Cahiers de linguistique théorique et appliquée

TIRAGE À PART

10

1973

Fasc. 1

ÉDITIONS DE L'ACADÉMIE DE LA RÉPUBLIQUE SOCIALISTE DE ROUMANIE
1973

THE ALGEBRA OF DISTRIBUTIONALLY DEFINED CLASSES *

BY

P. H. ROOSEN-RUNGE

1. CLASSES AS CLOSED SETS

The problem of constructing distributionally defined grammatical classes can be loosely stated as follows:

> Given a language L of strings over some alphabet A and some finite set M of sentence segments, such that each sentence in L can be 'spelled' as a unique sequence $m_1m_2...m_k$ of elements of M, select from all possible subsets of M those which incorporate in some useful way information about how elements of M occur in sentences of L.

The criterion to be employed in selecting such subsets is clearly the nub of the problem; a typically vague formulation has been to say that elements of M are to be assigned to the same class if they "pattern alike".

It hardly needs saying that classification problems of this sort are not peculiar to linguistics. They arise whenever we have a set of objects which are known to have certain properties, where the objects are to be clustered together in terms of the properties they have in common. For this reason, we begin by considering the structure of the object-property relationship in quite abstract terms.[1]

(1) *Definition*: Let the finite set of objects to be clustered or classified be denoted by M and let E be a set of properties such that:

> each object m has at least one of the properties in E, and each $e \in E$ is a property of some object in M.

* This research was partially supported by the National Research Council of Canada, Grant A5551.

[1] The notion of class which results is obviously "weak" since it is applicable to any objects and properties meeting certain minimal restrictions. This weakness is itself not without significance, however, for it makes clear the degree of theoretical elaboration required before one can, in any essential way, exploit such elementary linguistic features of the problem as, for example, that the elements of M are substrings of the elements of L.

Certain subsets of M are then distinguished by the fact that they can be specified in terms of a list of those properties in E which all elements of the subset have in common. To put it another way, such subsets are the extensions in M of conjunctions of properties in E, and are therefore of the form:

$$\{m \in M : m \text{ has } e_1, e_2, \ldots,\}$$

for some set of properties $e_1, e_2, \ldots \in E$. The following definitions lead to an algebraic characterization of the structure of these sets:

Definitions:

(2) Let $P(A)$ denote the lattice [2] of all subsets of A partially ordered by set inclusion. Define

$$\eta : P(M) \to P(E) \text{ and } \mu : P(E) \to P(M) \text{ by}$$

$$\eta(X) = \{e \in E : m \text{ has } e \text{ for all } m \in X\},$$

$$\mu(Y) : \{m \in M : m \text{ has } e \text{ for all } e \in Y\}.$$

(3) $X \in P(M)$ is *closed* if $\mu\eta(X) = X$. ($\mu\eta(X)$ is termed the *closure* of X and we will usually abbreviate it by $\bar{X}$.) $Y \in P(E)$ is *closed* if $\eta\mu(Y) = Y$.

(4) Let *M* and *E* denote the collections of closed subsets of M and E respectively.

In the terminology of Birkhoff, the mappings μ and η constitute a Galois connection between $P(M)$ and $P(E)$, and $\mu\eta$ and $\eta\mu$ are dual closure operators on $P(M)$ and $P(E)$. The following basic results will be extensively used in the remainder of the paper (for proofs, see Birkhoff, [1967] or Szász, [1963]):

Proposition 1: (A) For all $X \subseteq M$, $Y \subseteq E$,

(a) $X \subseteq X'$ implies $\eta(X') \subseteq \eta(X)$, and $Y \subseteq Y'$ implies $\mu(Y') \subseteq \mu(Y)$;

(b) $\eta(X)$ and $\mu(Y)$ are closed;

(c) $X \subseteq \mu\eta(X)$ and $Y \subseteq \eta\mu(Y)$,

$X \subseteq X'$ implies $\mu\eta(X) \subseteq \mu\eta(X')$, and

$Y \subseteq Y'$ implies $\eta\mu(Y) \subseteq \eta\mu(Y')$.

[2] A *lattice* is a partially ordered set in which any two elements have a greatest lower bound (*meet*) and least upper bound (*join*). As an example, in the set $P(A)$, whose elements are subsets of A, the join of any two elements is their union, the meet is their intersection. Since these unions and intersections are again subsets of A, $P(A)$ is a lattice.

(B) M and E are complete lattices [3] under the partial ordering $\subseteq$, and for any subset X of M or E, the meet operation (Π) is that of set intersection, i.e.

$$\Pi X = \bigcap X.$$

The join operation (Σ) is not, in general, that of set-union; however, $\bigcup X \subseteq \Sigma X$.

(C) Each of the lattices M and E are dually isomorphic to each other under the mappings μ and η; hence, for $Y \subseteq E$,

$$\Sigma \mathrm{Y} = \eta(\Pi\{\mu(\mathrm{Y}) : \mathrm{Y} \in \mathrm{Y}\})$$

$$\Pi \mathrm{Y} = \eta(\Sigma\{\mu(\mathrm{Y}) : \mathrm{Y} \in \mathrm{Y}\}),$$

and dual results for $X \subseteq M$.

In order to derive some non-trivial structure for the concept of 'class', some initial restrictions must be imposed; it would not lead us very far to allow any subset of M to be a potential class. In what follows, we shall restrict the range of choices for classes to the closed subsets of M which are determined by the properties in E. Henceforth, then, *class* will be a synonym for closed set. This restriction has the following intuitive interpretation: we can think of E together with the connective "and" as the vocabulary of a very primitive language for expressing what a set of objects in M have in common. To describe a set of objects in this language means to specify a set of properties which hold for all and only the objects in that set; and it follows immediately from (3) that the closed subsets of M are exactly those so describable.

We observe that one effect of restricting the concept of 'class' to closed sets is to make the distinction between objects and their properties purely nominal, at least with respect to the structure of the classes which they generate. By Prop. 1(C), the lattice M is just the lattice E turned upside-down; hence, all relationships among classes of objects are mirrored by corresponding relationships among classes of properties.

The concept of closure was introduced as a basis for defining grammatical classes by Sestier [1960] and has reappeared, apparently independently, in several other models of syntactic description. (For a review and citations, see Marcus [1967, 1969].) As a result there has been a kind of theoretical 'inflation', particularly as regