_{r_2}^2\}$, and $R_1$ and $R_2$ are congruences on $J_n$. Then let $R_3 = R_1 \hat{} R_2$; $R_3$ is a finite congruence on $J_n$.

$$\text{Let } \mathcal{S}_3 = \eta \left( \bigcup_{\substack{1 \le i \le k_1 \\ 1 \le j \le r_2}} [C_i^1 \cap C_j^2] \quad \cup \quad \bigcup_{\substack{1 \le i \le r_1 \\ 1 \le j \le k_2}} [C_i^1 \cap C_j^2] \right)$$

Then $\mathcal{S}_3$ is a g-set in (M,B $\cup$ C), and

$$\mathcal{S}_3 = \eta(\bigcup_{1 \le i \le k_1} C_i^1) \cup \eta(\bigcup_{1 \le j \le k_2} C_j^2) = \mathcal{S}_1 \cup \mathcal{S}_2.$$

In other words, the collection of g-sets in a morphology M is closed under union.

For any sets A, B in a morphology M, we make the following definitions:

(1) The composition of A and B is the set
$$CAB = \{x \cdot y \mid x \in A,\ y \in B\}.$$

(2) The concatenation of A and B is the set
$$KAB = \{x * y \mid x \in A,\ y \in B\}.$$

(3) The shift of A is the set
$$SA = \{x' \mid x \in A\}.$$

Theorem 3.4: The collection of g-sets in a morphology M is closed under composition, concatenation, and shift.

Proof: Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be g-sets in M. By Lemma 3.2, we may assume that each is a g-set in (M,B), for some $B = \{b_1, \ldots, b_n\}$, with associated congruences $R_1, R_2$ on $J_n$. We define a partition $A = \{A_1, A_2, A_3, A_4\}$ of $J_n$, where

$$A_1 = \{Cxy \mid x,\ y \in J_n\}$$

$$A_2 = \{Kxy \mid x, y \in J_n\}$$

$$A_3 = \{Sx \mid x \in J_n\}$$

$$A_4 = W_n.$$

Then $A_1 = \bigcup_{\substack{1 \le i \le r_1 \\ 1 \le j \le r_2}} D_{ij}$, where $D_{ij} = \{Cxy \mid x \in C_i^1, y \in C_j^2\}$,

$A_2 = \bigcup_{\substack{1 \le i \le r_1 \\ 1 \le j \le r_2}} E_{ij}$, where $E_{ij} = \{Kxy \mid x \in C_i^1, y \in C_j^2\}$,

$A_3 = \bigcup_{1 \le i \le r_1} F_i$, where $F_i = \{Sx \mid x \in S_i^1\}$.

Define $R_3$ by:

$R_3 = \{D_{ij} \mid 1 \le i \le r_1, 1 \le j \le r_2\} \cup \{A_2, A_3, A_4\}$. $R_3$ is a congruence on $J_n$, and

$$\mathcal{J}_3 = n\left(\bigcup_{\substack{1 \le i \le k_1 \\ 1 \le j \le k_2}} D_{ij}\right) \text{ is precisely } C\mathcal{J}_1\mathcal{J}_2.$$

Define $R_4$ by:

$R_4 = \{E_{ij} \mid 1 \le i \le r_1, 1 \le j \le r_2\} \cup \{A_1, A_3, A_4\}$. $R_4$ is a congruence on $J_n$, and

$$\mathcal{J}_4 = n\left(\bigcup_{\substack{1 \le i \le k_1 \\ 1 \le j \le k_2}} E_{ij}\right) \text{ is the g-set } K\mathcal{J}_1\mathcal{J}_2.$$

Define $R_5$ by:

$R_5 = \{F_i \mid 1 \le i \le r_1\} \cup \{A_1, A_2, A_4\}$. $R_5$ is a congruence on $J_n$, and

$\mathcal{J}_5 = n\left(\bigcup_{1 \le i \le k_1} F_i\right)$ is the g-set $S\mathcal{J}_1$.

The next lemma is used repeatedly in the proofs of Chapter 4. It is a slight variant of an exercise in [7].

The proof is straightforward and is omitted.

Lemma 3.5: Suppose, for a grammar G, for strings x,y in G, x yields y by a leftmost derivation

(*) $$x = z_0 \to z_1 \to \ldots \to z_n = y.$$

Then: (1) if x = Cab for some strings a,b, then y = Cde for strings d,e, such that a yields d and b yields e, each by a subderivation of (*).

(2) if x = Kab for some strings a,b, then y = Kde for strings d,e, such that a yields d and b yields e, each by a subderivation of (*).

(3) if x = Sa for some string a, then y = Sd for some string d, such that a yields d by a subderivation of (*).

For a set A in a morphology M, we define the set T(A) of terms over A in M as the least set $T \subset M$ such that

(1) $A \subset T$.

(2) If $t_1, t_2 \in T$, then $t_1 \cdot t_2 \in T$.

(3) If $t_1, t_2 \in T$, then $t_1 * t_2 \in T$.

(4) If $t_1 \in T$, then $t_1' \in T$.

Theorem 3.6: If $\mathcal{S}_1$ is a g-set in (M,B), then the collection $T(\mathcal{S}_1)$ of terms over $\mathcal{S}_1$ is a g-set in (M,B).

Proof: It suffices to show that if L(G) is a recognizable set in $J_n$, then so is T(L(G)). Let $G = (V, W_n, P, \sigma)$ be in best form. Let $G' = (V, W_n, P', \sigma)$, where

$$P' = P \cup \{\sigma \to C\sigma\sigma,\ \sigma \to K\sigma\sigma,\ \sigma \to S\sigma\}.$$

We will show that L(G') = T(L(G)).

To show that $T(L(G)) \subset L(G')$, we show that L(G') satisfies conditions 1 through 4 above.

(1) $L(G) \subset L(G')$, since $P \subset P'$.

(2) If $t_1, t_2 \in L(G')$, then $\sigma \overset{*}{\Rightarrow}_{G'} t_1$, $\sigma \overset{*}{\Rightarrow}_{G'} t_2$. Hence by applying the production $\sigma \to C\sigma\sigma$, we have $\sigma \to C\sigma\sigma \Rightarrow Ct_1t_2$, so $Ct_1t_2 \in L(G')$.

(3) Similarly, if $t_1, t_2 \in L(G')$, then we have $\sigma \to K\sigma\sigma \Rightarrow Kt_1t_2$, so $Kt_1t_2 \in L(G')$.

(4) And again, if $t \in L(G')$, then $\sigma \underset{G'}{\Rightarrow} t$ and we have the derivation $\sigma \rightarrow S\sigma \Rightarrow St$, so $St \in L(G')$.

Next we show that $L(G') \subset T(L(G))$. Suppose there is an $x \in L(G')$ which is not in $T(L(G))$. The proof is by induction on m, where m is the least integer such that there is such an x, and x has a derivation of length m. Suppose m = 1. Then the derivation is $\sigma \rightarrow w_j = x$ for some j, $1 \leq j \leq n$, and some production in P, since otherwise x contains nonterminals and is not in L(G'). Hence $x \in L(G) \subset T(L(G))$, a contradiction. Suppose m>1. Then we have a derivation,

(*) $$\sigma \xrightarrow[\pi_1]{} x_1 \xrightarrow[\pi_2]{} x_2 \xrightarrow[\pi_3]{} \cdots \xrightarrow[\pi_m]{} x_m = x,$$

where the $\pi_i$ are productions in P'. Since G is in best form, either

(1) $\pi_1 \in P$, in which case $\pi_1 = \sigma \rightarrow \alpha$ for some $\alpha \neq \sigma$,

or (2) $\pi_1 \notin P$, in which case $\pi_1 = \sigma \rightarrow C\sigma\sigma$, $\sigma \rightarrow K\sigma\sigma$, or $\sigma \rightarrow S\sigma$.

In case 1, because of the form of G, (in particular, $\sigma$ does not appear on the right hand side of any production), no production not in P can be applied, and x is in L(G), a contradiction.

So case (2) must hold. If $\pi_1 = \sigma \rightarrow C\sigma\sigma$, then by Lemma 3.5, $x = Cy_1y_2$, where $\sigma$ yields $y_1$ and $\sigma$ yields $y_2$ by subderivations of (*). Each of these subderivations has length no greater than m-1. By the induction hypothesis, $y_1$ and $y_2$ are in $T(L(G))$. Hence by property (4) of T, $Cy_1y_2$ is in $T(L(G))$, a contradiction. An analogous argument holds if $\pi_1 = \sigma \rightarrow K\sigma\sigma$ or $\sigma \rightarrow S\sigma$. Hence we have a contradiction, and no such m can exist. Therefore $L(G') \subset T(L(G))$. This completes the proof.

Next we show that the morphology homomorphic image of a g-set is a g-set.

Theorem 3.7: For any morphologies $M_1$, $M_2$, if $\mathcal{L}_1$ is a g-set in $(M_1, B_1)$ and $h: M_1 \rightarrow M_2$ is a homomorphism, then $h(\mathcal{L}_1)$ is a g-set in $(M_2, h(B_1))$.

Proof: Note that since h preserves dimension, $h(B_1)$ is a finite set of phrases in $M_2$. Suppose $|B_1| = n$, $h(B_1) = m$. Let $R_1 = \{C_1,\ldots,C_r\}$ be the associated congruence on $J_n$, $\mathcal{S}_1 = \eta(\bigcup_{1\leq i\leq k} C_i)$. Let $h(B_1) = \{c_1,\ldots,c_m\}$, $W_m = \{z_1,\ldots,z_m\}$; $\eta':J_m\to M_2$ the homomorphism such that $\eta'(z_i) = c_i$, $1\leq i\leq m$; let $\psi:J_n\to J_m$ be the (unique) homomorphic extension of the mapping $\psi:W_n\to W_m$ such that, for $w_j \in W_n$, $1\leq j\leq n$, $\eta'\cdot\psi(w_j) = h(w_j)$. For $1\leq k\leq r$, denote $\psi(C_k)$ by $D_k$. Let $E_0 = \{x \text{ in } J_m \mid x \notin \psi(J_n)\}$. For each non-empty subset I of $\{1,2,\ldots,r\}$, let $E_I = \{x \text{ in } J_m \mid x \text{ is in precisely the sets } C_i \text{ for } i \in I\}$. Then $R_2 = \{E_0\} \cup \{E_I \mid I \subset_{I\neq\phi} \{1,2,\ldots,r\}\}$, is clearly a partition of $J_m$. To show it is a congruence:

(1) Suppose $x \in E_0$ and $y \in E_I$ for some $I = \{n_1,\ldots,n_k\}$. Then $Cxy \in E_0$; if not, there is a $z = Cz_1z_2$ in $J_n$ such that $\psi(z) = Cxy$, $\psi(z_1) = x$ and $\psi(z_2) = y$. But then $x \notin E_0$, a contradiction. A similar argument shows that Cyx, Kxy, Kyx, and Sx are in $E_0$.

(2) If $x \in E_0$ and $y \in E_0$, again Cxy, Cyx, Kxy, Kyx, and Sx are in $E_0$ by the same argument.

(3) Suppose $x \in E_I$, $j \in E_J$. Then we claim that $Cxy \in E_H$, where H is determined as follows: for $1\leq n\leq r$, $n \in H$ if and only if there is an $i \in I$ and a $j \in J$, such that for all $t \in C_i$, for all $u \in C_i$, $Ctu \in C_n$. Note that $Cxy \in E_0$ is not possible, since x and y are in $h(J_n)$. Suppose $n \in H$. Then $x \in \psi(C_i)$, $y \in \psi(C_j)$ and $Cxy \in \psi(C_n)$. Suppose $Cxy \in \psi(C_n)$. Then there are elements $z_1,z_2$ in $J_n$ such that $Cxy = \psi(Cz_1z_2)$, $x = \psi(z_1)$, $y = \psi(z_2)$. Suppose $z_1 \in C_i$, $z_2 \in C_j$; then $i \in I$, $j \in J$, and $n \in H$. So $R_2$ is a finite congruence on $J_m$.

Let $\mathcal{S}_2 = \eta'(\bigcup_{I\ \{1,2,\ldots,k\}} D_I)$. $\mathcal{S}_2$ is then a g-set in $(M_2,h(B_1))$. To see that $h\eta = \eta'\psi$, it suffices to note that for $w_i \in W_n$, $h\eta(w_i) = \eta'\psi(w_n)$ by the definition of $\psi$. Also, it is clear that

$$I \subset \bigcup_{1,2,\ldots,k} D_I = \psi(\bigcup_{1\leq i\leq k} C_i)$$

Hence we have

$$\begin{aligned} h(S_1) &= h\eta(\bigcup_{1\leq i\leq k} C_i) \\ &= \eta'\psi(\bigcup_{1\leq i\leq k} C_i) \\ &= \eta'(\bigcup_{I \subset 1,2,\ldots,k} D_I) \\ &= \delta_2. \end{aligned}$$

Given a recognizable set L(G) in $J_n$, a <u>substitution</u> $\tau(L(G))$ is defined as follows: To each $w_j$ in $W_n$, correspond a recognizable set $L_j$. $\tau$ is a set map which corresponds to each term t in L(G) the collection of terms in $J_n$ formed by making all possible substitutions of occurrences of terms $w_j$ by terms in $L_n$. Then $\tau(L(G)) = \bigcup_{t \in L(G)} \tau(t)$. It is a well-known result in context-free languages that recognizable sets are closed under substitution.

A morphology M is <u>finitely generated</u> if it has a finite vocabulary V. In the remainder of the paper, by a morphology we will mean a locally finite, finitely generated morphology, and by a vocabulary, a finite, initialized vocabulary, unless specifically stated otherwise.

Suppose we want to discuss, for a fixed morphology M, all g-sets in (M,A) for all finite collections of phrases A. The next lemma allows us to restrict attention to g-sets in (M, V∪{1}), where V is a vocabulary for M. In what follows, we make a fixed ordering of V∪{1}, as follows: $V \cup \{1\} = \{v_1, v_2, \ldots, v_{n-1}, 1\}$, so that the associated homomorphism $\eta: \mathcal{J}_n \to M$ is specified by: $\eta(w_i) = v_i$, $1 \leq i \leq n-1$

$\eta(w_n) = 1$.

<u>Lemma 3.8</u>: If M is a morphology with vocabulary V, and $\delta$ is a g-set in (M,B), then $\delta$ is a g-set in (M, V∪{1}).

<u>Proof</u>: Since V is a vocabulary, the map $\eta: \mathcal{J}_n \to M$ is onto. Hence for each $b \in B$ there is a term t in $J_n$ such that $\eta(t) = b$. Suppose B has m elements, and $\delta = \eta(L(G))$, where $G = (U, W_m, P, \sigma)$ is a grammar in best form on $J_m$.

Define a grammar $G' = (U, W_m, P', \sigma)$, where P' contains (1) all productions in P except those of the form $\alpha \to w_j$, $1 \leq j \leq m$.

(2) For each production in P of the form $\alpha \to w_j$, where $\eta(w_j) = b$, the production $\alpha \to t$, where $\eta(t) = b$.

Let $\mathcal{L}' = \eta(L(G))$. It is easily seen that $\mathcal{L} = \mathcal{L}'$.

Theorem 3.9: Let M be a morphology with vocabulary V. Then every submorphology M' of M is a g-set in $(M, V \cup \{1\})$.

Proof: Let $V' = \{u_1, \ldots, u_{m-1}\}$ be a vocabulary for M'. Then let $\eta: J_n \to M$ be the homomorphism such that $\eta(w_i) = u_i$, $1 \leq i \leq n-1$, and $\eta(w_n) = 1$. Then $R = \{J_n\}$ is a congruence on $J_n$, and $\eta(J_n) = M'$. Hence M' is a g-set in $(M, V' \cup \{1\})$. The theorem then follows from Lemma 3.8.

Theorem 3.10: Let M be a morphology, with vocabulary V. Every g-set $\mathcal{L}$ in $(M, V \cup \{1\})$ is the homomorphic image of a g-set in a free morphology M.

Proof: Let M', with vocabulary V', be the free morphology associated with (M,V) constructed in the proof of Corollary 2.17, and let $\theta: M' \to M$ be the homomorphism of that corollary. Note that there is a one to one correspondence under $\theta$ between elements of V' and V, and $\theta(1) = 1$. Associated with $\mathcal{L}$ is the congruence $R = \{C_1, \ldots, C_r\}$ on $J_n$ (where V has n-1 elements), and the map $\eta: J_n \to M$ determined by $\eta(w_i) = v_i$, $1 \leq i \leq n-1$ and $\eta(w_n) = 1$. All we need to do is define $\eta': J_n \to M'$ as the (unique) homomorphism such that $\eta'(w_i) = \theta^{-1}(v_i) \cap V'$, for $1 \leq i \leq n-1$, which is precisely one element since $\theta$ is 1-1 on V'; and $\eta'(w_n) = 1$. Then $\mathcal{L}' = \eta'(\bigcup_{1 \leq i \leq s} C_i)$ is a g-set in M', which is free, and $\theta(\mathcal{L}') = \mathcal{L}$, by the construction.

Examples. Let us now look at some examples of the generation of grammatical sets.

Example 3.11: Let M be the linear morphology generated by $A = \{(alb), (ab)\}$. Let $G = (V, W_2, P, \sigma)$ be a grammar generating L(G) in $J_2$, where $\eta(w_1) = (alb)$, $\eta(w_2) = (ab)$, $V = \{\sigma\}$, and P contains the productions

$$\sigma \to Cw_1\sigma$$
$$\sigma \to w_2$$

Then L(G) consists of strings of the form

$$Cw_1Cw_1 \ldots Cw_1w_2, \text{ for } n>0,$$

and (L(G)) is the collection of elements in M

$$\underbrace{(a1b)\cdot(a1b)\cdot\ldots\cdot(a1b)}_{n}\cdot(ab) = \underbrace{aa\ldots a}_{n+1}\,\underbrace{bb\ldots b}_{n+1}, \text{ for } n\geq 0.$$

Hence we have generated the context-free language $\{a^nb^n \mid n\geq 1\}$, as a g-set in (M,A).

From now on, when no confusion will arise, we will substitute for the symbols $w_i \in W_n$ in the productions of a grammar G, the expressions $\eta(w_i)$ of M.

Example 3.12: $= \{a^nb^nc^n \mid n>0\}$. This language is well-known to be context-sensitive, but not context-free. Let M be the linear morphology generated by A = {(a1b2c3), (a1), (b2), (c3), (a), (b), (c)}. Let G = $(V,W_7,P,\sigma)$ be the grammar on $J_7$, where V = {σ, α} and P contains

(1) $\sigma \to C(a1b2c3)\alpha$
(2) $\sigma \to CKK(a1)(b2)(c3)\alpha$
(3) $\sigma \to KK(a)(b)(c)$.

Then L(G) consists of strings

$$\underbrace{C(a1b2c3)CKK(a1)(b2)(c3)CKK(a1)(b2)(c3)\ldots CKK(a1)(b2)(c3)}_{n}KK(a)(b)(c)$$

for $n\geq 0$, and $= \eta(L(G)) = \{a^nb^nc^n \mid n>0\}$, is a g-set in (M,A).

Example 3.13: $= \{aba^2b, aba^2ba^3ba^4b, \ldots\}$, also known not to be context-free. Let M be the linear morphology generated by A = {($aba^2b$), (2a1baa1b), (aa1), (aa), (abaab)}. Define G = $(V,W_5,P,\sigma)$ on $J_5$ by: V = {σ,γ}, P contains

(1) $\sigma \to aba^2b$
(2) $\sigma \to C(2a1baa1b)\gamma$
(3) $\sigma \to CK(aa1)(2a1baa1b)\gamma$

(4) $\gamma \to K(aa)(abaab)$.

Then $\eta(L(G)) = \mathcal{L}$, a g-set in (M,A).

Example 3.14: Let M be any morphology, with blanks denoted by $\underline{1},\underline{2},\ldots$. Let $A = \{\underline{1}\}$. Let $G = (V,W_1,P,\sigma)$ be the grammar on $J_n$ defined by: $V = \{\sigma\}$, P contains:

(1) $\sigma \to K\underline{1}S\sigma$

(2) $\sigma \to \underline{1}$.

Then L(G) contains all strings of the form $K\underline{1}SK\underline{1}SK\underline{1}S\ldots K\underline{1}S\underline{1}$, and $\eta(L(G)) = \mathcal{L} = \{\underline{1},\underline{1}*\underline{2},\underline{1}*\underline{2}*\underline{3},\ldots\}$.

Example 3.15: $\mathcal{L} = \{a^n | n \geq 1\}$. Let M be the linear morphology generated by $A = \{(a\underline{1}),(a)\}$. Let $G = (V,W_2,P,\sigma)$, where $V = \{\sigma,\alpha\}$ and P contains:

(1) $\sigma \to C(a\underline{1})\sigma$

(2) $\sigma \to (a)$.

Example 3.16: $\mathcal{L} = \{(\underline{1}), (\underline{11}), (\underline{111}),\ldots\}$. Let M be generated by $A = \{(\underline{1}), (\underline{12})\}$. Let $G = (V,W_2,P,\sigma)$, where $V = \{\sigma,\alpha\}$, and P contains:

(1) $\sigma \to C(\underline{12})\alpha$

(2) $\alpha \to CK(\underline{1})(\underline{12})\alpha$

(3) $\alpha \to (\underline{1})$

(4) $\sigma \to (\underline{1})$

Example 3.17: $\mathcal{L} = \{a^q | q \text{ not prime}\}$. Let $\mathcal{L}_1 = \{a^n | n>1\}$, which is generated as in Example 3.15, except that production (3) is eliminated. Let $\mathcal{L}_2 = \{(\underline{11}),(\underline{111}),\ldots\}$, which is generated as in Example 3.16 except that production (4) is eliminated. Then $C\mathcal{L}_1\mathcal{L}_2 = \{a^{mn} | m, n>1\} = \mathcal{L}$, and the proof of Theorem 3.7 provides a way of getting the necessary recognizable set.

Example 3.18: $\mathcal{L} = \{(\underline{1}), (\underline{12}), (\underline{123}),\ldots\}$. Let M be the linear morphology generated by $A = \{(\underline{1}), (\underline{12})\}$. Let $G = (V,W_2,P,\sigma)$, where $V = \{\sigma,\alpha\}$ and P contains:

(1) $\sigma \to C(\underline{12})$

(2) $\alpha \to K(\underline{1})C(\underline{12})S$

(3) $\alpha \to K(\underline{1})(\underline{2})$

(4) $\sigma \rightarrow (\underline{1})$.

Regularity. We would like to define a collection of particularly well-behaved g-sets in a morphology M, which we will call regular. In the case of phrase structure languages, the good behavior of regular sets is a consequence of the fact that they represent the union of congruence classes of a finite congruence on the free monoid (under juxtaposition) generated by the set of terminals. We will use a closely related idea. Suppose M is finitely generated, with vocabulary $V = \{v_1,\ldots,v_{n-1}\}$ and we consider $J_n$, with associated homomorphism defined by, for $w_i \in W_n$,

$$\eta(w_i) = v_i,\ 1 \leq i \leq n-1$$

$$\eta(w_n) = 1.$$

Then $\eta: J_n \rightarrow M$ is clearly onto. For a g-set S in $(M, V \cup \{1\})$, S is the union of congruence classes of a finite congruence on M if and only if $\eta^{-1}(S)$ is a recognizable set in $J_n$. It will be fruitful to choose certain recognizable subsets A of $J_n$, and define a notion of A-regularity as follows:

Let M be a morphology with vocabulary $V = \{v_1,\ldots,v_{n-1}\}$, and $J_n$, $\eta$ as above. Let A be a recognizable set in $J_n$. Then a g-set in $(M, V \cup \{1\})$ is A-regular if

(1) $S \subset \eta(A)$

(2) $\eta^{-1}(S) \cap A$ is a recognizable set in $J_n$. Then we have:

Theorem 3.19: Let L be the collection of A-regular g-sets in $(M, V \cup \{1\})$. Then L is closed under finite intersection.

Proof: Let $S_1$ and $S_2$ be such g-sets. Then

$$[\eta^{-1}(S_1) \cap A] \cap [\eta^{-1}(S_2) \cap A] = [\eta^{-1}(S_1) \cap \eta^{-1}(S_2)] \cap A$$

$$= \eta^{-1}(S_1 \cap S_2) \cap A$$

is recognizable, since it is the intersection of recognizable sets. Now $\eta[\eta^{-1}(S_1 \cap S_2) \cap A] \subset S_1 \cap S_2 \cap \eta(A) = S_1 \cap S_2$, since $S_1 \subset (A)$ and $S_2 \subset (A)$. Suppose that $x \in S_1 \cap S_2$. Then $x \in \eta(A)$. Hence there is a y in A such that $\eta(y) = x$. Since $y \in \eta^{-1}(S_1 \cap S_2)$, $y \in (\eta^{-1}(S_1 \cap S_2) \cap A$, and $x \in \eta[\eta^{-1}(S_1 \cap S_2) \cap A]$.

Hence $S_1 \cap S_2 \subset \eta[\eta^{-1}(S_1 \cap S_2) \cap A]$; so $S_1 \cap S_2 = \eta[\eta^{-1}(S_1 \cap S_2) \cap A]$, hence is an A-regular g-set in $(M, V \cup \{1\})$. The theorem follows easily by induction.

Theorem 3.20: Let L be the collection of A-regular g-sets in $(M, V \cup \{1\})$. Then L is closed under finite union.

Proof: Let $S_1$ and $S_2$ be such sets. $S_1 \cup S_2$ is a g-set by Theorem 3.3. $S_1 \cup S_2 \subset \eta(A)$, since $S_1 \subset \eta(A)$ and $S_2 \subset \eta(A)$. To see that $S_1 \cup S_2$ satisfies property (2),

$$\eta^{-1}(S_1 \cup S_2) \cap A = [\eta^{-1}(S_1) \cup \eta^{-1}(S_2)] \cap A$$
$$= [\eta^{-1}(S_1) \cap A] \cup [\eta^{-1}(S_2) \cap A],$$

which is recognizable since recognizable sets are closed under finite union. The theorem then follows easily by induction.

Theorem 3.21: If $S_1$ and $S_2$ are g-sets in $(M, V \cup \{1\})$, $S_1$ is Y-regular, and $S_2 = \eta(A)$, for some recognizable subset A of Y, then $S_1 \cap S_2$ is a g-set in $(M, V \cup \{1\})$.

Proof: Since $S_1$ is Y-regular, $S_1 = \eta[\eta^{-1}(S_1) \cap Y]$, and $\eta^{-1}(S_1) \cap Y$ is recognizable. Hence $\eta^{-1}(S_1) \cap Y \cap A$ is recognizable. Then

$$S_3 = \eta[\eta^{-1}(S_1) \cap Y \cap A] \subset S_1 \cap \eta(Y) \cap \eta(A)$$
$$= S_1 \cap S_2.$$

If $x \in S_1 \cap S_2$, then there is a $y \in A$ such that $x = \eta(y)$. Since $A \subset Y$, $\eta(y) \in Y$. Since $\eta(y) \in S_1$, $y \in \eta^{-1}(S_1)$. Hence $y \in \eta^{-1}(S_1) \cap Y \cap A$, and $x \in \eta[\eta^{-1}(S_1) \cap Y \cap A]$. So $S_1 \cap S_2 \subset S_3$; hence $S_1 \cap S_2 = S_3$, and is a g-set since $S_3$ is.

Theorem 3.22: If S is a Y-regular g-set in (M,B) for any recognizable set Y in $J_n$, then $\eta(Y) \setminus S$ is a Y-regular g-set in (M,B).

Proof: Since recognizable sets are closed under intersection and complementation, $X = [J_n \setminus (\eta^{-1}(S) \cap Y)] \cap Y$ is recognizable. We claim that $\eta(X) = \eta(Y) \setminus S$. If y is in $\eta(X)$, there is a t in X such that $\eta(t) = y$. Since t is in Y, $y = \eta(t)$ is in $\eta(Y)$. Suppose y is in S; then t is in $\eta^{-1}(S) \cap Y$ and hence

not in X, a contradiction. Hence y is not in S, so y is in $\eta(Y)\setminus S$, and we conclude that $\eta(X) \subset \eta(Y)\setminus S$. On the other hand, if y is in $\eta(Y)\setminus S$, then $y = \eta(t)$ for some t in Y, if t is in $\eta^{-1}(S) \cap Y$, then $\eta(t) = y$ is in S, a contradiction; hence t is in $J_n\setminus(\eta^{-1}(S) \cap Y)$, so t is in X and y is in $\eta(X)$. Hence $\eta(Y)\setminus S \subset \eta(X)$, and $\eta(Y)\setminus S = \eta(X)$, as claimed.

To show that $\eta(Y)\setminus S$ is Y-regular, first it is obvious that $\eta(Y)\setminus S \subset \eta(Y)$. Now

$$\eta^{-1}[(\eta(Y)\setminus S)] \cap Y = \eta^{-1}[\eta(X)] \cap Y.$$

Suppose t is in $\eta^{-1}(\eta(X)) \cap Y$, and t is not in X. Then t is not in $J_n\setminus(\eta^{-1}(S) \cap Y)$. Hence t is in $\eta^{-1}(S)\setminus Y$; but then $\eta(t)$ is in S, which is not possible since

$$\eta\eta^{-1}[\eta(X) \cap Y] = \eta(X) = \eta(Y)\setminus S,$$

a contradiction. Hence t must be in X, and $\eta^{-1}(\eta(X)) \cap Y \subset X$. Since $X \subset Y$, and $X \subset \eta^{-1}\eta(X)$, we have $X \subset \eta^{-1}[\eta(X)] \cap Y$. So $X = \eta^{-1}[\eta(Y)\setminus S] \cap Y$, and since X is recognizable, $\eta(Y)\setminus S$ is Y-regular.

Factorizations. Let M, V, $\mathcal{J}_n$, $\eta$ be as in the previous section. We define recursively a recognizable set $F_V$ (or simply F, where V is understood) in $\mathcal{J}_n$, called the V-factorizations (or factorizations) of M in $\mathcal{J}_n$. F will be the least subset of $\mathcal{J}_n$ such that:

(1) $W_n \subset F$.

(2) $\{\underbrace{SS\ldots S}_{r} w_n \mid r>0\} \subset F$.

(3) For $1 \leq i \leq n-1$, if $\deg(\eta(w_i)) = r>1$ and $t_1,\ldots,t_r$ are in F, then $Cw_i \underbrace{KK\ldots K}_{r-1} t_1 t_2 \ldots t_r$ is in F.

(4) For $1 \leq i \leq n-1$, if $\deg(\eta(w_i)) = 1$, and t is in F, then $Cw_i t$ is in F.

We remark that $\eta(F)$ is precisely the collection of phrases in M, since there is a natural correspondence between the V-factorizations in $\mathcal{J}_n$ and those defined earlier; namely,

for a phrase x in M, with V-factorization $G = (v_i, G_1, \ldots, G_k)$, there is a term $t = Cv_iKK\ldots Kt_1t_2\ldots t_r$ in F such that $\eta(t) = x$ and $\eta(t_j) = G_j$, $1 \leq j \leq r$.

Theorem 3.23: The collection of V-factorizations of M in $\mathcal{J}_n$ is a recognizable set.

Proof: Let R contain the following sets:

(1) $B_i = \{w_i\}$, for $1 \leq i \leq n-1$.

(2) $B_n = \{\underbrace{SS\ldots S}_{k} w_n \mid k \geq 0\}$

(3) For $1 \leq i \leq n-1$,

$$C_i = \begin{cases} \{Cw_i\underbrace{KK\ldots K}_{r-1}t_1t_2\ldots t_r \mid t \in F\}, \text{ if degree } \eta(w_i) = r > 1 \\ \{Cw_it \mid t \in F\}, \text{ if degree } (\eta(w_i)) = 1. \end{cases}$$

(4) Let $s = \max \{\deg (\eta(w_i)) \mid 1 \leq i \leq n-1\}$. Then for $2 \leq j \leq s$,

$$D_j = \{\underbrace{KK\ldots K}_{j-1}t_1t_2\ldots t_j \mid t_1, t_2, \ldots, t_j \in F\},$$

(5) $E = \mathcal{J}_n \setminus [(\bigcup_{1 \leq i \leq n} B_i) \cup (\bigcup_{1 \leq i \leq n-1} C_i) \cup (\bigcup_{2 \leq j \leq s} D_j)]$.

It is easy to see that R is a partition of $\mathcal{J}_n$, and that $F = [\bigcup_{1 \leq i \leq n} B_i] \cup [\bigcup_{1 \leq i \leq n-1} C_i]$. We need only ascertain that R is a congruence. The tables below show the results of application of the operations C, K, S to set in R, and are trivial to verify.

| X | SX |
|---|---|
| $B_i, 1 \leq i \leq n-1$ | E |
| $B_n$ | $B_n$ |
| $C_i, 1 \leq i \leq n-1$ | E |
| $D_j$, $2 \leq j \leq s$ | E |
| E | E |

| C | $B_j$, $1 \leq j \leq n-1$ | $B_n$ |
|---|---|---|
| $B_i$, $1 \leq i \leq n-1$ | $C_i$, if deg $\eta(w_i) = 1$<br>E, otherwise | $C_i$, if deg $\eta(w_i) = 1$<br>E, otherwise |
| $B_n$ | E | E |
| $C_i$, $1 \leq i \leq n-1$ | E | E |
| $D_j$, $2 \leq j \leq s$ | E | E |
| E | E | E |

| C | $C_j$, $1 \leq j \leq n-1$ | $D_k$, $2 \leq k \leq s$ | E |
|---|---|---|---|
| $B_i$, $1 \leq i \leq n-1$ | $C_i$, if deg $\eta(w_i) = 1$<br>E, otherwise | $C_i$, if deg $\eta(w_i) = k$<br>E, otherwise | E |
| $B_n$ | E | E | E |
| $C_i$, $1 \leq i \leq n-1$ | E | E | E |
| $D_j$, $2 \leq j \leq s$ | E | E | E |
| E | E | E | E |

| K | $B_j$, $1 \leq j \leq n-1$ | $B_n$ |
|---|---|---|
| $B_i$, $1 \leq i \leq n-1$ | $D_2$ | $D_2$ |
| $B_n$ | $D_2$ | $D_2$ |
| $C_i$, $1 \leq i \leq n-1$ | $D_2$ | $D_2$ |
| $D_j$, $2 \leq j \leq s$ | $D_{k+1}$, if $k<s$<br>E, otherwise | $D_{k+1}$, if $k<s$<br>E, otherwise |
| E | E | E |

| K | $C_j$, $1 \leq j \leq n-1$ | $D_k$, $2 \leq k \leq s$ | E |
|---|---|---|---|
| $B_i$, $1 \leq i \leq n-1$ | $D_2$ | E | E |
| $B_n$ | $D_2$ | E | E |
| $C_i$, $1 \leq i \leq n-1$ | $D_2$ | E | E |
| $D_j$, $2 \leq j \leq s$ | $D_{k+1}$, if $k<s$<br>E, otherwise | E | E |
| E | E | E | E |

We will examine the F-regular g-sets in more detail in Chapter 4, when we consider g-sets in linear morphologies.

Concatenative depth. For terms in $\mathcal{J}_n$, we define concatenative depth (K-depth) recursively as follows:

(i) K-depth $(w_i) = 1$ for $w_i \in W_n$.

(ii) For $t_1, t_2 \in J_n$, K-depth $(Ct_1,t_2) = \max$ {K-depth $(t_1)$, K-depth $(t_2)$}

(iii) For $t_1, t_2 \in J_n$, K-depth $(Kt_1,t_2) = \max$ {K-depth $(t_1)$,

K-depth $(t_2)$, dim $\eta(Kt_1t_2)\}$.

(iv) For $t \in J_n$, K-depth $(St)$ = K-depth $(t)$.

A subset B of $\mathcal{J}_n$ has <u>finite K-depth r</u> if r is the least integer such that each element of B has K-depth no greater than r. If no such r exists, then B has <u>infinite K-depth</u>.

<u>Theorem 3.24</u>: For any integer $n \geq 1$, the collection of terms t in $\mathcal{J}_n$ such that K-depth $(t)$ = n is a recognizable set in $\mathcal{J}_n$.

<u>Proof</u>: For $1 \leq i, j \leq n$, let $C(i,j) = \{t \text{ in } \mathcal{J}_n \mid$ K-depth $(t) = i$, dim $\eta(t) = j\}$. Let $D = \{t \text{ in } \mathcal{J}_n \mid$ K-depth $(t) > n\}$. Then $R = \{C(i,j) \mid 1 \leq i, j \leq n\} \cup \{D\}$ is a partition of $\mathcal{J}_n$. To show that R is a congruence on $\mathcal{J}_n$:

(1) If $x \in C(i,j)$, $y \in C(k,p)$, then $Cxy \in C(m,j)$ where $m = \max\{i,k\}$.

(2) If $x \in C(i,j)$, $Sx \in C(i,j)$.

(3) If $x \in C(i,j)$, $y \in C(k,p)$, then

(1) if $j+p \leq n$, $Kxy \in C(m,j+p)$, where $m = \max\{i,k,j+p\}$.

(2) if $j+p > n$, $Kxy \in D$.

<u>Corollary 3.25</u>: Let $D = \{n_1, \ldots, n_t\}$ be a finite collection of integers. Then $J = \{t \in \mathcal{J}_n \mid$ K-depth $(t) \in D\}$ is a recognizable set in $\mathcal{J}_n$.

<u>Proof</u>: Recognizable sets are closed under union.

We will need the notion of K-depth, as well as that of the dimension and degree of a set in Chapter 4.

A subset B of a morphology M will be called <u>r-dimensional</u> if r is the least integer such that each element in B has dimension at most r. A set C in $\mathcal{J}_n$ is r-dimensional if r is the least integer such that, for each element x in C, the dimension of $\eta(x)$ is no greater than r. (Note that the definition is unambiguous, since, for all the homomorphisms $\eta: \mathcal{J}_n \to M$ which we use, the dimension of $\eta(x)$ is the same.) In each case, if no such r exists, the set is <u>infinite-dimensional</u>.

Analogously, a subset B of a morphology M (respectively, the algebra $\mathcal{J}_n$) has degree r if r is the least integer such that the degree of x (respectively $\eta(x)$) is no greater than r for all x in B. Otherwise, B has infinite degree.

Ambiguity. We want to consider two kinds of ambiguity which can arise in the generation of a grammatical set; the first, which is analogous to the ambiguity arising in phrase structure languages, and is related to the properties of the recognizable sets, we will call structural ambiguity; the second, which has to do with the properties of the particular morphology we are dealing with, we will call morphological ambiguity.

Let M be a morphology with vocabulary $V = \{v_1,\ldots,v_{n-1}\}$, and let M' be its associated free morphology with vocabulary $V' = \{v'_1,\ldots,v'_{n-1}\}$, and onto homomorphism $\theta$: $M' \to M$ such that $\theta(v'_i) = v_i$ for $1 \leq i \leq n-1$. We will need the following fact.

Theorem 3.26: If $\eta: \mathcal{J}_n \to M$ is a homomorphism, then there are homomorphisms $\alpha: \mathcal{J}_n \to M'$ and $\theta: M' \to M$ such that $\theta\alpha = \eta$.

Proof: Let $\theta$ be the homomorphism of Corollary 2.17. Let $\alpha$ be the homomorphism determined by: for w in $W_n$, let $\alpha(w)$ be that element of the vocabulary V' of M' such that $\theta\alpha(w) = \eta(w)$ in the vocabulary V of M. Then it is easy to see that $\theta$ and $\alpha$ are the required maps.

We will consider only g-sets over $(M, V \cup \{1\})$, where $V = \{v_1,\ldots,v_{n-1}\}$ is a fixed ordering of V; consider homomorphic images of recognizable sets in $\mathcal{J}_n$, where $\eta: \mathcal{J}_n \to M$ is determined by $\eta(w_i) = v_i$, $1 \leq i \leq n-1$ and $\eta(w_n) = 1$.

Now suppose A is a recognizable set in $\mathcal{J}_n$. We will call A structurally unambiguous under $\eta$ if the map $\alpha: \mathcal{J}_n \to M'$ is one to one on A. Otherwise A is structurally ambiguous under $\eta$. A g-set $\mathcal{S}$ in $(M, V \cup \{1\})$ is structurally unambiguous if there exist a structurally unambiguous recognizable set A in $\mathcal{J}_n$ such that $\mathcal{S} = \eta(A)$. Otherwise, $\mathcal{S}$ is structurally ambiguous.

A g-set $\mathcal{A}$ in M will be called morphologically unambiguous if $\mathcal{A} = \theta(\mathcal{A}')$, for some g-set $\mathcal{A}'$ in $(M', V' \cup \{1\})$. Otherwise, $\mathcal{A}$ is morphologically ambiguous.

Theorem 3.27: If $\mathcal{A}$ is a g-set in a free morphology M, then $\mathcal{A}$ is morphologically unambiguous.

Proof: If M is free, then by Theorem 2.16, the map $\theta: M' \to M$ is an isomorphism. By Theorem 3.10, $\mathcal{A} = \theta(\mathcal{A}')$ for some g-set $\mathcal{A}'$ in M, and $\theta$ is one to one on $\mathcal{A}'$.

Theorem 3.28: If $\mathcal{A}$ is an F-regular g-set in $(M, V \cup \{1\})$, where M is any morphology with vocabulary V and F is the collection of V-factorizations of M in $\mathcal{J}_n$, then $\mathcal{A}$ is structurally unambiguous.

Proof: Since $\mathcal{A}$ is F-regular, $\mathcal{A} = \eta[\eta^{-1}(\mathcal{A}) \cap F] = \theta\alpha[\eta^{-1}(\mathcal{A}) \cap F]$, where $\eta^{-1}(\mathcal{A}) \cap F$ is recognizable. We note that $\alpha$ is one to one on F; for M' is free with reduced vocabulary V'; hence each phrase in M' has precisely one V'-factorization, and the V'-factorizations are in one to one correspondence with the terms in F.

Corollary 3.29: If $\mathcal{A} = \eta(A)$, where A is recognizable, and $A \subset F$, then $\mathcal{A}$ is structurally unambiguous.

In the theory of context-free languages, a context-free grammar is unambiguous if each element of the language it generated has precisely one leftmost derivation; otherwise it is ambiguous. A context-free language is unambiguous if there is an unambiguous grammar generating it; otherwise it is inherently ambiguous.

This type of ambiguity is analogous to the structural ambiguity defined for half-ring grammars and grammatical sets. As a matter fact, we can simulate the context-free generating process with a morphology whose semigroup under composition is the free semigroup generated by a collection of terminal symbols (whose composition is concatenation); then the context-free languages are the g-sets generated by using only composition rules. Then the usual amoiguity corresponds exactly to our concept of structural ambiguity.

In Chapter 4, we will show that all context-free languages can be generated as structurally unambiguous g-sets in linear morphologies. The example which follows is a context-free language known to be inherently ambiguous. It can be generated as a g-set which is both structurally and morphologically unambiguous.

Example 3.30: $= \{a^i bwba^i ba^j \mid i,j \geq 1\}$ $\{a^i bwba^j ba^j \mid i,j \geq 1\}$.

Let M be the linear morphology generated by V = $\{(\underline{1}bwb\underline{1}b\underline{2}), (\underline{1}bwb\underline{2}b\underline{2}), (a\underline{1}), (a)\}$. Let $\eta: \mathcal{J}_4 \to M$ be determined by:

$$\eta(w_1) = (\underline{1}bwb\underline{1}b\underline{2})$$
$$\eta(w_2) = (\underline{1}bwb\underline{2}b\underline{2})$$
$$\eta(w_3) = (a\underline{1})$$
$$\eta(w_4) = (a).$$

Let M' be the free morphology associated with M, with vocabulary V' = $\{c\underline{12}, d\underline{12}, e\underline{1}, f\}$, where

$$\theta(c\underline{12}) = (\underline{1}bwb\underline{1}b\underline{2})$$
$$\theta(d\underline{12}) = (\underline{1}bwb\underline{2}b\underline{2})$$
$$\theta(e\underline{1}) = (a\underline{1})$$
$$\theta(f) = (a).$$

Let $G = (U, W_4, P, \sigma)$ be the grammar on $\mathcal{J}_4$ such that U = $\{\sigma, \alpha\}$ and P contains

(1) $\sigma \to Cw_1K$
(2) $\sigma \to Cw_2K$
(3) $\alpha \to Cw_3\alpha$
(4) $\alpha \to w_4$

Then $L(G) \subset F$, hence is structurally unambiguous by Corollary 3.29, and $\eta(L(G)) = \mathcal{J}$.

However, morphological ambiguity remains; for example, consider the two elements $Cw_1Kw_4w_4$ and $Cw_2Kw_4w_4$ of L(G). We have

$$\alpha(Cw_1Kw_4w_4) = (c\underline{12})\cdot(f*f)$$
$$= cff$$
$$\text{and } \alpha(Cw_2Kw_4w_4) = (d\underline{12})\cdot(f*f)$$
$$= dff;$$

$$\begin{aligned}
\text{but } \eta(Cw_1Kw_4w_4) &= \theta\alpha(Cw_1Kw_4w_4)\\
&= \theta(cff)\\
&= (\underline{1}bwb\underline{1}b\underline{2})\cdot(a*a)\\
&= abwbaba\\
\text{and } \eta(Cw_2Kw_4w_4) &= \theta\alpha(Cw_2Kw_4w_4)\\
&= \theta(dff)\\
&= (\underline{1}bwb\underline{2}b\underline{2})\cdot(a*a)\\
&= abwbaba,
\end{aligned}$$

so $\theta$ is not one to one on $\alpha(L(G))$.

Now we let $G' = (U', W_4, P', \sigma)$ be the somewhat more complex grammar on $W_4$ defined by: $U' = \{\sigma, \alpha, \tau\}$, where $P'$ contains:

(1) $\sigma \to Cw_1\alpha$
(2) $\alpha \to Cw_3\alpha$
(3) $\alpha \to w_4$
(4) $\sigma \to Cw_1CKw_4Cw_3\alpha$
(5) $\sigma \to Cw_1CKCw_3\alpha w_4$
(6) $\sigma \to Cw_2CKw_4Cw_3\alpha$
(7) $\sigma \to Cw_2CKCw_3\alpha w_4$
(8) $\sigma \to Cw_1CKw_3Cw_3\tau\alpha$
(9) $\sigma \to Cw_1CKCw_3\tau w_3\alpha$
(10) $\sigma \to Cw_2CKw_3Cw_3\tau\alpha$
(11) $\sigma \to Cw_2CKCw_3\tau w_3\alpha$
(12) $\tau \to Cw_3\tau$
(13) $\tau \to w_3$

It is tedious but straightforward to show that $\mathcal{L} = \eta(L(G))$, $G$ is structurally unambiguous, and $\theta$ is one to one on $\eta(L(G))$. Hence $\mathcal{L}$ is both structurally and morphologically unambiguous as a g-set in $(M,V)$.

# CHAPTER IV

## LINGUISTIC SETS

For linguistic purposes, it turns out that grammatical sets are not precisely the objects we want to deal with. In particular, Example 3.16 and Example 3.18 show that g-sets may contain elements of positive degree. We may think of these elements as well-formed, but only partially formed sentences, since they contain unfilled blanks. For example,

The cowpoke kicked his pony in the _____.

requires the addition of, say, "morning," "rain," "corral," or "flank" to become a complete sentence, though its structure so far is acceptable, as compared with

Cowpoke _____ pony the the his in kicked.

which presumably we would not generate as an element of a g-set at all. We want to restrict a linguistic set, then, to those elements of a g-set which are "completely filled in," that is, those of degree zero.

In our linguistic application, a sentence is a one-dimensional element. A concatenation of two or more one-dimensional elements may be thought of as a string of sentences, or a paragraph.

In a morphology M, let E be the collection of elements of dimension 1. We have this fact:

Lemma 4.1: If $\mathcal{J}$ is a g-set in (M,A), so is $\mathcal{J} \cap E$.

Proof: Let $R = \{\eta^{-1}(E), \eta^{-1}(M \setminus E)\}$. Then R is a finite congruence on $\mathcal{J}_\eta$, as shown by the tables below, which are easily verified.

| C | $\eta^{-1}(E)$ | $\eta^{-1}(M\setminus E)$ |
|---|---|---|
| $\eta^{-1}(E)$ | $\eta^{-1}(E)$ | $\eta^{-1}(E)$ |
| $\eta^{-1}(M\setminus E)$ | $\eta^{-1}(M\setminus E)$ | $\eta^{-1}(M\setminus E)$ |

| K | $\eta^{-1}(E)$ | $\eta^{-1}(M\setminus E)$ |
|---|---|---|
| $\eta^{-1}(E)$ | $\eta^{-1}(M\setminus E)$ | $\eta^{-1}(M\setminus E)$ |
| $\eta^{-1}(M\setminus E)$ | $\eta^{-1}(M\setminus E)$ | $\eta^{-1}(M\setminus E)$ |

| S | |
|---|---|
| $\eta^{-1}(E)$ | $\eta^{-1}(E)$ |
| $\eta^{-1}(M\setminus E)$ | $\eta^{-1}(M\setminus E)$ |

Suppose $\mathcal{S} = \eta(\bigcup_{j=1}^{n} C_j)$, where $R' = \{C_1,\ldots,C_s\}$ is a finite congruence on $\mathcal{S}_n$. Then

$$R'' = \bigcup_{i=1}^{s} \{C_i \cap \eta^{-1}(E),\ C_i \cap \eta^{-1}(M\setminus E)\}$$

is a finite congruence. Define the g-set $\mathcal{S}'$ by:

$$\mathcal{S}' = \eta\Big(\bigcup_{j=1}^{k} (C_j \cap \eta^{-1}(E))\Big).$$

Then

$$\begin{aligned}\mathcal{S}' &= \bigcup_{j=1}^{k} (\eta(C_j \cap \eta^{-1}(E)))\\ &\subset \bigcup_{j=1}^{k} (\eta(C_j) \cap E)\\ &= \Big(\bigcup_{j=1}^{k} \eta(C_j)\Big) \cap E\\ &= \mathcal{S} \cap E.\end{aligned}$$

If x is in $\mathcal{S} \cap E$, then there is a j, and there is a y in $C_j$, such that $\eta(y) = x$ and y is in $\eta^{-1}(E)$. Hence y is in $C_j \cap \eta^{-1}(E)$, and $\eta(y)$ is in $(C_j \cap \eta^{-1}(E))$; hence y is in $\mathcal{S}'$. So $\mathcal{S}' \supset \mathcal{S} \cap E$, and $\mathcal{S}' = \mathcal{S} \cap E$.

It is not true that if $\mathcal{S}$ is a g-set of dimension k greater than one, then $\mathcal{S} = KK\ldots K\mathcal{S}_1\mathcal{S}_2\ldots\mathcal{S}_k$ for some g-sets $\mathcal{S}_i$,

$1 \leq i \leq k$, as shown by the following example.

Let M be a free morphology with ordered vocabulary $V = \{v_1, \ldots, v_{n-1}\}$ and let F be the collection of V-factorizations in $\mathcal{J}_n$, where $\eta(w_i) = v_i$, $1 \leq i \leq n-1$, and $\eta(w_n) = 1$. Then F is generated by the grammar $G = (\{\sigma, \alpha\}, W_n, P, \sigma)$, with productions

(1) $\sigma \rightarrow w_j$, $1 \leq j \leq n$,

(2) $\sigma \rightarrow \alpha$

(3) $\alpha \rightarrow S\alpha$

(4) $\alpha \rightarrow 1$

(5) $\sigma \rightarrow Cw_j \underbrace{KK\ldots K}_{r-1} \underbrace{\sigma\sigma\ldots\sigma}_{r}$, for each $w_j$,

where $r = \deg \eta(w_j)$.

Now L(G) is the collection of factorizations in $\mathcal{J}_n$, and $\eta(L(G))$ is the collection of phrases in M. Let $G' = (\{\sigma, \alpha, \sigma'\}, W_n, P', \sigma')$, where $P' = P \cup \{\sigma' \rightarrow CK\underline{1}\underline{1}\sigma\}$. Let $\mathcal{S}' = \eta(L(G'))$. Then $\mathcal{S}' = \{x * x \mid x \in \eta(L(G))\}$. $\mathcal{S}'$ has dimension two.

Suppose $\mathcal{S}' = K\mathcal{S}_1\mathcal{S}_2$ for some g-sets $\mathcal{S}_1, \mathcal{S}_2$. Let $v_1$ and $v_2$ be the distinct elements of V such that $\eta(w_1) = v_1$ and $\eta(w_2) = v_2$. Since $v_1$ is in $\mathcal{S}$, $v_1 * v_1$ is in $\mathcal{S}'$; hence $v_1$ must be in $\mathcal{S}_1$. Similarly, $v_2$ must be in $\mathcal{S}_2$. But since $\mathcal{S}' = K\mathcal{S}_1\mathcal{S}_2$, then $v_1 * v_2$ is in $\mathcal{S}'$, a contradiction, since $v_1 \neq v_2$.

This illustration shows that the structuring possibilities of g-sets reach beyond the sentence level. However, we consider only the one-dimensional case in this paper, which is that case corresponding to the construction of isolated sentences. Lemma 4.1 shows that we may either consider sets $\mathcal{S} \cap E$ where $\mathcal{S}$ is an arbitrary g-set, or simply g-sets $\mathcal{S}'$ of dimension one.

With this motivation, we define a <u>linguistic set</u>

(1-set) $\Gamma$ in the (M,A) as a set of the form $\mathcal{S} \cap D$, where $\mathcal{S}$ is a g-set in (M,A) and D is the collection of formulas in M.

Properties of linguistic sets. First we find some simple closure properties.

Theorem 4.2: If $\Gamma_1$ and $\Gamma_2$ are 1-sets in (M,A), then so is $\Gamma_1 \cup \Gamma_2$.

Proof: Suppose $\Gamma_1 = \mathcal{S}_1 \cap D$, $\Gamma_2 = \mathcal{S}_2 \cap D$, for g-sets $\mathcal{S}_1$, $\mathcal{S}_2$. Then $\Gamma_1 \cup \Gamma_2 = (\mathcal{S}_1 \cup \mathcal{S}_2) \cap D$, and by Theorem 3.3, $\mathcal{S}_1 \cup \mathcal{S}_2$ is a g-set, hence the result follows.

Theorem 4.3: If $\Gamma$ is a linguistic set in (M,A) and $h: M \to M'$ is a degree preserving homomorphism, then $h(\Gamma)$ is a linguistic set in (M', h(A)).

Proof: Let D be the set of formulas of M, D' those in M'. For some g-set $\mathcal{S}$, $\Gamma = \mathcal{S} \cap D$. By Theorem 3.7, $h(\mathcal{S})$ is a g-set in (M',h(A)). Now

$$\begin{aligned} h(\Gamma) &= h(\mathcal{S} \cap D) \\ &\subset h(\mathcal{S}) \cap h(D) \\ &\subset h(\mathcal{S}) \cap D', \end{aligned}$$

since $h(D) \subset D'$ (homomorphisms never increase degree).

Suppose x is in $h(\mathcal{S}) \cap D'$. Then there is a y in such that $h(y) = x$, $\dim(h(y)) = 1$, and $\deg(h(y)) = 0$. Since all homomorphisms preserve dimension, $\dim(y) = 1$; since h preserves degree, $\deg(y) = 0$. Hence y is in D, so y is in $\mathcal{S} \cap D$ and x is in $h(\Gamma)$. So $h(\mathcal{S}) \cap D' \subset h(\Gamma)$. This concludes the proof that $h(\Gamma) = h(\mathcal{S}) \cap D'$, which is a linguistic set in M'.

We notice in passing that 1-sets are not closed under concatenation, and are trivially closed under composition and shift, since $C\Gamma_1\Gamma_2 = \Gamma_1$ and $S\Gamma_1 = \Gamma_1$ for 1-sets $\Gamma_1$ and $\Gamma_2$.

Homogeneous variables and restricted linguistic sets. Now we arrive at the final condition which will yield the class

of sets we had in mind for linguistic applications. In the generation of sentences from rewriting rules, the variables in the grammars will represent grammatical categories, just as they do in the linguistic applications of context-free languages.

In Chapter 1, we suggested that transitive verbs be considered as two-blank predicates, as

( 1 carried 2 ),

to be composed with a 2-tuple (x,y), where x is a subject and y is an object. Hence we would like the variable $\nu$, which yields the grammatical category "transitive verb," to yield only one-dimensional elements of degree two. We will also want a variable $\alpha$ which yields precisely 2-tuples of the form (subject,object); these will all be two-dimensional elements $x*y$ such that deg (x) = 0 and deg (y) = 0, that is, x and y are "completely filled in."

In similar fashion, other grammatical categories will naturally have some fixed specifications of dimension and degree. Therefore, we will define homogeneous variables, which yield only elements of "fixed specifications." The condition of being generable by homogeneous variables will be the final requirement we make for the linguistic model.

The sets we propose as models for the syntax of language, then, are these: linguistic sets $\mathcal{L} \cap D$, where $\mathcal{L}$ is a grammatical set in (M,A) for a linear morphology M and some finite set of phrases A, $\mathcal{L}$ is generated by a grammar all of whose variables are homogeneous, and D is the collection of formulas in M.

We now make precise the notion of homogeneous variable. Let M be a linear morphology with (ordered) vocabulary $V = \{v_1,\ldots,v_{n-1}\}$ and let $\eta:\mathcal{J}_n \to M$ be the homomorphism which maps $w_i$ to $v_i$ for $1 \leq i \leq n-1$, and $w_n$ to 1. Then let $H = (U,W_n,P,\sigma)$ be a grammar. For a variable $\alpha$ in H, we will call $\alpha$ homogeneous if there is associated with it an r-tuple

of finite sets of integers $(N_1,\ldots,N_r)$, called its specifications, such that whenever $\alpha$ yields $x$ in $L(H)$,

1) $\eta(x)$ has dimension $r$ and

2) for $1\leq i\leq r$, $N_i$ is precisely the collection of blanks of which $i\cdot\eta(x)$ is not free.

As an example, if $\alpha$ is homogeneous, $\alpha$ yields $x$, and $x = a\underline{1}b\underline{3}*b\underline{2}c*b\underline{1}a\underline{4}$, then the specifications of $\alpha$ are $(\{1,3\},\{2\},\{1,4\})$.

Then a g-set $\mathcal{S}$ in $(M, V\cup\{1\})$ will be homogeneous if it is the interpretation under $\eta$ of a recognizable set generated by a grammar all of whose variables are homogeneous. An 1-set $\Gamma$ in $(M, V\cup\{1\})$ will be homogeneous if it is $\mathcal{S}\cap D$, for some homogeneous g-set $\mathcal{S}$, where D is the collection of formulas in M.

A natural restriction on the form of productions in the grammar generating a grammatical set $\mathcal{S}$ will guarantee that $\mathcal{S}$ can be generated by a grammar all of whose variables are homogeneous. The restriction is this: we will not allow generating rules containing the operator symbol S.

Given a pair $(M, V\cup\{1\})$, where V is an ordered vocabulary of M with n-1 elements, let $G = (U,W_n,P,\sigma)$ be a grammar such that P contains no productions in which S appears. Then $\eta(L(G)) = \Gamma$ will be called a restricted grammatical set (rg-set) and $T = \mathcal{S}\cap D$ a restricted linguistic set (rl-set) where D is the collection of formulas of M. [Note that $\eta$ here is the usual homomorphism mapping $w_i$ to $v_i$, $1\leq i\leq n-1$, and mapping $w_n$ to 1.]

We may assume when desired that G is in best form (see discussion in Chapter 3).

Let $\overline{S}$ be the collection of terms in $\mathcal{J}_n$ containing the symbol S. Then the equivalent formulation using finite congruences on $\mathcal{J}_n$ is this: the restricted g-sets $\Gamma$ are precisely those such that $R = \{C_1,\ldots,C_r\}$ is a congruence on $\mathcal{J}_n$, $\bigcup_{j=1}^{k} C_j\cap\overline{S} = \phi$, and $\Gamma = \eta(\bigcup_{j=1}^{k} C_j)$. Also, since

$R' = \{\bar{S}, \mathcal{J}_n \setminus \bar{S}\}$ is a finite congruence on $\mathcal{J}_n$, if we are given any congruence R" on $\mathcal{J}_n$, then the use of the congruence $R' \wedge R"$ will allow us to obtain as a g-set the "restricted part" of any g-set. This procedure is equivalent to removing from the generating grammar G (in best form) all rules containing the symbol S on the right-hand side.

Now we will embark on a sequence of proofs which will show that the restricted linguistic sets are precisely the ones we had in mind. The main result is contained in Theorem 4.10. Lemmas 4.4 and 4.5 are needed in the proof of Theorem 4.6.

Lemma 4.4: Let $G = (U, W_n, P, \sigma)$ be a restricted grammar in best form such that $\eta(L(G))$ is one-dimensional. Let $A_0 = \{\sigma\} \cup \{\alpha \mid \sigma \to \alpha \text{ is in } P\}$. For $i \geq 0$, let $A_{i+1} = A_i \cup \{\beta \in U \mid \alpha \to C\beta\gamma \text{ is in } P \text{ for some } \gamma \text{ in } U, \alpha \text{ in } A_i\}$. Let m be the number of variables in U. Then $\bigcup_{i \geq 0} A_i = A_m$, and for each $\beta$ in $A_m$, for each x such that $\beta$ yields x, $\dim(\eta(x)) = 1$.

Proof: Let $|A_i|$ denote the number of elements in $A_i$. Suppose that for some $i \geq 0$, $A_i = A_{i+1}$. Then for all $k > 1$, $A_i = A_{i+k}$. If $k = 2$, suppose $A_i \neq A_{i+2}$. Then there is a production $\alpha \to C\beta\gamma$ in P such that $\alpha$ is in $A_{i+1}$, and $\beta$ is not in $A_{i+1}$; hence $\alpha$ is not in $A_i$, a contradiction, since $A_i = A_{i+1}$. If the hypothesis holds for all $j < k$, suppose $A_i \neq A_{i+k}$. Then, again, for some $\alpha, \beta, \gamma$, $\alpha \to C\beta\gamma$ is in P, $\alpha$ is in $A_{i+k-1}$, and $\beta$ is not in $A_{i+k-1}$; hence $\alpha$ is not in $A_{i+k-1}$, a contradiction of our assumption; so if for some $i \geq 0$, $A_i = A_{i+1}$, then $A_i = A_j$ for all $j > i$.

Since $A_i \neq A_{i+1}$ if and only if $|A_i| < |A_{i+1}|$, then for some $j \leq m$, $A_j = A_{j+1} = A_m$, which proves the first assertion. The second assertion follows by induction on the length m of a derivation

$$\alpha = x_0 \xrightarrow{\pi_1} x_1 \xrightarrow{\pi_2} \cdots \xrightarrow{\pi_m} x_m = x, \text{ where } \alpha \in A_m.$$

Suppose m = 1. Then $\pi_1$ is $\alpha \to w_j$ and dim $[\eta(w_j)] = 1$.
Suppose the assertion holds for all x such that there is a derivation of x of length $<m$.

Case 1. $\pi_1$ has the form $\sigma \to \beta$; then $\beta$ yields x by a derivation of length less than m, and dim $[\eta(x)] = 1$, by the induction hypothesis.

Case 2. $\pi_1$ has the form $\alpha \to C\beta\gamma$; then $x = Ct_1t_2$ and $\beta$ yields $t_1$, $\gamma$ yields $t_2$, both by subderivations of length less than m. Hence dim $[\eta(t_1)] = 1 = \dim [\eta(t_1)\cdot\eta(t_2)] = \dim (\eta(Ct_1t_2)) = \dim \eta(x)$.

Case 3. $\pi_1$ has the form $\alpha \to K\beta\gamma$. This is not possible for a variable $\alpha$ in $A_m$, since $\alpha$ is one-dimensional; for, suppose it is. Since G is reduced, there is some y in L(G) such that

(*) $\qquad \alpha \to K\beta\gamma \Rightarrow Kt_1t_2 = y;\ \dim (\eta(y)) \geq 2.$

Let j be the least integer such that $\alpha$ is in $A_j$. Then there is a derivation

$$\sigma \to C\delta_1\delta_2 \to CC\delta_3\delta_4\delta_2 \to \cdots \to \underbrace{CC\cdots C}_{j}\delta_{(2j-1)}\delta_{(2j-2)}\cdots\delta_6\delta_4\delta_2,$$

where the $\delta_i$ are in V, and $\delta_{2j-1} = \alpha$. Now apply to $\alpha$ the sequence (*), yielding

(**) $\qquad \sigma \Rightarrow CC\ldots CKt_1t_2\delta_{(2j-2)}\cdots\delta_6\delta_4\delta_2.$

Again, since G is reduced, there are productions in P which can be applied to the variables in (**) to yield a term z in $\mathcal{J}_n$, and dim $\eta(z) \geq 2$. This contradicts the fact that L(G) is one-dimensional. Hence no productions of the form $\alpha \to K\beta\gamma$ appear in P for $\alpha$ in A. This completes the proof of the second assertion.

Lemma 4.5: If $G = (V, W_n, P, \sigma)$ is a restricted grammar generating $\mathcal{L}$ in (M,A), then for all $\alpha \in V$, and for all t in $\mathcal{J}_n$ such that $\alpha$ yields t, deg $(\eta(t)) \leq r$, where $r = \max \{\deg a \mid a \in A\}$.

Proof: By induction on the length m of a derivation.

Assume G is in best form. Let t be in $\mathcal{J}_n$, with derivation

$$\alpha \xrightarrow[\pi_1]{} x_1 \xrightarrow[\pi_2]{} x_2 \rightarrow \ldots \xrightarrow[\pi_m]{} x_m = t.$$

Suppose m = 1. Then $\pi_1$ is $\alpha \rightarrow w_j$, and $\eta(w_j) = a$ for some a in A, hence deg $\eta(w_j) \leq r$.

Suppose the hypothesis holds for all derivations of length less than m. We consider cases corresponding to the possible forms of $\pi$.

Case 1. $\pi_1$ is $\sigma \rightarrow \alpha$; then $\alpha$ yields t by a subderivation of length less than m, hence deg $\eta(t) \leq r$ by the induction hypothesis.

Case 2. $\pi_1$ is $\alpha \rightarrow C\beta\gamma$; then (by Lemma 3.5) $t = Ct_1t_2$, where $\gamma$ yields $t_2$ by a subderivation of length less than m. Hence deg $(\eta(t_2)) \leq r$. By Lemma 2.5, deg $(\eta(Ct_1t_2))$ = deg $(\eta(t_1) * \eta(t_2)) \leq$ deg $(\eta(t_2))$. Hence deg $\eta(t) \leq r$.

Case 3. $\pi_1$ is $\alpha \rightarrow K\beta\gamma$; then $t = Kt_1t_2$, and (again by Lemma 3.5) $\beta$ yields $t_1$ and $\gamma$ yields $t_2$ by subderivations each of length less than m. Hence deg $\eta(t_1) \leq r$, deg $\eta(t_2) \leq r$. By Lemma 2.6, deg $(\eta(t))$ = deg $(\eta(Kt_1t_2))$ = deg $(\eta(t_1) * \eta(t_2))$ = max {deg $(\eta(t_1))$, deg $(\eta(t_2))$} $\leq r$.

Case 4. $\pi_1$ is $\alpha \rightarrow w_j$. Then m = 1, and we have dealt with this case.

Theorem 4.6: Every one-dimensional restricted g-set has finite K-depth.

Plan of Proof: Given a one-dimensional rg-set $\mathcal{S} = \eta(L(G))$ in (M,A), we construct from $G = (V, W_n, P, \sigma)$ a new grammar $G' = (U, W_n, P', \sigma(1,1))$ such that L(G') has finite K-depth and $\eta(L(G)) = \eta(L(G'))$. In the construction of G', all variables in U are of the form $\alpha(n_1, n_2)$ for certain positive integers $n_1$, $n_2$. They correspond to variables $\alpha$ in V, in the sense that collectively, the variables $\alpha(n_1, n_2)$ yield in M precisely those terms which $\alpha$ does; in particular, $\alpha(n_1, n_2)$ yields those elements of M which are derived from $\alpha$ in G and which have dimension $n_2 - n_1 + 1$. From this fact it will follow that the dimension of L(G') is one and

that $L(G')$ has finite K-depth. To show that $L(G) \subset L(G')$, we choose x in $L(G)$ and attempt to match to a leftmost derivation (A) of x, a leftmost derivation (B) of z in $L(G')$ such that $\eta(z) = \eta(x)$.

In the process of constructing (B), one production at a time, from (A), we develop for convenience an intermediate derivation $(\hat{A})$. It matches (A) in a sense to be defined precisely, except that some symbols in $(\hat{A})$ are "roofed", and matches (B) when the roofed symbols are erased. If the construction of (B) can be successfully carried out according to our algorithm, then we obtain a z in $L(G')$ such that $\eta(z) = \eta(x)$, and may conclude that $L(G) \subset L(G')$. The proof that the construction is always successful consists of a tedious examination of cases. The general plan for showing the reverse inclusion is similar. We will make repeated use of Lemma 3.5, without explicit mention, in the following fashion:

Given a derivation

$$(*) \qquad \sigma \xrightarrow[\pi_1]{} x_1 \xrightarrow[\pi_2]{} x_2 \rightarrow \cdots \xrightarrow[\pi_n]{} x_n = x,$$

where the $\pi_i$ denote productions, if $x_1 = C\beta\gamma$ (we could illustrate with $K\beta\gamma$ or $S\beta$ as well) then $x = Ct_1t_2$ for some $t_1$, $t_2$ such that $\beta$ yields $t_1$ and $\gamma$ yields $t_2$ by appropriate subderivations of (*).

Proof: Let $\Gamma$ be a one-dimensional rg-set in (M,A), where $A = \{a_1,\ldots,a_n\}$. Let $r = \max\{\deg(x) \mid x \in A\}$. We assume r greater than 0, for if $r = 0$, and $\Gamma$ is one-dimensional, then $\Gamma \subset A$, is finite, and clearly can be generated by a grammar of K-depth. Let $G = (V,W_n,P,\sigma)$ be a grammar in best form in $\mathcal{G}_n$ such that $\eta(L(G)) = \Gamma$. We construct from G a new grammar G' such that $\eta(L(G')) = \Gamma$, and $L(G')$ has K-depth no greater than r.

Let $V = A_m \cup B$, where $B = V \setminus A_m$, and $A_m$ is the set of Lemma 4.4. To each $\alpha$ in V, correspond a set $V_\alpha$ as follows:

(1) for $\alpha$ in $A_m$, $V_\alpha = \{\alpha(s,s) \mid 1 \leq s \leq r\}$.

(2) for $\alpha$ in B, $V_\alpha = \{\alpha(n_1,n_2) \mid 1 \leq n_1 \leq n_2 \leq r\}$.

Let $U = \bigcup_{\alpha \in V} V_\alpha$. Let $G' = (U, W_n, P', \sigma(1,1))$, where P' contains:

(1) $\sigma(1,1) \rightarrow \alpha(1,1)$, if $\sigma \rightarrow \alpha$ is in P.

(2) $\alpha(n_1,n_2) \rightarrow C\beta(n_1,n_2)\gamma(1,n_3)$, if $\alpha \rightarrow C\beta\gamma$ is in P, $\alpha(n_1,n_2)$ is in $V_\alpha$, $\gamma(1,n_3)$ is in $V_\gamma$, and $\beta(n_1,n_2)$ is in $V_\beta$.

(3) (i) $\alpha(n_1,n_2) \rightarrow K\beta(n_1k)\gamma(k+1,n_2)$, if $\alpha(n_1,n_2)$ is in $V_\alpha$, $\beta(n_1,k)$ is in $V_\beta$, $\gamma(k+1,n_2)$ is in $V_\gamma$, and $\alpha \rightarrow K\beta\gamma$ is in P.

(ii) $\alpha(n_1,r) \rightarrow \beta(n_1,r)$, if $\alpha(n_1,r)$ is in $V_\alpha$, $\beta(n_1,r)$ is in $V_\beta$, and $\alpha \rightarrow K\beta\gamma$ is in P.

(4) $\alpha(s,s) \rightarrow w_j$ if $\alpha(s,s)$ is in $V_\alpha$ and $\alpha \rightarrow w_j$ is in P.

Claim: If $\alpha(n_1,n_2)$ yields x for x in L(G'), then dim $\eta(x) = n_2-n_1+1$.

Proof of claim: By induction on the length m of a leftmost derivation,

$$x_0 = \alpha(n_1,n_2) \xrightarrow[p_1]{} x_1 \xrightarrow[p_2]{} x_2 \longrightarrow \cdots \xrightarrow[p_m]{} x_m = x.$$

If m = 1, then $p_1$ is $\alpha(n_1,n_2) \rightarrow w_j$. By an inspection of P', we see that $n_1 = n_2$, hence $n_2-n_1+1 = 1$. Since $\eta(w_j)$ is a phrase, the hypothesis is satisfied for m = 1.

Now suppose the hypothesis holds for $k \leq m$, and consider a derivation of length m+1.

Case 1. $p_1$ is $\alpha(n_1,n_2) \rightarrow C\beta(n_1,n_2)\gamma(1,s)$. Then $x = Ct_1t_2$, where $\beta(n_1,n_2)$ yields $t_1$, $\gamma(1,s)$ yields $t_2$; further, the subderivation of $t_1$ from $\beta(n_1,n_2)$ has length no greater than m. Hence dim $\eta(t_1) = n_2-n_1+1$. But by Lemma 2.5, dim $\eta(Ct_1t_2)$ = dim $(\eta(t_1)\cdot\eta(t_2))$ = dim $\eta(t_1)$, so the desired conclusion holds.

Case 2. $p_1$ is $\alpha(n_1,n_2) \rightarrow K\beta(n_1,k)\gamma(k+1,n_2)$. Then $x = Kt_1t_2$, where $\beta(n_1,k)$ yields $t_1$, $\gamma(k+1,n_2)$ yields $t_2$, both by subderivations of length less than m+1. Hence dim $\eta(t_1)$ =

$k-n_1+1$, dim $\eta(t_2) = n_2-k$, and therefore by Lemma 2.6,

$$\dim \eta(x) = \dim (\eta(Kt_1t_2)) = \dim (\eta(t_1)) + \dim \eta(t_2))$$
$$= k-n_1+1+n_2-k$$
$$= n_2-n_1+1.$$

Case 3: $p_1$ is $\alpha(n_1,r)\rightarrow\beta(n_1,r)$. Then $\beta(n_1,r)$ yields $x$ by a derivation of length $m$, and dim $(x) = r-n_1+1$, as required.

Case 4. $p_1$ is $\alpha(n_1,n_1)\rightarrow w_j$; occurs only when $m = 1$.

Hence in all possible cases, dim $\eta(x) = n_2-n_1+1$, as required.

Claim: L(G')) has K-depth $\leq r$.

Proof of claim: We have assumed $r\geq 1$. We show by induction on the length $m$ of a derivation that for any $\alpha(n_1,n_2) \in U$, if $\alpha(n_1,n_2)$ yields $x$, where $x$ is in $J_n$, then K-depth $(x)\leq r$. Let $\alpha(n_1,n_2) \xrightarrow{\pi_1} x_1 \xrightarrow{\pi_2} x_2 \rightarrow \ldots \xrightarrow{\pi_m} x_m = x$ be such a derivation. If $m = 1$, then $\pi_1$ is $\alpha(n_1,n_2)\rightarrow w_j$ for some $w_j \in W_n$. Hence $x = w_j$, and K-depth $(x) = 1\leq r$.

Suppose the hypothesis holds for all derivations of length less than $m$. We will examine the four cases corresponding to the possible forms of $\pi_1$.

Case 1. $\pi_1$ is $\alpha(n_1,n_2)\rightarrow w_j$; then $m = 1$, and this case has been dealt with.

Case 2. $\pi_1$ is $\alpha(n_1,n_2)\rightarrow C\beta(n_1,n_2)\gamma(1,s)$; then $x = Ct_1t_2$, where $\beta(n_1,n_2)$ yields $t_1$ and $\gamma(1,s)$ yields $t_2$ by subderivations each of length less than $m$. Hence by the induction hypothesis, K-depth $(t_1)\leq r$ and K-depth $(t_2)\leq r$. Now K-depth $(x)$ = K-depth $(Ct_1t_2)$ = max {K-depth $(t_1)$, K-depth $(t_2)$}$\leq r$.

Case 3. $\pi_1$ is $\alpha(n_1,n_2)\rightarrow K\beta(n_1,s)\gamma(s+1,n_2)$; then $x = Ct_1t_2$, where $\beta(n_1,s)$ yields $t_1$ and $\gamma(s+1,n_2)$ yields $t_2$, by subderivations each of length less than $m$. Hence K-depth $(t_1)\leq r$ and K-depth $(t_2)\leq r$. Since K-depth $(Kt_1t_2)$ =

max {K-depth $(t_1)$, K-depth $(t_2)$, dim $\eta(Kt_1t_2)$}, we have K-depth $(Kt_1t_2) \leq r$ from the fact that dim $\eta(Kt_1t_2) = n_2-n_1+1 \leq r$.

Case 4. $\pi_1$ is $\alpha(n_1,n_2) \to \beta(n_1,n_2)$; then $\beta(n_1,n_2)$ yields x by a derivation of length less than m, hence K-depth $(x) \leq r$.

Claim: L(G') is one-dimensional.

Proof of claim: For all x in L(G'), we have $\sigma_{(1,1)}$ yields x. Hence dim $(\eta(x)) = 1-1+1 = 1$.

Now let $\mathcal{L}' = L(G')$. We will show that $\mathcal{L} = \mathcal{L}'$. First we show that $\mathcal{L} \subset \mathcal{L}'$. Let x be an element of L(G), and let

$$(A)\quad x_0 = \sigma \xrightarrow[\pi_1]{G} x_1 \xrightarrow[\pi_2]{G} x_2 \xrightarrow{G} \cdots \xrightarrow[\pi_{n-1}]{G} x_{n-1} \xrightarrow[\pi_n]{G} x_n = x$$

be a leftmost G-derivation, where the $\pi_i$ are productions in P, $1 \leq i \leq n$. We will attempt to construct a matching derivation

$$(B)\quad y_0 = \sigma(1,1) \xrightarrow[p_1]{G'} z_1 \xrightarrow[p_2]{G'} z_2 \longrightarrow \cdots \xrightarrow[p_{n-1}]{G'} z_{n-1} \xrightarrow[p_n]{G'} z_n = z$$

for productions $p_i$ in P', such that $\eta(x) = \eta(z)$. [In (B), for convenience we adopt the convention that either (i) $p_i \in P'$ or (ii) $p_i$ is a "place-holding" symbol only, and $z_{i-1} = z_i$.] As we proceed, we will have use also for a "dummy" derivation

$$(\hat{A})\quad y_0 = \sigma \xrightarrow[q_1]{} y_1 \xrightarrow[q_2]{} \cdots \xrightarrow[q_{n-1}]{} y_{n-1} \xrightarrow[q_n]{} y_n = y$$

which will be constructed along with (B), in such a way that it differs from (A) only in that (possibly) some variables $\alpha$ in A appear as $\hat{\alpha}$ in $(\hat{A})$. The symbols $\hat{\alpha}$ will be called roofed symbols. The process of construction follows:

1. Let i = 1; let $y_0 = \sigma$; let $z_0 = \sigma(1,1)$. By the form of G, $\pi_1$ is $\sigma \to \alpha$ for some $\alpha$; let $p_1$ and $q_1$ be $\sigma(1,1) \to \alpha(1,1)$.

2. If $x_i$ and $y_i$ are identical except that (possibly) some symbols in $y_i$ are roofed, then call $x_i$ and $y_i$ almost identical. In such case, continue. Otherwise, the construction has failed.

3. Let e(y) be the string resulting from the erasure of all roofed symbols in $y_i$. For any strings $X = \beta_1 \ldots \beta_s$, $Y = \gamma_1 \ldots \gamma_t$, for any i,j, $1 \leq i \leq s$, $1 \leq j \leq t$, we say that $\beta_i$ matches $\gamma_j$ if (i) $i = j$ and (ii) either (a) $\beta_i = \gamma_j$ or (b) $\beta_i$ is a variable and $\gamma_j \in V_{\beta_i}$.

If $s = t$ and $\beta_i$ matches $\gamma_i$ for $1 \leq i \leq s$, then we say X matches Y.

If $e(y_i)$ matches $z_i$, continue. Otherwise the construction has failed.

4. For each variable $\alpha(n_1,n_2)$ in $z_i$, examine the matching variable $\alpha$ in $x_i$. The string $x_i$ has the form

$$x_i = u\ \alpha\ v, \text{ where } u,v \in (V \cup W_n \cup \{C,K\})^*.$$

The word x has the form $x = t_1 t_2 t_3$, where $t_1$, $t_2$, $t_3 \in (W_n \cup \{C,K\})^*$, and by an appropriate subderivation of (A), u yields $t_1$, $\alpha$ yields $t_2$, and v yields $t_3$.

4.1. If $n_2 < r$, and $\dim(\eta(t_2)) \neq n_2 - n_1 + 1$, the construction has failed. If $n_2 = r$, and $\dim(\eta(t_2)) < n_2 - n_1 + 1$, then the construction has failed. Otherwise continue.

4.2. To each occurrence of a variable $\alpha(n_1,n_2)$ in $z_i$ with matching variable $\alpha$ in $x_i$ as above, we correspond a collection of terms in $\mathcal{J}_n$ called the substitutes of $\alpha(n_1,n_2)$ and denoted by sub $(\alpha(n_1,n_2))$. Let sub $(\alpha(n_1,n_2))$ be the collection $\eta^{-1}((1^*\ldots{}^*n_2-n_1+1)\cdot\eta(t_2))$.

The substitutes of $z_i$ [sub $(z_i)$] will be the collection of all terms in $\mathcal{J}_n$ which can be formed by replacing each variable $\alpha(n_1,n_2)$ in $z_i$ by some element of sub $\alpha(n_1,n_2)$ for that occurrence of $\alpha(n_1,n_2)$ in $z_i$.

If, for all t in sub $(z_i)$, $\eta(t) = \eta(x)$, continue; otherwise the construction has failed.

5. If $i = n$, the construction is complete, and successful. Otherwise, add 1 to i and continue.

6. Next we choose $p_{i+1}$ and $q_{i+1}$. We distinguish four cases, depending on the form of $\pi_{i+1}$ in P.

Case 1. $x_i = u\ \alpha\ v$, $\pi_{i+1}$ is $\alpha \to C\beta\gamma$.

1A. The matching occurrence of $\alpha$ in $y_i$ is roofed. Let $q_{i+1}$ be $\hat{\alpha} \to \hat{C}\hat{\beta}\hat{\gamma}$, and let $p_{i+1}$ be a place-holder only, so that $z_i = z_{i+1}$.

1B. The matching occurrence of $\alpha$ in $y_i$ is not roofed. Let $q_{i+1}$ be $\alpha \to C\beta\gamma$. To choose $p_{i+1}$, note that by step 3, there is a matching symbol $\alpha(n_1,n_2)$ in $z_i$. Examine (A). With notation as in step 4, we have $x = t_1t_2t_3$, and $\alpha$ yields $t_2$. Since (A) is a leftmost derivation, we now know that the first step in the derivation of $t_2$ from $\alpha$ is $\pi_{i+1}$; that is, the associated sub-derivation has the form $\alpha \to C\beta\gamma \to \ldots \to Ct_4t_5 = t_2$, for some terms $t_4$, $t_5$ in $J_n$. Let $s = \dim(\eta(t_5))$, and let $p_{i+1}$ be $\alpha(n_1,n_2) \to C\beta(n_1,n_2)\gamma(1,s)$. Let us make sure that this production is in P'. Since $\alpha(n_1,n_2)$ has appeared, it is in V; further, if $\beta(n_1,n_2) \notin V_\beta$, then $\beta \in A_m$ and $n_1 \neq n_2$. However, $\beta$ yields $t_4$, where $\dim \eta(t_4) = 1$ by Lemma 4.6; hence $\dim \eta(Ct_4t_5) = 1$. But by step 4.1, since $t_2 = Ct_4t_5$, we know that $\dim \eta(Ct_4t_5) \geq n_2 - n_1 + 1$. This, along with the fact that $n_1 \leq n_2$, gives $n_1 = n_2$, a contradiction. Hence $\beta(n_1,n_2)$ is in $V_\beta$ and $p_{i+1}$ is in P'.

Case 2. $x_i = u\ \alpha\ v$, $\pi_{i+1}$ is $\alpha \to K\beta\gamma$.

2A. The matching occurrence of $\alpha$ in $y_i$ is roofed. Let $q_{i+1}$ be $\hat{\alpha} \to \hat{K}\hat{\beta}\hat{\gamma}$, and let $p_{i+1}$ be a place-holder only, so that $z_i = z_{i+1}$.

2B. The matching occurrence of $\alpha$ in $y_i$ is not roofed. Then there is a matching variable $\alpha(n_1,n_2)$ in $z_i$. Examine (A). The subderivation $\alpha \quad t_2$ now can be seen to have the form $\alpha \to K\beta\gamma \to Kt_4t_5 = t_2$, for some $t_4$, $t_5$ in $J_n$. Suppose $s_1 = \dim(\eta(t_4))$, and $s_2 = \dim(\eta(t_5))$. Then $\dim(\eta(t_2)) = s_1 + s_2$, by Lemma 2.6. We distinguish three cases, depending on the value of $n_2$ and of $n_1 + s_1 - 1$.

2B(i). $1 \leq n < r$. Then by step 4, $n_2 - n_1 + 1 = s_1 + s_2$. Let $p_{i+1}$ be $\alpha(n_1,n_2) \rightarrow K\beta(n_1,n_1+s_1-1)\gamma(n_1+s_1,n_2)$ and let $q_{i+1}$ be $\alpha \rightarrow K\beta\gamma$.

To see that $p_{i+1} \in P'$: if not, then either (1) $\beta(n_1,n_1+s_1-1)$ is not in $V_\beta$, $\beta \in A_m$ and $s_1 > 1$, or (2) $\gamma(n_1+s_1,n_2)$ is not in $V_\gamma$, $\gamma \in A_m$ and $s_2 > 1$, or both. In the first case, we have $\beta$ yields $t_4$, and dim $\eta(t_4) > 1$, a contradiction; in the second case we have a similar contradiction.

2B(ii). $n_2 = r$, $n_1 + s_1 - 1 < r$. Let $p_{i+1}$ be $\alpha(n_1,r) \rightarrow K\beta(n_1,n_1+s_1-1)\gamma(n_1+s_1,r)$ and $q_{i+1}$ be $\alpha \rightarrow K\beta\gamma$.

Again, if $p_{i+1} \notin P'$, then either $\beta \in A_m$ and $s_1 > 1$, a contradiction since $\beta$ yields $t_4$ and dim $\eta(t_4) = s_1$; or $\gamma \in A_m$ and $r - s_1 - n_1 > 0$. But by step 4, $n_1 + s_1 + s_2 - 1 \geq r$; that is, $s_2 \geq r - s_1 - n_1 + 1 > 1$. So $s_2 > 1$ and $\gamma$ yields $t_5$, where dim $\eta(t_5) = s_2$, a contradiction.

2B(iii). $n_2 = r$, $n_1 + s_1 - 1 \geq r$. Let $p_{i+1}$ be $\alpha(n_1,r) \rightarrow \beta(n_1,r)$, and let $q_{i+1}$ be $\alpha \rightarrow \hat{K}\beta\hat{\gamma}$.

If $p_{i+1} \notin P'$, then $\beta \in A_m$, and $n_1 < r$. Combining this with the inequality $n_1 + s_1 - 1 \geq r$, we conclude $s_1 > 1$, a contradiction, since $\beta$ yields $t_4$, dim $\eta(t_4) = s_1$. So $p_{i+1} \in P$.

Case 3. $x_i = u \cdot \alpha v$, $\pi_{i+1}$ is $\alpha \rightarrow w_j$.

3A. The matching occurrence of $\alpha$ in $y_i$ is roofed. Let $q_i$ be $\hat{\alpha} \rightarrow \hat{w}_j$, and let $p_i$ be a placeholder, so that $z_{i+1} = z_i$.

3B. The matching occurrence of $\alpha$ in $y_i$ is not roofed. Then suppose $\alpha(n_1,n_2)$ is the matching occurrence in $z_i$. It is now clear that the subderivation by which $\alpha$ yields $t_2$ is precisely $\alpha \xrightarrow[\pi_{i+1}]{} w_j = t_2$. Since dim $(\eta(t_2))$ = dim $(\eta(w_j)) = 1$, by step 4 we have:

(i) if $n_2$ r, $n_2 - n_1 + 1 = 1$ hence $n_1 = n_2$;

(ii) if $n_2 = r$, $n_2 - n_1 + 1 \leq 1$, which also yields $n_1 = n_2$, since $n_2 \geq n_1$.

So, in any case, $n_1 = n_2$, and we let $p_{i+1}$ be $\alpha(n_1,n_1)\rightarrow w_j$, and let $q_{i+1}$ be $\alpha\rightarrow w_j$.

Case 4. $\pi_{i+1} = \sigma\rightarrow\alpha$. Because of the form of G, this case appears if and only if i = 0; hence we need not consider it.

Now return to step 2. This completes the detail of the construction. To clarify the construction, we present an example below of a possible derivation (A) and the associated derivations ($\hat{A}$) and (B), when r = 2.

(A) $\sigma \xrightarrow[\pi_1]{} \alpha \xrightarrow[\pi_2]{} C\beta\gamma \xrightarrow[\pi_3]{} Cw_2\gamma \xrightarrow[\pi_4]{} Cw_2K\xi\tau \xrightarrow[\pi_5]{} Cw_2KK\alpha\gamma\tau \xrightarrow[\pi_6]{}$

$Cw_2KKw_1\gamma\tau \xrightarrow[\pi_7]{} Cw_2KKw_1w_2\tau \xrightarrow[\pi_8]{} Cw_2KKw_1w_2w_3.$

($\hat{A}$) $\sigma \xrightarrow[q_1]{} \alpha \xrightarrow[q_2]{} C\beta\gamma \xrightarrow[q_3]{} Cw_2\gamma \xrightarrow[q_4]{} Cw_2\hat{K}\xi\hat{\tau} \xrightarrow[q_5]{} Cw_2KK\alpha\gamma\hat{\tau} \xrightarrow[q_6]{}$

$Cw_2\hat{K}Kw_1\gamma\hat{\tau} \xrightarrow[q_7]{} Cw_2\hat{K}Kw_1w_2\hat{\tau} \xrightarrow[q_8]{} Cw_2\hat{K}Kw_1w_2w_3.$

(B) $\sigma(1,1) \xrightarrow[p_1]{} \alpha(1,1) \xrightarrow[p_2]{} C\beta(1,1)\gamma(1,2) \xrightarrow[p_3]{} Cw_2\gamma(1,2) \xrightarrow[p_4]{}$

$Cw_2\xi(1,2) \xrightarrow[p_5]{} Cw_2K\alpha(1,1)\gamma(1,1) \xrightarrow[p_6]{} Cw_2K2_1\gamma(1,1) \xrightarrow[p_7]{}$

$Cw_2Kw_1w_2 \xrightarrow[p_8]{} Cw_2Kw_1w_2.$

If, for each x in L(G), the construction can be successfully carried out, then we obtain a z in L(G') such than $\eta(x) = \eta(z)$. For notice that sub $(z_n) = \{z_n\} = \{z\}$ since $z_n$ contains no variables, and by step 4, $\eta(z) = \eta(x)$. Hence we may conclude that $\mathcal{L} \subset \mathcal{L}'$.

We will next show that the construction can always be successfully completed. If it fails, it must fail at step 2, 3, or 4, for some i>0. We will show by induction on i that such failure is not possible. Suppose i = 1. Steps 2 and 3 are trivially satisfied. We have $z_1 = \alpha(1,1)$; $t_2 = x$. Since $\mathcal{L}$ is one-dimensional, dim $(\eta(x)) = 1$, satisfying the first condition of step 4. If r>1, then sub $[\alpha(1,1)] =$

{x}, and condition 4.2 is satisfied. If r = 1, then sub [α(1,1)] = $\eta^{-1}((1)\cdot\eta(x))$, and for t in sub [α(1,1)], $\eta(t) = 1\cdot\eta(x) = \eta(x)$, since dim $(\eta(x)) = 1$. Hence the construction never fails for i = 1.

Suppose, for x in L(G), with leftmost derivation (A), the construction fails for the first time for some i+1, $i \geq 1$, at step 2. We have $x_i = u\ \alpha\ v$, $y_i = u'\ \alpha\ v'$ or $u'\ \hat{\alpha}\ v'$, where u and u' are almost identical and v and v' are almost identical. An inspection of the choice of $q_i$ shows that, if $\pi_{i+1}$ is $\alpha \to t$, whatever the form of t, the production $q_{i+1}$ is $\alpha \to t'$ or $\hat{\alpha} \to t'$, for some t' such that t and t' are almost identical. Hence $x_{i+1} = utv$ and $y_{i+1} = u't'v'$ are almost identical, a contradiction, and there is no failure at step 2.

Suppose there is a failure at step 3. If $x_i = u\ \alpha\ v$, and $y_i = u'\ \hat{\alpha}\ v'$, then $q_i$ yields only roofed variables, so $e(y_{i+1}) = e(y_i)$. Also, $p_i$ is only a place-holder, so $z_{i+1} = z_i$, and since $e(y_i)$ matches $z_i$, $e(y_{i+1})$ matches $z_{i+1}$.

If $x_i = u\ \alpha\ v$ and $y_i = u'\ \alpha\ v'$, $z_i = u''\alpha(n_1,n_2)v''$, where u" matches e(u'), v" matches e(v'), then the possible forms for $y_{i+1}$, $e(y_{i+1})$, and $z_{i+1}$ are:

| $y_{i+1}$ | $e(y_{i+1})$ | $z_{i+1}$ |
|---|---|---|
| $u'C\beta\gamma v'$ | $e(u')C\beta\gamma e(v')$ | $u''C\beta(n_1,k)\gamma(k+1,n_2)v''$ |
| $u'K\beta\gamma v'$ | $e(u')K\beta\gamma e(v')$ | $u''K\beta(n_1,k)\gamma(k+1,n_2)v''$ |
| $u'\hat{K}\beta\hat{\gamma}v'$ | $e(u')\beta e(v')$ | $u''\beta(n_1,r)v''$ |
| $u'w_jv'$ | $e(u')w_je(v')$ | $u''w_jv''$ |

In each case, $e(y_{i+1})$ matches $z_{i+1}$; hence another contradiction. The algorithm does not fail at step 3.

Then the construction must fail at step 4. We assume $p_{i+1}$ is not a place-holder, since otherwise step 4 is identical to the i-th step 4, hence succeeds as before.

Again we have $x_i = u\ \alpha\ v$, $z_i = u''\alpha(n_1,n_2)v''$, $\pi_{i+1}$ is $\alpha \to t$, $p_{i+1}$ is $\alpha \to t''$ for some strings $t,t''$, and $x_{i+1} = utv$, $z_{i+1} = u''t''v''$. Since the construction succeeded for $i$, it can fail on condition 4.1 only for variables $\beta(k_1,k_2)$ which appear in $t''$. The subderivation by which $\alpha$ yields $t_2$ is now seen to be $\alpha \xrightarrow[\pi_{i+1}]{} t \Rightarrow t_2$, and dim $\eta(t_2) = n_2-n_1+1$, if $n_2<r$; dim $\eta(t_2)\geq n_2-n_1+1$, if $n_2 = r$. Again we must distinguish cases depending on the form of $t$.

Case 1. $t = C\beta\gamma$; then $t_2 = Ct_4t_5$, with subderivations $\beta \to t_4$, $\gamma \to t_5$. By the choice of $p_{i+1}$, $t'' = C\beta(n_1,n_2)\gamma(1,s)$, where

$$s = \begin{cases} \dim \eta(t_5), & \text{if } \dim \eta(t_5) \leq r \\ r, & \text{if } \dim \eta(t_5) > r. \end{cases}$$

By Lemma 2.5, dim $(\eta(t_4))$ = dim $(\eta(t_2))$; by the previous application of step 4, dim $(\eta(t_2)) = n_2-n+1$ if $n_2<r$, and if $n_2 = r$, then dim $(\eta(t_2))\geq n_2-n_1+1$. By the choice of $s$, dim $(\eta(t_5)) = s = s-1+1$. So this case does not fail.

Case 2. $t = K\beta\gamma$; then $t_2 = Kt_4t_5$, with subderivations $\beta \to t_4$, $\gamma \to t_5$. Then by the choice of $p_{i+1}$, either

(1) $n_2 = r$, $t'' = \beta(n_1,r)$, in which case $n_1+\dim(\eta(t_4))-1\geq r$,

or (2) $n_2 = r$, $t'' = K\beta(n_1k)\gamma(k+1,r)$, in which case dim $\eta(t_4) = k-n_1+1$, dim $\eta(t_5)\geq r-k$,

or (3) $n_2<r$, $t'' = K\beta(n_1,k)\gamma(k+1,n_2)$, and dim $\eta(t_4) = k-n_1+1$; dim $\eta(t_5) = n_2-k$.

In each case, 4.1 is satisfied.

Case 3. $t = w_j$; then $t'' = w_j$, and no untested variable appears. So condition 4.1 is satisfied.

Now the only condition the construction may fail to satisfy is 4.2.

We will assume, then, that for some $w$ in sub $(z_{i+1})$, $(w) \neq \eta(x)$. This implies that $w$ is not in sub $(z_i)$, by

the minimality of i+1. The only way this can happen is that $p_{i+1}$ is $\alpha(n_1,n_2)\to t''$ for some string t", where sub (t") is not contained in sub $[\alpha(n_1,n_2)]$ for the occurrence of $\alpha(n_1,n_2)$ to which $p_{i+1}$ is applied. We will show by an examination of all possible forms of t that this is not possible, and thereby will conclude that for all w in sub $(z_{i+1})$, $\eta(w) = \eta(x)$. This contradiction will complete the proof that the construction is always possible.

There are several cases, corresponding to the possible forms of t and t".

Case 1. $t = C\beta\gamma$, $t'' = C\beta(n_1,n_2)\gamma(1,s)$. For $\tau$ in sub (t"), $\tau = Cab$ for a in sub $[\beta(n_1,n_2)]$ and b in sub $[\gamma(1,s)]$. For such a,b, a is in $\eta^{-1}((1*\ldots*n_2-n_1+1)\cdot\eta(t_4))$, and b is in $\eta^{-1}((1*\ldots*s)\cdot\eta(t_5))$.

1A. $s<r$. Then

$$\eta(Cab) = (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot(1*\ldots*s)\cdot\eta(t_5).$$

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot\eta(t_5),$$

since dim $(\eta(t_5)) = s$,

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(Ct_4t_5);$$

since $Ct_4t_5 = t_2$, Cab is in sub $[\alpha(n_1,n_2)]$.

1B. $s = r$. Then note that (by Lemma 4.5), deg $(\eta(t_4))\leq r$, hence $\eta(t_1)\cdot(1*\ldots*r) = (t_4)$, and

$$\eta(Cab) = (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot(1*\ldots*r)\cdot\eta(t_5)$$

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot\eta(t_5), \text{ as before.}$$

Hence Cab is in sub $[\alpha(n_1,n_2)]$.

Case 2. $t = K\beta\gamma$.

2A. $n_2 = r$, $t'' = \beta(n_1,r)$. Again there must be an a in sub $[\beta(n_1,r)]$ which is not in sub $[\alpha(n_1,n_2)]$ for the occurrence of $\alpha$ in question. For such a, a is in $\eta^{-1}((1*\ldots*r-n_1+1)\cdot\eta(t_4))$. But by the construction, dim $\eta(t_4)\geq r-n_1+1$; hence

$$(1*\ldots*r-n_1+1)\cdot\eta(t_4) = (1*\ldots*r-n_1+1)\cdot(\eta(t_4)*\eta(t_5))$$

$$= (1*\ldots*r-n_1+1)\cdot(\eta(Kt_4t_5))$$

$$= (1*\ldots*r-n_1+1)\cdot(\eta(t_2));$$

hence

$$a \in \eta^{-1}((1*\ldots*r-n_1+1)\cdot\eta(t_2)) = \text{sub } [\alpha(n_1,n_2)].$$

2B. $n_2 = r$; $t'' = K\beta(n_1,k)\gamma(k+1,r)$.

If $\tau$ is in sub $(t'')$, $\tau = Kab$ for some a in sub $[\beta(n_1,k)]$, and some b in sub $[\gamma(k+1,r)]$. For such a,b, we have $a \in \eta^{-1}((1*\ldots*k-n_1+1)\cdot\eta(t_4))$,

$b \in \eta^{-1}((1*\ldots*r-k)\cdot\eta(t_5))$, and

$$\eta(Kab) = [(1*\ldots*k-n_1+1)\cdot\eta(t_4)]*[(1*\ldots*r-k)\cdot\eta(t_5)]$$

$$= (1*\ldots*r-n_1+1)\cdot\eta(t_4)*\eta(t_5)),$$

since dim $\eta(t_4) = k-n_1+1$ and dim $\eta(t_5) \geq r-k$,

$$= (1*\ldots*r-n_1+1)\cdot(\eta(Kt_4t_5)).$$

Hence Kab is in sub $[\alpha(n_1,n_2)]$.

Case 3. $t = w_j$, $t'' = w_j$. Then $n_1 = n_2$, $t_2 = w_j$. Since $\eta(w_j)$ is a phrase, sub $[\alpha(n_1,n_2)] = \eta^{-1}(1\cdot\eta(t_2)) = \eta^{-1}(\eta(w_j))$. Hence $w_j$ is in sub $[\alpha(n_1,n_2)]$.

So the construction did not fail for i>1 at any step; hence all constructions can be completed successfully.

This completes the proof that $\mathcal{L} \subset \mathcal{L}'$.

To show that $\mathcal{L}' \subset \mathcal{L}$, let z be in L(G'), with leftmost derivation

(B) $z_0 = \sigma(1,1) \xrightarrow[p_1]{} z_1 \xrightarrow[p_2]{} z_2 \longrightarrow \ldots \xrightarrow[p_n]{} z_n = z.$

We construct a matching derivation

(A) $x_0 = x_0' = \sigma \xrightarrow[\pi_1]{} x_1 \xrightarrow{\Pi_1} x_1' \xrightarrow[\pi_2]{} x_2 \longrightarrow \ldots \xrightarrow[\pi_n]{} x_n \xrightarrow{\Pi_n} x_n' =$

x, where $\pi_i$, $1 \leq i \leq n$ are in P, and the expressions $\Pi_i = \pi_{i1},\ldots,\pi_{im_i})$ represent (possibly empty) sequences of productions $\pi_{ij}$ in P.

We will again have use for a dummy derivation

(A') $y_0 = y_0' = \sigma \xrightarrow[1]{} y_1 \xrightarrow{Q_1} y_1' \xrightarrow[q_2]{} y_2 \longrightarrow \ldots \xrightarrow[q_n]{} y_n \xrightarrow{Q_n} y_n' = y,$

where $Q_i = (q_{i1},\ldots,q_{im_i})$ is a sequence which we construct from $\Pi_i$.

We will show that $\eta(z) = \eta(x)$. The construction is similar to the earlier one.

1. Let $i = 1$; let $\sigma = x_0 = x_0' = y_0 = y_0'$. An inspection of P' shows that $p_1$ is $\sigma(1,1)\rightarrow\alpha(1,1)$ for some $\alpha$; let $\pi_1$ and $q_1$ be $\sigma\rightarrow\alpha$, and let $\Pi_1$ and $Q_1$ be empty.

2. If $x_i'$ and $y_i'$ are almost identical, continue. Otherwise the construction has failed.

3. If $e(y_i')$ matches $z_i$, continue; otherwise the construction has failed.

4. Now we define sub $(x_i')$. We define a substitution for an occurrence of a variable $\alpha$ in $x_i'$ as follows:

(1) if $\alpha$ is in $A_m$, sub $(\alpha) = \eta^{-1}(1\cdot\eta(t_2))$, where as before we have $\alpha(n_1,n_2)$ yields $t_2$ in (B) for the matching variable $\alpha(n_1,n_2)$ in (B).

(2) if $\alpha$ is not in $A_m$, and $n_2<r$, where $\alpha(n_1,n_2)$ is the matching variable in (B), and $\alpha(n_1,n_2)$ yields $t_2$ in (B), then sub $(\alpha) = \eta^{-1}((1*\ldots*n_2-n_1+1)\cdot\eta(t_2))$.

(3) if $\alpha$ is not in $A_m$ and $n_2 = r$, then sub $(\alpha) = \bigcup_{k\geq 0} \eta^{-1}([(1*\ldots*r-n_2+1)\cdot\eta(t_2)]*b_1*\ldots*b_k)$, where for $1\leq i\leq k$, $b_i$ is any phrase in M.

When all possible substitutions have been made for each variable, call the resulting collection of terms sub $(x_i')$. If, for all t in sub $(x_i')$, $\eta(t) = \eta(z)$, then continue. Otherwise the construction has failed.

5. If $i = n$, the construction is successful. Otherwise, add 1 to i and continue.

6. Let us now choose $\pi_{i+1}$, $q_{i+1}$, $\Pi_{i+1}$ and $Q_{i+1}$. We consider four cases, depending on the form of $P_{i+1}$.

Case 1. $p_{i+1}$ is $\alpha(n_1,n_2)\rightarrow C\beta(n_1,n_2)\gamma(1,s)$. Let $q_{i+1}$ and $\pi_{i+1}$ be $\alpha\rightarrow C\beta\gamma$. Let $\Pi_{i+1}$ and $Q_{i+1}$ be placeholders, i.e. empty sequences.

Case 2. $p_{i+1}$ is $\alpha(n_1,n_2)\rightarrow K\beta(n_1,k)\gamma(k+1,n_2)$. Let $q_{i+1}$ and $\pi_{i+1}$ be $\alpha\rightarrow K\beta\gamma$, and let $\Pi_{i+1}$ and $Q_{i+1}$ be placeholders.

Case 3. $p_{i+1}$ is $\alpha(n_1,r)\rightarrow\beta(n_1,r)$. Then, by the construction of P', there is a variable $\gamma$ in V, and a production $\pi$ in P, such that $\pi$ is $\alpha\rightarrow K\beta\gamma$. Let $\pi_{i+1}$ be $\pi$ for any such $\pi$, and let $q_{i+1}$ be $\alpha\rightarrow\hat{K}\beta\hat{\gamma}$. Since G is in best form, there is an element u in $\mathcal{J}_n$ and a sequence of productions $\Pi_{i+1} = (\pi_{(i+1)1},\ldots,\pi_{(i+1)m_{(i+1)}})$ such that $\gamma$ yields u by the leftmost application of these productions. Apply this sequence to $\gamma$ in (A), forming $x'_{i+1}$. Let the corresponding roofed sequence by $Q_{i+1}$, which when applied yields $y'_{i+1}$.

Case 4. $p_{i+1}$ is $\alpha(s,s)\rightarrow w_j$. Let $q_{i+1}$ and $\pi_{i+1}$ be $\alpha\rightarrow w_j$, and let $\Pi_{i+1}$ and $Q_{i+1}$ be placeholders.

This completes the construction. Now when we have shown that it is always possible, we may conclude that $\mathcal{J}' \subset \mathcal{J}$; for, when i = n, there are no variables in $z_i$, and sub $(z'_n) = \{z'_n\} = \{z\}$; hence $\eta(z) = \eta(x)$.

It is easy to see by an argument analogous to that in the first half of the proof that no failure in the construction can come at steps 2 or 3.

We consider step 4, and show by induction on i that no failure can occur there. Suppose i = 1. Then for some $\alpha$ in V, $z_1 = \alpha$, and $t_2 = x$. By the construction of G', the production $\sigma(1,1)\rightarrow\alpha(1,1)$ appears in P' if and only if $\gamma(1,s)$ yields $t_5$ by appropriate subderivations.

If $\tau$ is in sub (t"), then $\tau = Cuv$ for some u in sub ($\beta$), some v in sub ($\gamma$). Note that in this case, $\Pi_{i+1}$ is the empty sequence, and $z_{i+1} = z'_{i+1}$.

1A. $\beta$ and $\gamma$ are both in $A_m$: Then $n_1 = n_2$, by the construction of P', and sub $(\beta) = \eta^{-1}(1\cdot\eta(t_4))$. Also, s = 1, and sub $(\gamma) = \eta^{-1}(1\cdot\eta(t_5))$. By Lemma 4.4, dim $\eta(t_5)$ = 1 and dim $\eta(t_4)$ = 1. Hence for all u in sub ($\beta$), for all v in sub ($\gamma$),

$$\eta(Cuv) = 1\cdot\eta(t_4)\cdot 1\cdot\eta(t_5)$$
$$= 1\cdot\eta(t_4)\cdot\eta(t_5)$$

$= 1 \cdot \eta(Ct_4t_5)$

$= 1 \cdot \eta(t_2)$; hence

Cuv is in sub ($\alpha$), a contradiction.

1B. $\beta$ is in $A_m$, $\gamma$ is not in $A_m$: Then sub $\beta = \eta^{-1}(1 \cdot \eta(t_4))$.

1B. (i). $s<r$. sub $(\gamma) = \eta^{-1}((1*\ldots*s) \cdot \eta(t_5))$

$= \eta^{-1}(\eta(t_5))$,

since dim $\eta(t_5) = s$.

$\eta(Cuv) = 1 \cdot \eta(t_4) \cdot \eta(t_5)$

$= 1 \cdot \eta(Ct_4t_5)$

$= 1 \cdot \eta(t_2)$;

hence Cuv is in sub ($\alpha$).

1B. (ii). $s = r$.

sub $(\gamma) = \bigcup_{k \geq 0} ([1*\ldots*s \cdot \eta(t_5)]*b_1*\ldots*b_k)$.

$\eta(Cuv) = 1 \cdot \eta(t_4) \cdot \{[(1*\ldots*r) \cdot \eta(t_5)]*b_1*\ldots*b_k\}$ for some $k \geq 0$, some phrases $b_i$, $1 \leq i \leq k$.

$= 1 \cdot \eta(t_4) \cdot (1*\ldots*r) \cdot [[(1*\ldots*r) \cdot \eta(t_5)]*b_1*\ldots*b_k]$, since by Lemma 4.5, deg $(\eta(t_4)) \leq r$,

$= 1 \cdot \eta(t_4) \cdot (1*\ldots*r) \cdot \eta(t_5)$, since dim $\eta(t_5) \geq r$

$= 1 \cdot \eta(t_4) \cdot \eta(t_5)$

$= 1 \cdot \eta(Ct_4t_5)$

$= 1 \cdot \eta(t_2)$;

hence Cuv is in sub ($\alpha$).

1C. $\beta$ is not in $A_m$, $\gamma$ is in $A_m$: Then sub $(\gamma) = \eta^{-1}(1 \cdot \eta(t_5))$.

1C. (i). $n_2<r$. Then sub $\beta = \eta^{-1}[(1*\ldots*n_2-n_1+1) \cdot \eta(t_4)]$

$\eta(Cuv) = (1*\ldots*n_2-n_1+1) \cdot \eta(t_4) \cdot 1 \cdot \eta(t_5)$.

$= (1*\ldots*n_2-n_1+1) \cdot \eta(t_4) \cdot \eta(t_5)$, since by Lemma 4.4, dim $\eta(t_5) = 1$;

$= (1*\ldots*n_2-n_1+1) \cdot \eta(t_2)$; hence Cuv is in sub ($\alpha$).

1C. (ii). $n_2 = r$. Then

$$\text{sub } \beta = \bigcup_{k\geq 0} \eta^{-1}[[[1*\ldots*n_2-n_1+1]\cdot\eta(t_4)]*b_1*\ldots*b_k].$$

$$\eta(Cuv) = [(1*\ldots*n_2-n_1+1)\cdot\eta(t_4)]*b_1*\ldots*b_k*(1\cdot\eta(t_5)),$$

$$= [(1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot\eta(t_5)]*[b\cdot\eta(t_5)]*\ldots*[b_k\cdot\eta(t_5)].$$

$$= [(1*\ldots*n_2-n_1+1)\cdot\eta(t_2)]*[b_1\cdot\eta(t_5)]*\ldots*[b_k\cdot\eta(t_5)],$$

which is in sub ($\alpha$), since the fact that $\beta$ is not in $A_m$ implies that $\alpha$ is not in $A_m$.

1D. Neither $\beta$ nor $\alpha$ is in $A_m$:

1D. (i). $n_2<r$, $s<r$.

$$\eta(Cuv) = (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot(1*\ldots*s)\cdot\eta(t_5)$$

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot\eta(t_5), \text{ since dim } \eta(t_5) = s,$$

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(t_2);$$

hence Cuv is in sub ($\alpha$).

1D. (ii). $n_2<r$, $s = r$.

$$\eta(Cuv) = [(1*\ldots*n_2-n_1+1)\cdot\eta(t_4)]\ [[(1*\ldots*r)\cdot\eta(t_5)]*b_1*\ldots*b_k],$$

for some phrases $b_i$, $1\leq i\leq k$,

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot(1*\ldots*r)\cdot\eta(t_5),$$

since deg $(\eta(t_4))\leq r$ and dim $(\eta(t_5))\geq r$,

$$= (1*\ldots*n_2-n_1+1)\cdot\eta(t_4)\cdot\eta(t_5);$$

hence $\eta(Cuv)$ is in sub ($\alpha$).

1D. (iii). $n_2 = r$, $s<r$.

$$\eta(Cuv) = [[(1*\ldots*r-n_1+1)\cdot\eta(t_4)]*b_1*\ldots*b_k]\cdot(1*\ldots*s)\cdot\eta(t_5)$$

$$= [[(1*\ldots*r-n_1+1)\cdot\eta(t_4)]*b_1*\ldots*b_k]\cdot\eta(t_5),$$

since dim $(\eta(t_5)) = s$

$$= [(1*\ldots*r-n_1+1)\cdot\eta(t_4)\cdot\eta(t_5)]*[b_i\cdot\eta(t_5)]*\ldots*[b_k\cdot\eta(t_5)],$$

$$= [(1*\ldots*r-n_1+1)\cdot\eta(t_2)]*[b_1\cdot\eta(t_5)]*\ldots*[b_k\cdot\eta(t_5)];$$

hence Cuv is in sub ($\alpha$).

1D. (iv). $n_2 = r$, $s = r$.

$$\eta(Cuv) = \{[(1*\ldots*r-n+1)\cdot\eta(t_4)]*b_1*\ldots*b_k\}\cdot\{[(1*\ldots*r)\cdot\eta(t_5)]*c_1*\ldots*c_j\},$$

for some $j$, $k\geq 0$, some phrase $b_i$, $1\leq i\leq k$, $c_s$, $1\leq s\leq j$,

$= \{[(1*\ldots*r\text{-}n\text{+}1)\cdot\eta(t_4)]*b_1*\ldots*b_k\}\cdot\eta(t_5)$,
since deg $\eta(u)\leq r$,

$= \{[(1*\ldots*r\text{-}n\text{+}1)\cdot\eta(t_4)\cdot\eta(t_5)]*[b_1\cdot\eta(t_5)]*\ldots*[b_k\cdot\eta(t_5)]$;

hence Cuv is in sub ($\alpha$).

Case 2. $t = K\beta(n_1,k)\gamma(k+1,n_2)$. Note that $\alpha$ can not be in $A_m$. Then $t'' = K\beta\gamma$, and $z'_{i+1} = z_{i+1}$. As a subderivation of (B), we have

$$\alpha(n_1,n_2) \to K\beta(n_1,k)\gamma(k+1,n_2) \Rightarrow Kt_4t_5 = t_2,$$

where $\beta(n_1,k)$ yields $t_4$ and $\gamma(k+1,n_2)$ yields $t_5$, dim $\eta(t_4) = k\text{-}n_1\text{+}1$, dim $\eta(t_5) = n_2\text{-}k$. If $\tau$ is in sub ($t''$), then $\tau = Kab$ for some a in sub ($\beta$), some b in sub ($\gamma$).

2A. $n_2 < r$.

$\eta(Kab) = \eta(a)*\eta(b)$

$= [(1*\ldots*k\text{-}n\text{+}1)\cdot\eta(t_4)]*(1*\ldots*n_2\text{-}k)\cdot\eta(t_5)$

$= (1*\ldots*n_2\text{-}n_1\text{+}1)\cdot(\eta(t_4)*\eta(t_5))$

$= (1*\ldots*n_2\text{-}n_1\text{+}1)\cdot\eta(t_2)$,

hence Kab is in sub ($\alpha$).

2B. $n_2 = r$.

$\eta(Kab) = [(1*\ldots*k\text{-}n_1\text{+}1)\cdot\eta(t_4)]*[(1*\ldots*n_2\text{-}k)\cdot\eta(t_5)]*b_1*\ldots*b_s$,
for some phrases $b_i$, $1\leq i\leq s$, some $s\leq 0$;

$= [(1*\ldots*n_2)\cdot(\eta(t_4)*\eta(t_5))]*b_1*\ldots*b_s$,
since dim $\eta(t_4) = k\text{-}n_1\text{+}1$, dim $\eta(t_5) = n_2\text{-}k$,

$= [(1*\ldots*n_2)\cdot(\eta(t_2))]*b_1*\ldots*b_s$,

so Kab is in sub ($\alpha$).

Case 3. $t = \beta(n_1,r)$. Then $t'' = K\beta\gamma$ for some variable $\gamma$ in V. There is an associated sequence $\Pi_{i+1}$ of productions in G by which $\gamma$ yields some term a of $\mathcal{J}_n$. As a subderivation of (B), we have $\alpha(n_1,r) \xrightarrow[p_{i+1}]{} \beta(n_1,r) \Rightarrow t_2$, where dim $\eta(t_2) = r\text{-}n_1\text{+}1$.

We also have:

$x_i = u\alpha(n_1,r)v$ $\qquad$ $x_{i+1} = u\beta(n_1,r)v$

$z_i = u'\alpha v'$ $\qquad$ $z_{i+1} = u'K\beta\gamma v'$

$z'_{i+1}$ = u'Kβav', where no variables appear in u or u'. If τ is in sub (t"), then τ = Kua for some u in sub (β).

3A. β is in $A_m$: Then r = $n_1$, and if u is in sub (β), η(u) = (1·η($t_2$)), and η(Kua) = [1·η($t_2$)]*η(a).

3B. β is not in $A_m$: Then for u in sub (β), η(u) = [(1*...*$n_1$-r+1)·η($t_2$)]*$b_1$*...*$b_k$, and η(Kua) = [(1*...*$n_1$-r+1)·η($t_2$)]*$b_1$*...*$b_k$*η(a).

In either case, since the production α→Kβγ is in P, α is not in $A_m$, so Kua is in sub (α).

Case 4. t = $w_j$. Then t" = $w_j$, and $z'_{i+1}$ = $z_{i+1}$. As a subderivation of (B), we have α($n_1$,$n_2$)→$w_j$ = $t_2$. Hence $w_j$ is in sub (α), and sub (t") = {$w_j$}.

So we conclude that the construction can not fail for any i+1, i≥0, at step 4.2, hence there can be no failure in the construction at any step. This completes the proof that $\mathcal{L}'$ ⊂ $\mathcal{L}$; along with the earlier result that $\mathcal{L}$ ⊂ $\mathcal{L}'$, we now have the final result: $\mathcal{L}'$ = $\mathcal{L}$.

Lemma 4.7: If G is a reduced grammar with homogeneous variables, and α→Cβγ is in P, then dim β = dim α and deg γ≥deg α.

Proof: Since G is reduced, there are elements $t_1$, $t_2$ in $\mathcal{J}_n$ such that α→Cβγ ⇒ C$t_1t_2$, where β yields $t_1$, and γ yields $t_2$. Then dim α = dim η(C$t_1t_2$) = dim η($t_1$) = dim β, and deg α≥deg η($t_2$) = deg γ, by Lemma 2.5.

Lemma 4.8: If G is a reduced grammar with homogeneous variables, and α→Kβγ appears in P, then dim α = dim β + dim γ, and deg (α) = max {deg β, deg γ}.

Proof: There are $t_1$, $t_2$ in $J_n$ such that α→Kβγ ⇒K$t_1t_2$, where β yields $t_1$ and γ yields $t_2$. Then dim (α) = dim η(K$t_1t_2$) = dim η($t_2$) + dim η($t_1$) = dim β + dim γ, and deg (α) = deg η(K$t_1t_2$) = max {deg η($t_1$), deg η($t_2$)} = max {deg γ, deg β}.

For the remainder of this paper, we will consider restricted linguistic sets in linear morphologies only.

A morphology M will from now on mean a linear, finitely generated, locally finite morphology. The following lemma follows immediately from the definition of a linear morphology.

Lemma 4.9: Let x be a phrase in a linear morphology. Let $M = \{i \mid x \text{ is not free of the } i\text{-th blank}\}$. Then i is in M if and only if the integer i appears in the string x. Next, given a linear morphology pair (M,A), where $A = \{a_1,\ldots,a_n\}$ with associated map $\eta(w_i) = a_i$, $1 \leq i \leq n$, we define a special finite congruence $R_r$ on $\mathcal{J}_n$. Let $r = \max_{1 \leq i \leq n} \{\deg(a_i)\}$.

Partition $\mathcal{J}_n$ as follows:

$D = \{x \text{ in } \mathcal{J}_n \mid x \text{ contains the symbol S}\}$

$A = \{x \text{ in } \mathcal{J}_n \setminus D \mid x \text{ has K-depth greater than r}\}$

$B_1 = \{x \text{ in } \mathcal{J}_n \setminus (D \cup A) \mid \dim \eta(x) = 1\}$

$\vdots$

$B_r = \{x \text{ in } \mathcal{J}_n \setminus (D \cup A) \mid \dim \eta(x) = r\}$.

Clearly $\mathcal{J}_n = A \cup D \cup [\bigcup_{1 \leq j \leq r} B_j]$, and these sets are pairwise disjoint.

Now further partition each set $B_j$ as follows: let $(N_1,\ldots,N_j)$ be a j-tuple of sets $N_k$ of nonnegative integers such that for $1 \leq k \leq j$, either $N_k = \{0\}$ or $N_k \subset \{1,\ldots,r\}$. Let $\mathcal{J}_j$ be the collection of all such j-tuples. Then for each $(N_1,N_2,\ldots,N_j)$ in $\mathcal{J}_j$, let $B_j(N_1,N_2,\ldots,N_j) =$ $\{x \text{ in } B_j \mid \text{for } 1 \leq i \leq j, \text{ if } \deg(\underline{k} \cdot \eta(x)) = 0, \text{ then } N_k = \{0\}$ and if $\deg(\underline{k} \cdot \eta(x)) \neq 0$, then $N_k = \{i \mid \underline{k} \cdot \eta(x) \text{ is not free of the } i\text{-th blank}\}\}$.

It is easily seen that $\bigcup_{(N_1,\ldots,N_j) \in \mathcal{J}_j} B_j(N_1,\ldots,N_j) = B_j$ and that the sets $B_j(N_1,\ldots,N_j)$ are pairwise disjoint. Hence we have a finite partition of $\mathcal{J}_n$ containing the sets D, A, and $B_j(N_1,\ldots,N_j)$ for all $1 \leq j \leq r$, all j-tuples in $\mathcal{J}_j$. Call this collection of sets $R_r$. To whos that $R_r$ is a congruence on $\mathcal{J}_n$, we check the following tables:

(1)

| C | D | A | $B_i(M_1,\ldots,M_i)$ |
|---|---|---|---|
| D | D | D | D |
| A | D | A | A |
| $B_j(N_1,\ldots,N_j)$ | D | A | $B_j(P_1,\ldots,P_j)$, where for $1\leq k\leq j$, $P_k = \bigcup_{s\in N_k} M_{\bar{s}}$, where $\bar{s} \equiv x \pmod{i}$. |

(2)

| S | |
|---|---|
| D | D |
| A | D |
| $B_j(N_1,\ldots,N_j)$ | D |

(3)

| K | D | A | $B_i(M_1,\ldots,M_i)$ |
|---|---|---|---|
| D | D | D | D |
| A | D | A | A |
| $B_j(N_1,\ldots,N_j)$ | D | A | For $i+j>r$: A<br>For $i+j\leq r$:<br>$B_{i+j}(N_1,\ldots,N_j,M_1,\ldots,M_i)$ |

The entry $B_j(P_1,\ldots,P_j)$ in (1) representing the class of Cxy for x in $B_j(N_1,\ldots,N_j)$, y in $B_i(M_1,\ldots,M_k)$ is the only nontrivial calculation. To illumine the argument which follows, here is an example:

$$\eta(x) = (a\underline{1}b\underline{2}c\underline{1})*(b\underline{3})*(\underline{16}bcd)$$

$$\eta(y) = (a\underline{4}a\underline{1})*(cc\underline{2})*a*b$$

Then $x \in B_3(N_1,N_2,N_3)$, where $N_1 = \{1,2\}, N_2 = \{3\}$, $N_3 = \{1,6\}$, and $y \in B_4(M_1,M_2,M_3,M_4)$, where $M_1 = \{1,4\}$, $M_2 = \{2\}$, $M_3 = \{0\}$, $M_4 = \{0\}$.

$\eta(x)\cdot\eta(y) = \eta(Cxy) = (aa\underline{4}a\underline{1}bcc\underline{2}ca\underline{4}a\underline{1})*(ba)*(a\underline{4}a\underline{1}cc\underline{2}cd)$. Hence $Cxy$ is in $B_3(P_1,P_2,P_3)$, where $P_1 = \{1,2,4\}$, $P_2 = \{0\}$, $P_3 = \{1,2,4\}$. Note that $P_1 = M_1 \cup M_2$, $P_2 = M_3$, $P_3 = M_1 \cup M_{\overline{6}} = M_1 \cup M_2$.

Now for the argument. If $x$ is in $B_j(N_1,\ldots,N_j)$, $\eta(x) = x_1*\ldots*x_j$, where for $1\leq k\leq j$, $x_k$ is a string of symbols in which the integers in $N_k$, and no other integers, appear (by Lemma 4.9). Similarly, $\eta(y) = y_1*\ldots*y_i$, where for $1\leq t\leq i$, $y_t$ contains the integers $M_t$, and no others. Now

$$\eta(x)\cdot\eta(y) = [x_1\cdot(y_1*\ldots*y_i)]*[x_2\cdot(y_1*\ldots*y_i)]*\ldots*[x_j\cdot(y_1*\ldots*y_i)]$$

$= z_1*\ldots*z_j$, where $1\leq k\leq j$, $z_k$ is the result of substituting, for each integer $\underline{n}$ in $x_k$, the expression $y_{\overline{n}}$, where $\overline{n} \equiv n(\mathrm{mod}\ i)$. Hence an integer $\underline{m}$ appears in $z_k$ if and only if there is $n \in N_k$ such that $m \in M_{\overline{n}}$. This completes the demonstration that for $t_1$ in $B_j(N_1,\ldots,N_j)$ and $t_2$ in $B_i(M_1,\ldots,M_i)$, $Ct_1t_2$ is in $B_j(P_1,\ldots,P_j)$ as defined Table (1).

We eliminate the other details of showing $R_r$ represents a finite congruence on $\mathcal{J}_n$, since they are trivial.

Theorem 4.10: These are equivalent:

(1) $\Gamma$ is an rl-set in $(M,A)$, for some $A$.

(2) $\Gamma$ is an rg-set of dimension 1, degree 0 in $(M,A)$, for some $A$.

(3) $\Gamma$ is a homogeneous g-set of dimension 1, degree 0 in $(M,B)$, for some $B$.

Proof: (1)$\Rightarrow$(2). If $\Gamma$ is an rl-set in $(M,A)$, then $\Gamma = \mathcal{S}\cap D$ for some rg-set $\mathcal{S}$ in $(M,A)$, where $D$ is the collection of formulas in $M$. Let $E$ be the collection of one-dimensional elements of $M$. Then $R' = \{\eta^{-1}(E), \mathcal{J}_n \setminus \eta^{-1}(E)\}$ is a finite congruence on $\mathcal{J}_n$. Since $\mathcal{S}$ is a g-set, $\mathcal{S} = \eta(B)$ for some recognizable set $B$. Hence $B \cap \eta^{-1}(E)$ is recognizable, and since $\eta(B \cap \eta^{-1}(E)) = \eta(B) \cap E = \mathcal{S}\cap E$, $\mathcal{S}\cap E$ is a g-set in $(M,A)$.

Now, since $D\subset E$, $\Gamma = (\mathcal{S}\cap E)\cap D$, and $\mathcal{S}\cap E$ is a one-

dimensional g-set. Further, $\mathcal{J} \cap E$ is restricted, since $B \cap \eta^{-1}(E)$ contains no strings with operator symbols S if B contains none.

Next we apply Theorem 4.6 to $\mathcal{J} \cap E$, to conclude that $\mathcal{J} \cap E$ has finite K-depth r, for some positive integer r. We let $R_r$ be the special congruence defined above. We let $T = \{C_1, \ldots, C_s\}$ be the congruence associated with the recognizable set L(G') (with K-depth no greater than r) of Theorem 4.6, such that $L(G') = \bigcup_{1 \leq i \leq k} C_i$ and $\eta(L(G')) = \mathcal{J} \cap E$. Now form the congruence $R'' = R_r \wedge T$. By the construction of L(G'), we have $L(G') \subset B_1$. So $L(G') = \bigcup_{1 \leq i \leq k} (B_1 \cap C_1)$ and $L(G') \cap \eta^{-1}(D) = \bigcup_{1 \leq i \leq k} [B_1(\{0\}) \cap C_i]$, which is a recognizable set in $\mathcal{J}_n$. So

$$\begin{aligned} \eta[L(G') \cap \eta^{-1}(D)] &= \eta(L(G')) \cap D \\ &= \mathcal{J} \cap E \cap D \\ &= \mathcal{J} \cap D = \Gamma \end{aligned}$$

is a g-set in (M,A). Clearly the restricted property is not lost, and $\mathcal{J} \cap D$ has dimension one, degree zero, since it is contained in D.

(2)$\Rightarrow$(1). If $\Gamma$ is an rg-set of dimension one, degree zero, then $\Gamma = \Gamma \cap D$, hence $\Gamma$ is an rl-set.

(1)$\Rightarrow$(3). By the discussion in the proof that (1)$\Rightarrow$(2), we see that $\Gamma$ is generated by a recognizable set whose associated congruence is $R_r \wedge T$, and $\Gamma = \eta(L(G'))$, where $L(G') = \bigcup_{1 \leq i \leq k} (B_1\{0\} \cap C_i)$. By the results of Mezei and Wright we know that L(G') can be generated by a grammar G" in best form; in particular, each variable $\alpha \neq \sigma$ in G" has the property that, for some congruence class X in $R_r$ T, $X = \{t \text{ in } \mathcal{J}_n \mid \alpha \text{ yields } t\}$. Now we look at the classes X in $R_r \wedge T$. If $\alpha$ is a variable in G", and $\alpha$ corresponds to a class of the form $B_j(N_1, \ldots, N_j) \cap C_i$, then it is homogeneous. Now suppose $\alpha$ corresponds to a class $D \cap C_i$ or $A \cap C_i$. This cannot happen, since G" is reduced (it is in best form)

and L(G') is restricted, with finite K-depth. Since L(G') has dimension 1, degree 0, the specifications of $\sigma$ must be ({0}). Hence all variables in G" are homogeneous; L(G") = L(G'), and $\Gamma$ = L(G') is a homogeneous g-set of dimension 1, degree 0 in (M,A).

(3)$\Rightarrow$(2). Let $\Gamma = \eta(L(G))$ be such a grammatical set in $(M, V \cup \{1\})$, where $G = (U, W_n, P, \sigma)$. By a slight variant of a well-known result ( Page 34, 7 ), it can be shown that L(G) can be generated by a grammar whose productions are all of the form (i) $\alpha \to C\beta\gamma$,

(ii) $\alpha \to K\beta\gamma$

(iii) $\alpha \to S\beta$

or (iv) $\alpha \to w_j$;

the construction does not destroy the homogeneity of the variables. So we will assume that the productions in G have this form. Now suppose a production of the form $\alpha \to S\beta$ appears in G, where deg $\beta$ = 0. Then deg $\alpha$ = 0. We construct a new grammar G' which differs from G only in that these productions are replaced by productions $\alpha \to \beta$. Then the fact that $\eta(L(G)) = \eta(L(G'))$ follows easily by inductions on the length of derivations in G and G'; the essential fact is that if deg $\beta$ = 0 and $\beta$ yields x, then deg $\eta(x)$ = 0 and $\eta(Sx) = \eta(x)' = \eta(x)$.

A similar argument shows that if $\alpha \to C\beta\gamma$ appears in G', where deg $\beta$ = 0, then we may substitute the production $\alpha \to \beta$; note here that deg $\beta$ = 0 implies deg $\alpha$ = 0.

Without displaying these straightforward proofs, we assume, then, that $\Gamma = \eta(L(G))$, where $G = (U, W_n, P, \sigma)$ has homogeneous variables, and each production in P has the form (i) $\alpha \to w_j$ for some $w_j$ in $W_n$

or (ii) $\alpha \to \beta$, where deg $\alpha$ = deg $\beta$ = 0.

or (iii) $\alpha \to C\beta\gamma$, where deg $\beta \neq 0$.

or (iv) $\alpha \to K\beta\gamma$

or (v) $\alpha \to S\beta$, where deg $\beta \neq 0$.

Let r be the largest degree of a variable in U. We define a new grammar $G' = (U', W_m, P', \sigma^\circ)$ as follows: Let $W_m = \{w_j^i \mid 1 \leq i \leq r,\ 1 \leq j \leq n\}$ be a set of m symbols (where $m = nr$). To each variable $\alpha$ in U, we make correspond a set of symbols $U_\alpha = \{\alpha^i \mid 0 \leq i \leq r\}$. Let $U' = \bigcup_{\alpha \in U} U_\alpha$.

Let P' contain:

(1) if $\alpha \to w_j$ is in P, the productions $\alpha^i \to w_j^i$ for all $\alpha^i$ in $U_\alpha$.

(2) if $\alpha \to \beta$ is in P, the production $\alpha^\circ \to \beta^\circ$.

(3) if $\alpha \to C\beta\gamma$ is in P, the productions $\alpha^i \to C\beta^\circ\gamma^i$, for $0 \leq i \leq r$.

(4) if $\alpha \to K\beta\gamma$ is in P, the productions $\alpha^i \to K\beta^i\gamma^i$, for $0 \leq i \leq r$.

(5) if $\alpha \to S\beta$ is in P, the productions $\alpha^i \to \beta^{i+1}$, for $0 \leq i \leq r$.

Then G' is a restricted grammar. Now let $\eta' : \mathcal{J}_m \to M$ be the (unique) homomorphism such that for $w_j^i$ in $W_m$, $\eta'(w_j) = [\eta(w_j)]^{(i)}$. [We repeat an earlier convention: for x in M, denote x' by $x^{(1)}$ and $x^{(n')}$ by $x^{(n+1)}$; we will agree that $x(^\circ) = x$.] Then let $A = \eta'(W_m)$. Now $\eta'(L(G'))$ is a restricted grammatical set in (M,A). It remains to show that $\eta'(L(G')) = \eta(L(G))$.

Given a leftmost derivation

$$\sigma \xrightarrow[\pi_0]{} x_0 \xrightarrow[\pi_1]{} x_1 \to \cdots \to x_q = x \text{ in } G,$$

we construct a matching G' derivation

$$\sigma^\circ \xrightarrow[p_0]{} y_0 \xrightarrow[p_1]{} y_1 \to \cdots \to y_k = y$$

such that $\eta'(y) = \eta(x)$.

Let $q = 0$. Choose $p_0$ as follows.

(1) If $\pi_0$ is $\sigma \to w_j$, then let $p_0$ be $\sigma^\circ \to w_j^\circ$. If $\pi_0$ is $\sigma \to C\beta\gamma$, let $p_0$ be $\sigma^\circ \to C\beta^\circ\gamma^\circ$; if $\pi_0$ is $\sigma \to K\beta\gamma$, let $p_0$ be $\sigma^\circ \to K\beta^\circ\gamma^\circ$; if $\pi_0$ is $\sigma \to \beta$, let $p_0$ be $\sigma^\circ \to \beta^\circ$; if $\pi_0$ is $\sigma \to S\beta$, let $p_0$ be $\sigma^\circ \to \beta^1$.

(2) If $x_s$ differs from $y_2$ only in that (a) $y_s$ contains no symbols S, and (b) variables in $y_s$ carry superscripts, then continue; otherwise the construction has failed.

(3) For each variable $\beta^i$ appearing in $y_s$, find the matching variable $\beta$ in $x_s$. For some t in $\mathcal{J}_n$, $\beta$ yields t. For $\beta^i$, substitute $\underbrace{SS\ldots S}_{i}t$. For each terminal $w_j^i$ in $x_s$, substitute $\underbrace{SS\ldots S}_{i}w_j$. When all substitutions have been made, call the resulting string sub $(y_s)$. If $\eta(\text{sub } y_s) = \eta(x)$, continue. Otherwise the construction has failed.

(4) If s = k, the construction is complete. Otherwise, add 1 to s, and continue.

(5) Choose $p_s$. If $x_{s-1} = u\beta v$, for strings u and v, and $\pi_s$ is $\beta \to t$, we find the matching variable $\beta^i$ in $y_{s-1}$, and choose $p_s$ to be applied to $\beta^i$, depending on the form of t.

Case 1. $t = \gamma$. Then deg $\beta = 0$. If $i = 0$, let $p_s$ be $\beta^\circ \to \gamma^\circ$; otherwise the construction has failed.

Case 2. $t = C\gamma\delta$. Let $p_s$ be $\beta^i \to C\gamma^\circ\delta^i$.

Case 3. $t = K\gamma\delta$. Let $p_s$ be $\beta^i \to K\gamma^i\delta^i$.

Case 4. $t = w_j$. Let $p_s$ be $\beta \to w_j$.

Case 5. $t = S\gamma$. Let $p_s$ be $\beta^i \to \gamma^j{}^{i+1}$ if this production is in P'; otherwise the construction has failed.

Return to step 2.

Now if this construction is always successful, we have, for each x in L(G), a y in L(G') such that $\eta(x) = \eta'(y)$. For $\eta(\text{sub } y) = \eta'(y)$, since for all i,j, $\eta'(w_j^i) = \eta(\underbrace{SS\ldots S}_{i}w_j)$. Hence we will conclude that $\eta(L(G)) \subset \eta'(L(G'))$. We show by contradiction that the construction can always be successfully carried out. Assume the construction fails for some x in L(G). Let d be the least integer such that there is an x in L(G) for which the procedure fails at some

step for s = d.

An inspection of step 1 shows that for d = 0, the construction always works. So d must be greater than zero.

Suppose there is a failure at step 2. An examination of all possible choices of $p_d$ shows this is not possible, by the minimality of d.

Suppose the construction fails at step 3. At step s-1, we had $x_{s-1} = u\beta v$; $\pi_s$ is $\beta \to t$ for some string t, and $p_s$ is $\beta^i \to t'$ for some string t'. Since by the minimality of x, $\eta(\text{sub } y_{s-1}) = \eta(x)$, in showing that $\eta(\text{sub } y_s) = \eta(x)$ it will suffice to show that $\eta(\text{sub } \beta^i) = \eta(\text{sub } t')$. We consider cases depending on the form of t.

(1) $t = \gamma$. Then i = 0, $t' = \gamma^\circ$, and sub $\gamma^\circ$ = sub $\beta^\circ$.

(2) $t = C\gamma\delta$; then $t' = C\gamma^\circ\delta^i$; sub $\beta^i = \underbrace{SS\ldots S}_{i}Cz_1z_2$,

where $\gamma$ yields $z_1$ and $\delta$ yields $z_2$; sub $t' = Cz_1\underbrace{SS\ldots S}_{i}z_2$.

Then $(\text{sub } \beta^i) = [\eta(z_1)\cdot\eta(z_2)]^{(i)}$

$= \eta(z_1)\cdot\eta(z_2)^{(i)}$

$= \eta(t')$.

(3) $t = K\gamma\delta$; then $t' = K\gamma^i\delta^i$; sub $\beta^i = \underbrace{SS\ldots S}_{i}Kz_1z_2$,

where $\gamma$ yields $z_1$ and $\delta$ yields $z_2$; sub $t' = K\underbrace{SS\ldots S}_{i}t_1\underbrace{SS\ldots S}_{i}t_2$.

Then $\eta(\text{sub } \beta^i) = [\eta(z_1)*\eta(z_2)]^{(i)}$

$= \eta(z_1)^{(i)}*\eta(z_2)^{(i)}$

$= \eta(\text{sub } t')$.

(4) $t = w_j$; then $t' = w_j^i$. sub $\beta^i = \underbrace{SS\ldots S}_{i}w_j$ = sub t',

hence $\eta(\text{sub } \beta^i) = \eta(t')$.

(5) $t = S\gamma$; then $t' = \gamma^{i+1}$; sub $\beta^i = \underbrace{SS\ldots S}_{i+1}z$, where $\gamma$ yields z, and sub $(t') = \underbrace{SS\ldots S}_{i+1}z$.

Hence no failure can occur at step 3.

Then the construction must fail at step 5; that is, there must be some production called for which does not appear in P'.

Case 1. $t = \gamma$. Then deg $\beta = 0$, hence $i = 0$ and the desired production is in P'.

Cases 2, 3, 4, 5. If $i<r$, then all needed productions appear in P'. We will show that $\beta^r$ can never appear in the construction.

We will need a definition. We say that $\alpha^i$ produces $\beta^{i+k}$ if there is a derivation $\alpha \Rightarrow u$ such that $\beta^{i+k}$ is a symbol in u, and the derivation is formed under the following restrictions:

(1) if a production $\alpha^i \rightarrow C\gamma^\circ\delta^i$ appears, then we apply no further productions to $\gamma^\circ$.

(2) if a production $\alpha^i \rightarrow K\delta^i\xi^i$ appears, we choose either $\delta^i$ or $\xi^i$ for the continuation of the derivation, applying no further productions to the other.

The resulting string, then, will yield $\beta^{i+k}$ from $\alpha^i$ in a "direct" way, without additional productions which are irrelevant to the appearance of $\beta^{i+k}$.

It is clear that if $\beta^r$ appears in a derivation, there is some $\alpha^\circ$ which produces it. We will show that, if, for any i, $\alpha^i$ produces $\beta^{i+k}$, then the least non-zero integer appearing in the specifications of $\alpha$ is greater than k. Assuming this result for the moment, we then argue as follows. Suppose $\beta^r$ appears in a derivation. For some $\alpha$ in V, $\alpha^\circ$ produces $\beta^r$; hence the least non-zero integer in the specifications of $\alpha$ is at least r+1; if deg $\alpha \neq 0$, then deg $\alpha$ is greater than r, a contradiction, since we assumed r to be the maximum degree of variables in G.

Now, if deg $\alpha = 0$, we claim that there is some $\gamma^\circ$ which produces $\beta^r$ such that deg $\gamma^\circ \neq 0$. The only productions applicable to $\alpha^\circ$, if deg $\alpha = 0$, are of the form (1) $\alpha^\circ \rightarrow \beta^\circ$, where deg $\beta = 0$ or (2) $\alpha^\circ \rightarrow C\beta^\circ\delta^\circ$, where deg $\beta \neq 0$.

The application of a production of type 1 yields again a variable of degree zero with zero superscript. Hence we must at some point in the derivation apply a production of type 2, where deg ($\delta$)>0, in order to obtain the special type of derivation which produces $\beta^r$ from $\alpha^\circ$. But in that case, we have $\delta^\circ$ produces $\beta^r$, and $\delta^\circ$ has positive degree. So again we have arrived at a contradiction, and $\beta^r$ can not appear.

We conclude that there is no failure at step 5, so the construction is always possible, and $\eta(L(G)) \subset \eta'(L(G'))$.

It remains to show the earlier claim that, if $\alpha^i$ produces $\beta^{i+k}$, then the least non-zero integer appearing in the specifications of $\alpha$ is greater than k. Suppose the assertion is not true. Let s be the least integer such that, for some i, some k, some $\alpha$, some $\beta$, $\alpha^i$ produces $\beta^{i+k}$ by a special derivation of length s such that the assertion fails. Let us examine such a derivation, and consider several cases, depending on the form of the first production applied in the derivation. Clearly s is greater than zero.

Case 1. $\pi$ is $\alpha^i \rightarrow \gamma^i$. Then $\gamma^i$ produces $\beta^{i+k}$, contradicting the minimality of s.

Case 2. $\pi$ is $\alpha^i \rightarrow C\delta^\circ\gamma^i$; then again $\gamma^i$ produces $\beta^{i+k}$, a contradiction of the minimality.

Case 3. $\pi$ is $\alpha^i \rightarrow K\delta^i\gamma^i$; then either $\delta^i$ or $\gamma^i$ produces $\beta^{i+k}$, by a subderivation of length less than s, again a contradiction.

Case 4. $\pi$ is $\alpha^i \rightarrow \gamma^{i+1}$. Since $\gamma^{i+1}$ yields $\beta^{i+k}$ by a special derivation of length less than s, the least positive integer in the n-tuple of specifications of $\gamma$ is greater than k-1. But note that since $\pi$ is in P', the production $\alpha \rightarrow S\gamma$ is in P' further, deg $\gamma \neq 0$ and deg $\alpha \neq 0$. If $(N_1,\ldots,N_m)$ and $(M_1,\ldots,M_m)$ are the specifications of $\alpha$ and $\gamma$ respectively (notice they must both be m-tuples for some m, since for all x in a morphology L, dim x = dim x'), then

for all sets $N_j \neq \{0\}$, $N_j = \{m+1 \mid m \text{ is in } M_j\}$. Hence the least positive integer appearing in $(N_1,\ldots,N_m)$ is greater than k, as required.

This completes the proof of the claim, and hence the proof that $(L(G)) \subset \eta'(L(G'))$.

Next we show the reverse inclusion. Let $\alpha \xrightarrow[\pi]{i} z_1 \longrightarrow z_2 \longrightarrow \ldots \longrightarrow z_s = z$ be a leftmost derivation in G'. We will show by induction on s that $\alpha$ yields an x in L(G) such that $\eta(x)^{(i)} = \eta'(z)$.

Suppose s = 1. Then $\pi$ is $\alpha^{i}\rightarrow w_j^{i}$. By the construction, the production $\alpha\rightarrow w_j$ appears in P; and $\eta'(w_j^{i}) = [\eta(w_j)]^{(i)}$. So the assertion holds for s = 1.

Suppose s>1, and the assertion holds for k<s. We distinguish several cases, depending on the form of $\pi$.

Case 1. $\pi$ is $\alpha^{\circ}\rightarrow\beta^{\circ}$. Then by the induction hypothesis, $\beta$ yields x in L(G) such that $\eta(x) = \eta(z)$. Since $\alpha\rightarrow\beta$ is in P, by the construction (note that deg $\alpha$ = deg $\beta$ = 0), we have the desired result.

Case 2. $\pi$ is $\alpha^{i}\rightarrow C\beta^{\circ}\gamma^{i}$; then $\alpha\rightarrow C\beta\gamma$ is in P. Now $z = Cy_1y_2$, where $\beta^{\circ}$ and $\gamma^{i}$ yield $y_1$ and $y_2$ by subderivations of length less than s. Hence $\beta$ yields $x_1$ and $\gamma$ yields $x_2$ such that $\eta'(y_1) = \eta(x_1)$ and $\eta'(y_2) = \eta(x_2)^{(i)}$. Hence $\alpha$ yields $Cx_1x_2$, where

$$\begin{aligned}\eta(Cx_1x_2)^{(i)} &= [\eta(x_1)\cdot\eta(x_2)]^{(i)}\\ &= \eta(x_1)\cdot\eta(x_2)^{(i)}\\ &= \eta'(y_1)\cdot\eta'(y_2)\\ &= \eta'(Cy_1y_2)\\ &= \eta'(z), \text{ as required.}\end{aligned}$$

Case 3. $\pi$ is $\alpha^{i}\rightarrow K\beta^{i}\gamma^{i}$; then $\alpha\rightarrow K\beta\gamma$ is in P, and $z = Ky_1y_2$; by the induction hypothesis, $\beta$ yields $x_1$ and $\gamma$ yields $x_2$ such that $\eta(x_1)^{(i)} = \eta(y_1)$ and $\eta(x_2)^{(i)} = \eta(y_2)$. Hence

$$\begin{aligned}\eta(Kx_1x_2)^{(i)} &= [\eta(x_1)*\eta(x_2)]^{(i)}\\ &= \eta(x_1)^{(i)}*\eta(x_2)^{(i)}\end{aligned}$$

$$= \eta'(y_1)*\eta'(y_2)$$

$$= \eta'(Ky_1y_2)$$

$$= \eta'(z).$$

Case 4. $\pi$ is $\alpha^{i}\rightarrow\beta^{i+1}$. Then $\alpha\rightarrow S\beta$ is in P. By the induction hypothesis, $\beta$ yields x such that $\eta(x)^{(i+1)} = \eta'(z)$. Hence $\alpha$ yields Sx and

$$\eta'(Sx)^{(i)} = [\eta(x)']^{(i)}$$

$$= [\eta(x)]^{(i+1)}$$

$$= \eta'(z).$$

Hence the assertion holds for all s.

Applying this assertion to $\sigma^{\circ}$, we have $\eta'(L(G')) \subset \eta(L(G))$, which completes the proof.

F-regular restricted linguistic sets. We will look at a particularly well-behaved class of sets, the rl-sets in $(M, V \cup\{1\})$ which are F-regular, where F is the collection of V-factorizations of M in $\mathcal{J}_n$ defined in Chapter 3. We let $V = \{v_1,\ldots,v_{n-1}\}$ be a fixed ordering of V and $\eta(w_i) = v_i$, $1\leq i\leq n-1$, $\eta(w_n) = 1$, as usual. We obtain a simple form for productions in the grammars generating such sets.

Theorem 4.11: Every F-regular rl-set can be generated by a grammar whose productions are of the form

(i) $\sigma\rightarrow\beta$

(ii) $\alpha\rightarrow w_j$

or (iii) $\alpha\rightarrow Cw_j\underbrace{KK\ldots K}_{r-1}\alpha_1\alpha_2\ldots\alpha_r$,

for some $w_j$ in $W_n$, some variables $\alpha,\alpha_1,\ldots,\alpha_r$, some $r\geq 1$, where r is the degree of $\eta(w_j)$.

Proof: Let $G = (U,W_n,P,\sigma)$ be a grammar in best form generating such a set $\Gamma$ in $(M, V\cup\{1\})$. Then we define a new grammar $G' = (U,W_n,P',\sigma)$. Let P' be the collection of (1) productions $\sigma\rightarrow\beta$, where $\sigma\rightarrow\beta$ is in G, and (2) for all $\alpha \neq \sigma$, for all strings t such that $\alpha$ yields t and t is of the form (ii) or (iii), the production $\alpha\rightarrow t$. We note that P' is a

finite set, since V is finite, and the degree of elements in $(V \cup \{1\})$ is bounded.

It is clear that $L(G') \subset L(G)$. Now we want to show that $L(G) \subset L(G')$. First we will show by induction on the length of a derivation that, for each $\alpha$ in U, if $\alpha$ yields a term in F by a derivation in G, it yields the same term by a derivation in G'.

Suppose this is not true. Let m be the least integer for which there is some variable $\alpha$ and some term x in F for which the hypothesis does not hold, with a leftmost derivation in G of length m,

$$\alpha = x_0 \xrightarrow[\pi_1]{} x_1 \xrightarrow[\pi_2]{} \cdots \xrightarrow[\pi_m]{} x_m = x.$$

Suppose $m = 1$. Then $\pi_1$ must be $\alpha \to w_j$ for some terminal $w_j$; but $\alpha \to w_j$ is in P', so m is not 1. Suppose m is greater than 1. Since x is in F, $\pi_1$ must be of the form $\alpha \to C\beta\gamma$, and $\pi_2$ must have the form $\beta \to w_j$, and $x = Cw_jt$ for some string t.

Case 1. $\deg(\eta(w)) = 1$ and t is in F. In this case, $\gamma$ yields t by a derivation in G of length less than m, so by the minimality of m, $\gamma$ yields t in G'. We note that $\alpha \to Cw_j\gamma$ is in P', so $\alpha$ yields x in G'.

Case 2. t is not in F. Then, since $x = Cw_jt$ is in F, t has the form $\underbrace{KK\ldots K}_{r-1}t_1t_2\ldots t_r$ for some terms $t_i$ in F, and some $r>1$, where $r = \deg \eta(w_j)$.

Since G is in best form, and the derivation is leftmost, $\pi_3, \pi_4, \ldots, \pi_{r+1}$ must have the form $\xi \to K\delta\mu$ for some variables $\xi$, $\delta$, $\mu$, and $x_{r+1} = Cw_j\underbrace{KK\ldots K}_{r-1}\alpha_1\alpha_2\ldots\alpha_r$ for some variables $\alpha_i$.

By the construction, the production $\alpha \to x_{r+1}$ is in P'. Further, each $\alpha_i$ must yield $t_i$ (which is in F) by a subderivation of length less than m; hence $\alpha_i$ yields $t_i$ in G', by the minimality of m. Hence $\alpha$ yields x in G', a contradiction.

So, for each variable $\alpha$, if $\alpha$ yields a term in F by a derivation in G, it yeilds the term by a derivation in G'. But a term x is in L(G) precisely when there is a derivation L(G),

$$\sigma \to \alpha \Rightarrow x, \text{ and } x \text{ is in } F.$$

Now $\sigma \to \alpha$ is in P' whenever it is in P. Since x is in F and $\alpha$ yields x in G, then $\alpha$ yields x in G'. Hence $\sigma$ yields x in G' and x is in L(G'). So $L(G) \subset L(G')$, and we may conclude that $L(G) = L(G')$.

Theorem 4.12: In a free morphology M, with vocabulary V, if $\Gamma$ is a g-set in $(M, V \cup \{1\})$ generated by a grammar G with productions of the form specified in Theorem 4. then $\Gamma$ is an F-regular rg-set.

Proof: From the form of the productions it is clear that $L(G) \subset F$, and $\Gamma$ is restricted. Since M is free, V is monotectonic, hence for each phrase x in M, $\eta^{-1}(x) \cap F$ consists of precisely one element. Therefore, $\eta^{-1}(L(G)) \cap F = L(G)$, which is recognizable; also, since $L(G) \subset F$, $\eta(L(G)) \subset \eta(F)$. So $\Gamma$ is F-regular.

Lemma 4.13: If D is the collection of formulas in M with (initialized) vocabulary V, then D is an F-regular restricted linguistic set in $(M, V \cup \{1\})$.

Proof: Let $V = V_1 \cup V_2$, where $V_1$ consists of the elements of degree zero in V, and $V_2$ contains those of positive degree. We construct a grammar $G = (U, W_\eta, P, \sigma)$ such that $L(G) = \eta^{-1}(D) \cap F$. Then $\eta(L(G)) = D \cap \eta(F) = D$ since $\eta(F)$ contains all phrases, and

$$\begin{aligned} \eta^{-1}\eta(L(G)) \cap F &= \eta^{-1}[D \cap \eta(F)] \cap F \\ &= \eta^{-1}(D) \cap \eta^{-1}\eta(F) \cap F \\ &= \eta^{-1}(D) \cap F \\ &= L(G), \end{aligned}$$

hence $\eta(L(G)) = D$ is an F-regular g-set; it is also an 1-set since $D \cap D = D$. We will see that D is restricted from the form of the productions in G. We now specify G. Let $U = \{\sigma, \alpha\}$. Let P contain:

(1) $\sigma \to Cw_j \underbrace{KK\ldots K}_{r-1}\underbrace{\sigma\sigma\ldots\sigma}_{r}$,

where $r = \deg \eta(w_j)$, if $r>0$ , and $1 \leq j \leq n-1$.

(2) $\sigma \to w_k$ if $\deg \eta(w_k) = 0$.

By the form of the productions, $L(G) \subset F$ is clear. Now we show that, for t in F, $\deg \eta(t) = 0$ if and only if t is in L(G). First we show by induction on a leftmost derivation in G, that for t in L(G), $\deg \eta(t) = 0$. Let the derivation be

(*) $$\sigma \xrightarrow[\pi]{} x_1 \longrightarrow x_2 \longrightarrow \ldots \longrightarrow x_m = t.$$

Suppose $m = 1$. Then $\pi$ is $\alpha \to w_k$, and $\deg(\eta(w_k)) = 0$ by the construction of G. Suppose, for $m>1$, the hypothesis holds for all $k<m$. Then $\pi$ is $\sigma \to Cw_j \underbrace{KK\ldots K}_{r-1}\underbrace{\sigma\sigma\ldots\sigma}_{r}$. Then $t = Cw_jKK\ldots Kt_1t_2\ldots t_r$, where $\alpha$ yields $t_i$ by a subderivation of (*) of length less than m; hence by the induction hypothesis, $\deg \eta(t_i) = 0$ for all i. Therefore $\eta(t) = \eta(w_j)\cdot(\eta(t_1)*\ldots*\eta(t_r))$ has degree zero, since by Lemma 2.5,

$$\begin{aligned}\deg \eta(t) &\leq \deg(\eta(t_1)*\ldots*\eta(t_r)) \\ &= \max\{\deg \eta(t_i) \mid 1 \leq i \leq r\} \\ &= 0.\end{aligned}$$

This completes the first half of the proof.

Next we show, by induction on the depth of t (defined below) that if t is in F and $\deg(\eta(t)) = 0$, then t is in L(G). The depth of a term t in F is:

(1) if $t \in W_n$, depth (t) = 1

(2) if $t = \underbrace{SS\ldots S}_{r}w_n$ for some $r>0$, depth (t) = 1

(3) if $t = Cw_j\underbrace{KK\ldots K}_{r-1}t_1t_2$ for some $r>0$,

$$\text{depth } (t) = \max\{\text{depth } (t_i) \mid 1 \leq i \leq r\} + 1.$$

If depth (t) = 1, and $\deg \eta(t) = 0$, then $t = w_j$ for some $w_j$ such that $\eta(w_j) = 0$. An inspection of P shows that $w_j$ is in L(G) for such $w_j$.

Suppose for m>1, the hypothesis holds for all t with depth less than m. Then if L has depth m,

$$t = Cw_j \underbrace{KK\ldots K}_{r-1} t_1 t_2 \ldots t_r \text{ for some } r \geq 0,$$

where for each $t_i$, depth $(t_i)$ is less than m.

$\eta(t) = \eta(w_j) \cdot (\eta(t_1) * \ldots * \eta(t_r))$. Since V is initialized and $\eta(w_j)$ is in V and deg $\eta(w_j) = r$, we may conclude by Lemma 2.10 that deg $(\eta(t))$ = max {deg $\eta(t_i) \mid 1 \leq i \leq r$}. Therefore if deg $(\eta(t)) = 0$, we have deg $\eta(t_i) = 0$ for all i, $1 \leq i \leq r$. Then by the induction hypothesis, we have $\sigma$ yields $t_i$ for $1 \leq i \leq r$. Since the production

$$\sigma \to Cw_j \underbrace{KK\ldots K}_{r-1} \underbrace{\sigma\sigma\ldots\sigma}_{r}$$

is in P, we have the derivation

$$\sigma \to Cw_j \underbrace{KK\ldots K}_{r-1} \underbrace{\sigma\sigma\ldots\sigma}_{r} \quad Cw_j \underbrace{KK\ldots K}_{r-1} t_1 t_2 \ldots t_r$$

in P, as required. This completes the proof.

Hence $L(G) = \eta^{-1}(D) \setminus F$, and the earlier discussion completes the proof of the theorem.

Theorem 4.14: If $\Gamma_1$ and $\Gamma_2$ are F-regular rl-sets in $(M, V \cup \{1\})$, so are $\Gamma_1 \cup \Gamma_2$, $\Gamma_1 \cap \Gamma_2$, and $D \setminus \Gamma_1$, where D is the collection of formulas in M.

Proof: By Theorem 4.10, $\Gamma_1$ and $\Gamma_2$ are rg-sets. By Theorem 3.20, $\Gamma_1 \cup \Gamma_2$ is an F-regular g-set. The restricted property is preserved, since $\Gamma_1 = \eta(C)$ and $\Gamma_2 = \eta(D)$ for some recognizable sets C and D which do not contain strings with the symbol S, hence neither does the recognizable set $C \cup D$, and $\Gamma_1 \cup \Gamma_2 = (C \cup D)$. So $\Gamma_1 \cup \Gamma_2$ is an F-regular rg-set, hence an F-regular rl-set.

By Theorems 4.10 and 3.19, $\Gamma_1 \cap \Gamma_2$ is an F-regular g-set; and $\Gamma_1 \cap \Gamma_2$ has dimension 1, degree 0, so it is an 1-set. Again the restricted property is preserved; for

$$\Gamma_1 \cap \Gamma_2 = [\eta^{-1}(\Gamma_1 \cap \Gamma_2) \cap F],$$

and $\Gamma_1 \cap \Gamma_2$ has degree zero.

Now we show that if a term t in F contains the symbol S, then $\eta(t)$ has positive degree; we use induction on the depth m of a term t in F, defined as in the proof of Lemma 4.14. Suppose the depth of t is 1, and t contains S. Then $t = \underbrace{S\ldots S}_{k} w_n$, for some $k \geq 0$, and $\eta(t)$ is the blank k+1, which has positive degree. Hence the assertion holds for m = 1. Suppose the hypothesis holds for all terms of depth less than m. If t has depth m, $t = Cw_j K\ldots K t_1 t_2 \ldots t_r$ for $r \geq 0$, $t_i$ in F of depth less than m, for $1 \leq i \leq m$. If t contains S, then some $t_j$ must contain S; hence by the induction hypothesis $\eta(t_j)$ has positive degree. But $\eta(t) = \eta(w_j) \cdot (\eta(t_i) * \ldots * \eta(t_r))$, and since V is initialized and $\eta(w_j)$ is in V and has degree r, we conclude by Lemma 2.10 that

$\deg_\eta(t) = \max\{\deg(\eta(t_i)) \mid 1 \leq i \leq r\}$, which is positive. This concludes the proof of the assertion.

So if there is a term t in $\eta^{-1}(\Gamma_1 \cap \Gamma_2) \cap F$ containing the symbol S, then $\eta(t)$ has positive degree. This is a contradiction, since $\eta(t)$ is in $\Gamma_1 \cap \Gamma_2$, which has degree zero. Hence $\eta^{-1}(\Gamma_1 \cap \Gamma_2) \cap F$ is restricted, and therefore so is $\Gamma_1 \cap \Gamma_2$.

Next, by Theorem 3.22, $\eta(F) \setminus \Gamma_1$ is an F-regular g-set. By Lemma 4.13, D is an F-regular rg-set. Since F-regular g-sets are closed under intersection, $[\eta(F) \setminus \Gamma_1] \cap D = D \setminus \Gamma_1$ is an F-regular g-set. It is also an l-set, since $D \setminus \Gamma_1 \subset D$, which has dimension 1, degree 0. Now we need only show that $D \setminus \Gamma_1$ is restricted. To do this, we refer to the proofs of Lemma 4.13 and Theorem 3.19 and Theorem 3.22, and note that: $D \setminus \Gamma_1 = (\eta(F) \setminus \Gamma_1) \cap D = \eta(Y)$, where $Y = [\eta^{-1}(D) \cap A \cap F]$ is recognizable, and $\eta(F) \setminus \Gamma_1 = \eta(A)$. It remains only to show that Y is restricted. But $\eta^{-1}(D) \cap F$ is restricted, by the proof of Lemma 4.13, and clearly any subset of a restricted set in $\Phi_n$ is restricted. So Y is restricted, and $D \setminus \Gamma_1$ is an F-regular restricted linguistic set, as required.

Theorem 4.15: Every context-free language is the homomorphic image of an F-regular restricted linguistic set in a free morphology.

Proof: Let $H = (U,\Sigma,P,\sigma)$ be a context-free grammar (in the traditional sense) generating the context-free language $L(H)$. We may assume H is in Greibach normal form [11]; that is, all productions are of the form

(*) $$\alpha \to m\alpha_1\alpha_2\cdots\alpha_n,$$

for some variables $\alpha,\alpha_1,\ldots,\alpha_n$, for some $n\geq 0$, and for some terminal m. Number the productions in P as $p_1, p_2,\ldots,p_r$. Let $A = \{z_1,z_2,\ldots,z_r\}$ be a collection of distinct symbols. We will define a submorphology M' of the total linear morphology over A. It will be that submorphology generated by the set V', which contains, for each $p_i$ in P, the expression $(z_i\underline{1}\underline{2}\ldots\underline{n})$, if $p_i$ has the form (*). Now we define a reocgnizable set L(G) on $\mathcal{J}_r$, where $\eta:\mathcal{J}_r\to M$ is the homomorphism which maps $w_i$ to $z_i\underline{1}\underline{2}\ldots\underline{n}$ in V. Let $G = (U,W_r,P',\sigma)$, where P' contains r productions $q_i$, $1\leq i\leq r$, each derived from $p_i$ as follows:

if $p_i$ has the form $\alpha\to m\alpha_1\alpha_2\cdots\alpha_n$,
then $q_i$ is $\alpha\to Cw_iKK\ldots K\alpha_1\alpha_2\cdots\alpha_n$.

The form of the productions in G satisfies the hypothesis of Theorem 4.12, hence $\eta(L(G))$ is F-regular. Now $\eta(L(G))$ is a g-set in (M',V'), which is Lukasiewicz and hence free. Note that $\eta(L(G))$ is restricted. Now let M be the submorphology of the total linear morphology over $\Sigma$ generated by the set A which we now define by: $m\underline{1}\underline{2}\ldots\underline{n}$ is in A if and only if, for some variables $\alpha_1, \alpha_2,\ldots,\alpha_n$ in U, for some $n\geq 0$, for some $p_i$ in P, the right-hand side of $p_i$ is $m\alpha_1\alpha_2\cdots\alpha_n$.

We can define a homomorphism $\psi:M'\to M$ by specifying its values on V', since V' is a vocabulary for M' and M' is free. Let $\psi$ be determined by: $\psi(z_i\underline{1}\ldots\underline{n}) = m\underline{1}\ldots\underline{n}$, where $m\alpha_1\ldots\alpha_n$ is the right-hand side of $p_i$.

Now we claim that $\psi\eta(L(G))$ is the context-free language

$L(H)$. To see that $L(H) \subset \psi\eta(L(G))$, we show by induction on the length $k$ of a leftmost derivation in H that for any variable $\alpha$ in U, if $\alpha$ yields $x$ in $L(H)$ by a derivation in H, then $\alpha$ yields an element $y$ in $L(G)$ such that $\psi\eta(y) = x$. Let the derivation be

$$(**) \qquad \alpha \underset{\pi}{\xrightarrow{H}} x_1 \xrightarrow{H} \ldots \xrightarrow{H} x_k = x,$$

where $\pi$ denotes the first production applied. Suppose $k = 1$. Then $\pi$ is $\alpha \to m$, for some $m$ in $\Sigma$. This case is easy; if is $p_j$, then the production $\alpha \to w_j$ appears in P'; $\eta(w_j) = z_j$ and $\psi\eta(w_j) = \psi(z_j) = m$. Therefore the hypothesis holds for $k = 1$. Assume the hypothesis holds for $s<k$. Suppose $\pi$ is $p_j$, which is $\alpha \to m\alpha_1\alpha_2\ldots\alpha_n$. Then $x = mz_1z_2\ldots z_n$, where for $1 \leq i \leq n$, $\alpha_i$ yields $z_i$ by a subderivation of (**). Since these subderivations have length less than $k$, by the induction hypothesis each $\alpha_i$ yields $y_i$ by a derivation in G such that $\psi\eta(y_i) = z_i$. The production $\alpha \to Cw_jKK\ldots K\alpha_1\alpha_2\ldots\alpha_n$ is in P' by the construction; hence we have a G-derivation

$$\alpha \to Cw_j\underbrace{KK\ldots K}_{n-1}\alpha_1\alpha_2\ldots\alpha_n \Rightarrow Cw_j\underbrace{KK\ldots K}_{n-1}y_1y_2\ldots y_n.$$

We also have

$$\begin{aligned}\psi\eta[Cw_jKK\ldots Ky_1\ldots y_n] &= \psi\eta(w_j)\cdot(\psi\eta(y_1)*\ldots*\psi\eta(y_n)) \\ &= (m\underline{1}\ldots\underline{n})\cdot(z_1*\ldots*z_n), \text{ where the } z_i \text{ are phrases,} \\ &= mz_1z_2\ldots z_n, \text{ as required.}\end{aligned}$$

So $L(H) \subset \psi\eta(L(G))$. To show that $\psi\eta(L(G)) \subset L(H)$, we show by induction on the length of a leftmost derivation that for any variable $\alpha$, if $\alpha$ yields $y$ by a derivation in G, then $\alpha$ yields $\psi\eta(y)$ by a derivation in H. Let the derivation be

$$(***) \qquad \alpha \underset{\pi}{\xrightarrow{G}} y_0 \xrightarrow{G} \ldots \xrightarrow{G} y_m = y.$$

Suppose $k = 1$. Then $\pi$ is $\alpha \to w_j$ for some $w_j$ in $W_n$. By the construction, there is a production $\alpha \to m$ in P such that $\psi\eta(w_j) = m$.

Suppose the hypothesis holds for derivations of length less than k, and suppose $\pi$ is the production $\alpha \to Cw_j \underbrace{KK\ldots K}_{n-1}\alpha_1\ldots\alpha_n$.

Then $y = Cw_jKK\ldots Kt_1t_2\ldots t_n$, where $\alpha_i$ yields $t_i$ in G for $1 \leq i \leq n$ by a subderivation of (***), of length less than m. By the induction hypothesis, for each i, $\alpha$ yields $\psi\eta(t_i)$ by a derivation in H. By the construction, the production $p_j$ in H is $\alpha \to m\alpha_1\ldots\alpha_n$, where $\psi\eta(w_j) = (m\underline{1}\underline{2}\ldots\underline{n})$. So we have in H, $\alpha \to m\alpha_1\ldots\alpha_n \Rightarrow m[\psi\eta(t_1)]\ldots[\psi\eta(t_n)]$. But this is precisely $\psi\eta(y)$, for

$$\begin{aligned}\psi\eta(y) &= \psi\eta[Cw_j\underbrace{KK\ldots K}_{n-1}t_1t_2\ldots t_n] \\ &= \psi\eta(w_j)\cdot(\psi\eta(t_1)*\ldots*\psi\eta(t_n)) \\ &= (m\underline{1}\underline{2}\ldots\underline{n})\cdot(\psi\eta(t_1)*\ldots*\psi\eta(t_n)) \\ &= m[\psi\eta(t_1)]\ldots[\psi\eta(t_n)].\end{aligned}$$

So $L(H) \supset \psi\eta(L(G))$. We complete the proof by noting that since L(H) has dimension 1, degree 0, so does $\psi\eta(L(G))$; further, $\psi$ preserves degree, hence $\eta(L(G))$ is a linguistic set.

We remark that not all homomorphic images of F-regular rl-sets in free morphologies are context-free languages. We will show, without going into the finer details, how to construct as the homomorphic image of an rl-set in a free morphology, the set $C = \{xx \mid x \in L(H)\}$ for any context-free language L(H). It is well-known that this set is not context-free for arbitrary context-free languages.

So let L(H) be a context-free language. By Theorem 4.15, it is the homomorphic image of an F-regular rl-set $\Gamma$ in (M, V $\cup\{1\}$) where M is free. We add to V the element $(s\underline{1})$, where s is some symbol distinct from those in V, and let L be the (free) morphology generated by V $\cup\{s\underline{1}\}$, which is a vocabulary for L. $\Gamma$ is easily shown to be an F-regular rl-set in (L, V $\cup\{s\underline{1}\}\cup\{\underline{1}\}$). Now $\Gamma = \eta(L(G))$ for some $G = (U, W_n, P, \sigma)$. We define a new grammar G' =

$(U \cup \{\sigma'\}, W_n, P', \sigma')$, where $P' = P \cup \{\sigma' \to Cw_j\sigma\}$, and $w_j$ is such that $\eta(w_j) = s\underline{1}$. Then $L(G')$ consists of strings $Cw_jt$, where $\eta(t)$ is in $L(H)$ under the homomorphism h of Theorem 4.15. Extend h so that $h(s\underline{1}) = (\underline{11})$. Then $\eta(L(G'))$ is the collection of strings $(s\underline{1})\cdot\eta(t)$, and $h\eta(L(G'))$ the collection

$$(\underline{11})\cdot h\eta(t) = (\underline{11})\cdot x$$
$$= (xx),$$

for x in $L(H)$.

Theorem 4.16: Every F-regular rl-set in a free morphology is a context-free language.

Proof: Let $\Gamma$ be such a set in $(M, V \cup \{1\})$, where M is a submorphology of the total linear morphology over S, $\Gamma = \eta(L(G))$, $G = (U, W_n, P, \sigma)$. Then by Theorem 4.12, we may assume that the productions in G are of the form

(*) $$\alpha \to Cw_j \underbrace{KK\ldots K}_{r-1}\alpha_1\alpha_2\cdots\alpha_r,$$

where $\deg(\eta(w_j)) = r$.

Define a context-free grammar $H = (U, S, P', \sigma)$, where P' contains: for each production of the form (*) in P, the production

$$\alpha \to m\alpha_1\alpha_2\cdots\alpha_r,$$

where $\eta(w_j) = m\underline{12}\ldots\underline{r}$. Then we claim that $L(H) = \eta(L(G))$.

Let $\alpha$ be any variable in U. We show by induction on the length of a leftmost derivation that if $\alpha$ yields a string of terminals x, by a derivation in H, then $\alpha$ yields by a derivation in G a term t in $\mathcal{J}_n$ such that $\eta(t) = x$. Let $\alpha \xrightarrow[\pi]{} x_1 \longrightarrow x_2 \longrightarrow \ldots \longrightarrow x_s = x$ be a leftmost derivation in H. Suppose $s = 1$. Then $\pi$ is $\alpha \to m = x$ for some m in S such that $\alpha \to w_j$ is in P and $\eta(w_j) = m$. Hence the claim is true for $m = 1$. Suppose the hypothesis holds for $k<s$. Then $x_1 = m\alpha_1\ldots\alpha_r$, and $x = mz_1z_2\ldots z_r$, where for $1 \le i \le r$, $\alpha_i$ yields $z_i$ by a subderivation of length less than s. Hence by the induction hypothesis, for each $\alpha_i$, there is a term $t_i$ in $\mathcal{J}_n$

such that $\alpha_i$ yields $t_i$ by a G-derivation and $\eta(t_i) = z_i$. Since $\pi$ is $\alpha \to m\alpha_1\alpha_2\ldots\alpha_r$, by the construction the production $\alpha \to Cw_jKK\ldots K\alpha_1\ldots\alpha_r$ is in P, where $\eta(w_j) = m\underline{1}\underline{2}\ldots\underline{r}$. Hence we have the G-derivation

$$\alpha \to Cw_jKK\ldots K\alpha_1\alpha_2\ldots\alpha_r \Rightarrow Cw_jKK\ldots Kt_1t_2\ldots t_r = t$$

and $\eta(t) = (m\underline{1}\underline{2}\ldots\underline{r})\cdot(z_1 * z_2 * \ldots * z_r)$

$$= mz_1z_2\ldots z_r = x.$$

Hence, in particular, the hypothesis holds for the variable $\sigma$, so $L(H) \subset \eta(L(G))$.

Now we show by induction on the length of a leftmost derivation in G that, for any variable $\alpha$, if $\alpha$ yields t in $\mathcal{J}_n$, then $\alpha$ yields $\eta(t)$ by a derivation in H. Let the G-derivation be $\alpha \xrightarrow[\pi]{} t_1 \longrightarrow t_2 \longrightarrow \ldots \longrightarrow t_s = t$. Suppose $s = 1$. Then $\pi$ is $\alpha \to w_j$ for some $w_j$ in $W_n$. Further, since $\Gamma$ is an rl-set, and $\eta(w_j)$ is in $\Gamma$, $\eta(w_j)$ has degree zero. Since M is free, $\eta(w_j) = m$ for some symbol m. By the construction, $\alpha \to m$ is in P'. So the claim is true for $s = 1$. Suppose $s > 1$, and the hypothesis holds for $k < s$. Then $\pi$ is of the form $\alpha \to Cw_jKK\ldots K\alpha_1\alpha_2\ldots\alpha_r$, $t = Cw_jKK\ldots Kt_1t_2\ldots t_r$, and for $1 \leq i \leq r$, $\alpha_i$ yields $t_i$ by a subderivation of length less than s. By the construction, the production $\alpha \to m\alpha_1\ldots\alpha_r$ is in P, where $\eta(w_j) = (m\underline{1}\underline{2}\ldots\underline{r})$. Hence by the induction hypothesis we have the H-derivation

$$\alpha \to m\alpha_1\ldots\alpha_r \Rightarrow mz_1z_2\ldots z_r,$$

where $z_i = \eta(t_i)$, $1 \leq i \leq r$. Now

$$\begin{aligned}\eta(t) &= (m\underline{1}\underline{2}\ldots\underline{r})\cdot(\eta(t_1)*\ldots*\eta(t_r)) \\ &= (m\underline{1}\underline{2}\ldots\underline{r})\cdot(z_1*z_2*\ldots*z_r) \\ &= mz_1z_2\ldots z_n,\end{aligned}$$

so the claim holds for all s.

Applying this result to the variable $\sigma$, we have $\eta(L(G)) \subset L(H)$. Hence $\eta(L(G)) = L(H)$, and is a context-free language.

Theorem 4.17: All context-free languages are structurally unambiguous rg-sets.

Proof: We refer to the proof of Theorem 4.15. Let L(H) be a context-free language. The recognizable set L(G) of that proof, where $L(H) = \psi\eta[L(G)]$, is contained in the set of A-factorizations of M in $\mathcal{J}_n$. Note that A is a vocabulary for M. Hence, by Corollary 3.29, $\psi\eta[L(G)]$ is structurally unambiguous.

Theorem 4.18: Every restricted linguistic set is the homomorphic image of a restricted grammatical set in a free morphology.

Proof: Let $\Gamma$ be an rl-set in (M,A). Let M' be the free morphology associated with M, and let $\theta:M'\to M$ be the (onto) homomorphism of Corollary 2.17. Suppose $A = \{a_1,a_2,\ldots,a_n\}$ and $\Gamma = \eta(C)$ for a recognizable set C in $\mathcal{J}_n$, where $\eta(w_i) = a_i$, $1\leq i\leq n$. For each $a_i$ in A, let $a_i'$ be any element of the set $\eta^{-1}(a_i)$ in M'. Let $\eta':\mathcal{J}_n\to M'$ be the homomorphism determined by $\eta'(w_i) = a_i'$, $1\leq i\leq n$. Then $\eta'(C)$ is an rg-set in (M',A'), and by the construction,

$$\theta[\eta'(C)] = \eta(C) = \Gamma.$$

Substratum Properties. The formulas in linear morphologies are finite strings of (juxtaposed) symbols from some finite alphabet S, as are the words in context-free languages. We ignore the morphology structure, for the moment, and consider the formulas as elements in the free semigroup with unity (under juxtaposition) generated by S, which we denote by S*. $\Lambda$ represents the empty string in the semigroup; note that it is not an element of a linear morphology. This view allows us to examine properties usually associated with the languages whose underlying algebraic system is such a semigroup. In the case of linguistic sets, we will call such properties substratum properties.

Let S* and T* be semigroups over S, T respectively, as above. Let $h:S^*\to T^*$ be a (semigroup) homomorphism. Then if $\Gamma$ is a linguistic set in M, a submorphology of the total linear morphology over S, $\Gamma$ is contained in S*; further, if h is non-erasing, that is, if for all s in S, $h(s) \neq \Lambda$,

then $h(\Gamma) \subset M'$, the total linear morphology over T. In this case we call h a substratum homomorphism of the l-set $\Gamma$.

Theorem 4.19: The restricted linguistic sets in linear morphologies are closed under non-erasing substratum homomorphism.

Proof: Let $h:S^* \to T^*$ be such a homomorphism, and let $\Gamma$ be an rl-set in (M,A), where M is a submorphology of the total linear morphology over S. Let M' be the total linear morphology over T. Then construct the set B from A as follows: If a is in A, replace each occurrence of a symbol s in S with the string h(s) from T*. Note that the non-erasing restriction guarantees that h(s) is not the empty string, hence the element of B we construct is in M'. Suppose $\Gamma = \eta(C)$ for some recognizable set C in $\mathcal{J}_n$. Let $\eta':\mathcal{J}_n \to M'$ be determined by: if $\eta(w_k) = a_i$, then $\eta'(w_i)$ is that element of B produced by the above construction. It follows easily that $\eta'(C) = h(\Gamma)$.

Let $w = a_1 a_2 \ldots a_m$ be a phrase in a submorphology of the total linear morphology M' over S; where each $a_i$ is in $S \cup N$. Then the substratum reversal of w, written $w^R$, is the formula $a_m a_{m-1} \ldots a_2 a_1$. We extend this notion to all of M' be defining: $(x*y)^R = x^R * y^R$. The substratum reversal of an l-set is the collection of reversals of its elements, i.e.

$$\Gamma^R = \{w^R \mid w \in \Gamma\}.$$

Lemma 4.20: In a linear morphology, for elements x,y,

$$(1) \quad (x \cdot y)^R = x^R \cdot y^R$$

$$(2) \quad (x')^R = (x^R)'.$$

Proof: It suffices to prove the theorem when x is a phrase, since $(x*y)^R = x^R * y^R$. Suppose M' is a submorphology of the total linear morphology over S, $x = a_1 a_2 \ldots a_m$ is a phrase in M', where $a_i \in S \cup N$, $1 \leq i \leq m$, and $y = z_1 * z_2 * \ldots * z_s$ for phrases $z_k$ in M', $1 \leq k \leq s$. Then $x\,y = \hat{a}_1 \hat{a}_2 \ldots \hat{a}_m$, where for $1 \leq i \leq m$,

$$\hat{a}_i = \begin{cases} \hat{a}_i \text{ if } a_i \in S \\ z_{\bar{k}}, \text{ where } \bar{k} \equiv k \pmod{s} \text{ if } a_i = k \text{ for some } k \text{ in } N. \end{cases}$$

Then $(x \cdot y)^R = \hat{a}_m{}^R \hat{a}_{m-1}{}^R \ldots \hat{a}_2{}^R \hat{a}_1{}^R$. Now $x^R = a_m a_{m-1} \ldots a_2 a_1$ and $y^R = z_1{}^R * z_2{}^R * \ldots * z_s{}^R$; $x^R \cdot y^R = b_m b_{m-1} \ldots b_2 b_1$, where

$$b_i = \begin{cases} a_i \text{ if } a_i \in S \\ z_{\bar{k}}, \text{ where } \bar{k} \equiv k \pmod{s} \text{ if } a_i = k \text{ for some } k \text{ in } N. \end{cases}$$

In each case, $a_i{}^R = b_i$, so $(x \cdot y)^R = x^R \cdot y^R$.

Now we look at $(x')^R$. As before, $x = a_1 a_2 \ldots a_m$. Then $x' = b_1 b_2 \ldots b_m$, where

$$b_i = \begin{cases} a_i, \text{ if } a_i \text{ is in } S. \\ k+1, \text{ if } a_i = k \text{ for some } k \text{ in } N. \end{cases}$$

We also have $x'^R = b_m b_{m-1} \ldots b_2 b_1$.

$x^R = a_m a_{m-1} \ldots a_2 a_1$,

$x^{R'} = c_m c_{m-1} \ldots c_2 c_1$, where

$$c_i = \begin{cases} a_i, \text{ if } a_i \text{ is in } S \\ k+1, \text{ if } a_i = k \text{ for some } k \text{ in } N. \end{cases}$$

Hence $(x')^R = (x^R)'$.

Theorem 4.21: Linguistic sets in linear morphologies are closed under substratum reversal.

Proof: Let $\Gamma$ be an 1-set in $(M,A)$, where $M$ is a submorphology of $M'$, the total linear morphology over $S$.

We construct a set $B$ from $A$. If $a_i$ is in $A$, then $a_i = s_1 s_2 \ldots s_m$ for symbols $s_i$ in $S \cup N$, $1 \leq i \leq m$. Let $b_i = s_m s_{m-1} \ldots s_2 s_1$. Then let $B$ be the collection of elements $b_i$ so formed from elements in $A$. $B$ is a collection of phrases in $M'$. Suppose $\Gamma = \eta(C)$ for some recognizable set $C$ in $\mathcal{J}_n$. Define $\eta' : \mathcal{J}_n \to M'$ by: $\eta'(w_i) = b_i$. Then we claim that $\Gamma^R = \eta'(C)$.

It suffices to show that for all $t$ in $\mathcal{J}_n$, $\eta(t)^R = \eta'(t)$; this we do by induction on the operator depth $j$ of $t$. Suppose $j = 1$; then $t = w_i$ for some $w_i$ in $W_n$, and $\eta(t) = a_i$; then

$\eta'(t) = b_i = a_i^R$ by the construction. Hence the assertion holds for $j = 1$. Suppose $j>1$ and the hypothesis holds for $s<j$. We consider three cases, depending on the form of $t$.

Case 1. $t = Ct_1t_2$ for some $t_1$, $t_2$ in $\mathcal{J}_n$ with operator depth less than $j$. Now

$$\eta(t)^R = [\eta(t_1)\cdot\eta(t_2)]^R;$$

$$= \eta(t_1)^R\cdot\eta(t_2)^R \text{ by Lem. 4.20;}$$

$$= \eta'(t_1)\cdot\eta'(t_2), \text{ by the induction hypothesis}$$

$$= \eta'(Ct_1t_2), \text{ as required.}$$

Case 2. $t = Kt_1t_2$ for some $t_1$, $t_2$ in $\mathcal{J}_n$ with operator depth less than $j$. Then

$$\eta(t)^R = (\eta(t_1)*\eta(t_2))^R$$

$$= \eta(t_1)^R*\eta(t_2)^R, \text{ by definition;}$$

$$= \eta'(t_1)*\eta'(t_2), \text{ by the induction hypothesis,}$$

$$= \eta'(Kt_1t_2), \text{ as required.}$$

Case 3. $t = St_1$ for some $t_1$ in $\mathcal{J}_n$ with operator depth less than $j$. Then

$$\eta(t)^R = [\eta(t_1)]'^R$$

$$= [\eta(t_1)^R]' \text{ by Lemma 4.20;}$$

$$= [\eta'(t_1)]', \text{ by the induction hypothesis,}$$

$$= \eta'(St), \text{ as required.}$$

Hence for all $t$ in $\mathcal{J}_n$, $\eta(t)^R = \eta'(t)$. Now if $x$ is in $\Gamma$, $x = \eta(t)$ for some $t$ in $C$; $\eta'(t) = \eta(t)^R$ is in $\eta'(C)$. If $y$ is in $\eta'(C)$, then $y = \eta'(t) = \eta(t)^R$ for some $t$ in $C$. So $\Gamma^R = \eta'(C)$, and is an 1-set in $(M',B)$.

Let $x = a_1a_2\ldots a_n$ and $y = b_1b_2\ldots b_s$ be formulas in $M'$, the total linear morphology over $S$, where $a_i$, $b_j \in S$, $1\leq i\leq n$, $1\leq j\leq s$. Then the substratum product of $x$ and $y$, denoted $xy$, is the formula $z = z_1a_2\ldots a_nb_1b_2\ldots b_s$. If $X$ and $Y$ are two subsets of $M'$, we define the substratum product of $X$ and $Y$ to be $XY = \{xy \mid x \in X, y \in Y\}$.

Theorem 4.22: Restricted linguistic sets in linear morphologies are closed under substratum product.

Proof: Let $\Gamma_1$ be an rl-set in (M,A), where $\Gamma_1 = \eta(C)$ for some recognizable set in $\mathcal{J}_n$; let $\Gamma_2$ be an rl-set in (L,B), where $\Gamma_2 = \eta'(D)$ for some recognizable set D in $\mathcal{J}_m$. Suppose M and L are submorphologies of the total linear morphologies over S and S' respectively. Let P be the total linear morphology over S ∪ S'. Now we will generate $\Gamma_1\Gamma_2$ as an rl-set in $(P, A \cup B \cup \{(\underline{12})\})$. Fix an ordering for $A \cup B \cup \{(\underline{12})\} = \{d_1,\ldots,d_s\}$. Let $\eta'': \mathcal{J}_s \to P$ be the homomorphism determined by: $\eta''(w_i) = d_i$, $1 \leq i \leq s$. Suppose $C = L(G)$ and $D = L(H)$ for grammars in best form $G = (U, W_n, P, \sigma)$, $H = (U', W_m, P', \sigma')$. Assume U and U' are disjoint. Let $J = (U \cup U', W_s, P'', \sigma'')$ where P'' contains:

(1) $\sigma'' \to Cw_j K \sigma\sigma'$, for that $w_j$ such that $\eta''(w_j) = (\underline{12})$

(2) All productions in P and P' except those of the form $\alpha \to w_i$ for $w_i$ in $W_n$ or $W_m$.

(3) For each production $\alpha \to w_i$ in P, the production $\alpha \to w_k$, where $\eta''(w_k) = \eta(w_i)$.

(4) For each production $\alpha \to w_i$ in P', the production $\alpha \to w_k$, where $\eta''(w_k) = \eta'(w_i)$.

Then $\eta''(L(J))$ yields precisely those strings of the form $(\underline{12}) \cdot (x * y) = xy$, where x is in $\Gamma_1$ and y is in $\Gamma_2$.

Theorem 4.23: If $\Gamma$ is an rl-set in (M,A), then so is $\Gamma^+$.

Proof: Let $\Gamma = \Gamma_1 = \Gamma_2$ in the proof of Theorem 4.22. To the grammar J generating the product $\Gamma\Gamma$, add the productions $\sigma \to \sigma''$ and $\sigma' \to \sigma''$, to form the grammar J'. It is tedious but completely straightforward to show that $\eta''(L(J''))$ is precisely $\Gamma^+$.

Erasure Operators. In linguistic applications, although we want to reject sentences with unfilled blanks, it will be convenient, on occasion, to have a method for removing "extra" blanks, if the sentence is otherwise grammatically correct. For example, the sentence

The ___ duchess carried a ___ parasol.

is well-formed, and does not require for syntactical correctness the addition of modifiers in the blanks.

We now introduce an element $\varepsilon$, called an erasure operator, whose function is to eliminate unwanted blanks; that is, (The ___ duchess carried a ___ parasol)$\cdot\varepsilon$ = The duchess carried a parasol. We will call a morphology with such an element a morphology with erasure operator.

Formally, we introduce $\varepsilon$ into the total linear morphology $M = (M,\cdot,*,',\tau)$ over the set S. Let M' be the collection of all n-tuples, each of whose slots contains either a finite non-empty sequence of symbols in $S \cup N$, or the symbol $\varepsilon$. Then $M' = (M',\cdot,*,',(\underline{1}))$, the total linear morphology over S with erasure operator $\varepsilon$, is defined as follows.

Denote the n-tuple $x = (x_1,x_2,\ldots,x_n)$ by $x_1*x_2*\ldots*x_n$. For x,y in M', where $x = x_1*\ldots*x_n$ and $y = y_1*\ldots*y_s$,

(1) $x*y$ is the n+s-tuple $x_1*\ldots*x_n*y_1*\ldots*y_s$.

(2) $x\cdot y$ is the n-tuple $z_1*\ldots*z_n$, where $z_i$ is defined by:

(1) if $x_i = \varepsilon$, then $z_i = \varepsilon$.

(2) if $x_i \neq \varepsilon$, then $z_i$ is the result of (a) substituting for each blank k in $x_i$ the expression $y_{\bar{k}}$, where $\bar{k} \equiv k \pmod{s}$, if $y_{\bar{k}} \neq \varepsilon$: and (b) erasing the blank k in $x_i$ if $y_{\bar{k}} = \varepsilon$.

(3) $x'$ is the n-tuple $z_1*\ldots*z_n$, where

$$z_i = \begin{cases} \varepsilon, \text{ if } x_i = \varepsilon \\ \text{the result of substituting, for each blank k in } x_i, \text{ the blank k+1, otherwise.} \end{cases}$$

Thereby M' becomes a half-ring morphology, with M as a submorphology. Now let L, with vocabulary V, be any submorphology of M. Then the submorphology of M generated by $V \cup \{\varepsilon\}$ contains L. So we have

Theorem 4.24: Every linear morphology L can be extended

to a linear morphology with erasure operator $\varepsilon$.

Now let us consider $\varepsilon$ to be the empty sequence of symbols $\Lambda$. Then we may consider the matter of arbitrary substratum homomorphism.

Theorem 4.25: The collection of rl-sets in linear morphologies with erasure operators is closed under arbitrary substratum homomorphism.

Proof: We refer to the proof of Theorem 4.19. Given the situation in that proof, we may now construct the set B from A as follows: if $a_i = s_1 s_2 \ldots s_m$ for symbols $s_j$ in $S \cup N$, then

(1) if $a_i = \varepsilon$, then $b_i = \varepsilon$.

(2) if for some $s_i$, $s_i$ is in N, then $b_i$ is the result of (a) substituting, for each $s_j$ in S, the string $h(s_j)$, if $h(s_j) \neq \Lambda$; and (b) erasing $s_j$ if $h(s_j) = \Lambda$.

(3) if for all $s_j$, $s_j \in S$, then

(i) if $h(s_j) = \Lambda$ for all $s_j$ in $a_i$, $b_i = \varepsilon$.

(ii) if for some $s_j$, $h(s_j) \neq \Lambda$, then $b_i$ is defined as in rule 2.

With this change in the construction of B, the construction is identical with that of Theorem 4.19.

BIBLIOGRAPHY


1. Aho, Alfred V. Indexed grammars--an extension of context-free grammars, IEEE Conference Record of the 1967 Eighth Annual Symposium on Switching and Automata Theory, pp. 21-31.

2. Chomsky, N. On certain formal properties of grammars, Information and Control, Vol. 2, 1959, pp. 137-167.

3. Chomsky N. Three models for the description of a language, IRE Trans. Inform. Theory, Vol. IT2, 1956, pp. 113-124.

4. Davis, A. S. An axiomatization of the algebra of transformations over a set, Math. Annalen 164, 1966, pp. 372-77.

5. Davis, A. S. Half-ring moprhologies, Proceedings of the 1967 Leeds Summer Seminar on Logic (to appear).

6. Floyd, R. W. On the non-existence of a phrase structure grammar for ALGOL-60, Commun. Assoc. Computing Machinery, Vol. 5, 1962, pp. 483-484..

7. Ginsburg, S. "The Mathematical Theory of Context-Free Languages," McGraw Hill, New York, 1966.

8. Ginsburg, S., and Greibach, S. A. Abstract families of languages, System Development Corp. Rept. TM-738/031/00, April 17, 1967.

9. Ginsburg, S., Greibach, S. A., and Harrison, Michael A. Stack automata and compiling, Journal Assoc. Computing Machinery, Vol. 14, No. 1, pp. 172-201.

10. Ginsburg, S. and Rice, H. G. Two families of languages related to ALGOL, J. Assoc. Computing Machinery, Vol. 10, 1963, pp. 29-47.

11. Greibach, S. A. A new normal-form theorem for context-free phrase structure grammars, J. Assoc. Computing Machinery, Vol. 12, 1965, pp. 42-52.

12. Mezei, J. and Wright, J. B. Algebraic automata and context-free sets, Inf. and Control, Vo. 11, 1967, pp. 3-29.

13. Naur, P. (ed.) Report on the algorithmic language ALGOL-60, Common Assoc. Computing Machinery, Vol. 3, 1960, pp. 299-314.

14. Rosenkrantz, Daniel J. Programmed grammars, a new device for generating formal languages, IEEE Conference Record of the 1967 Eighth Annual Symposium on Switching and Automata Theory, pp. 14-20.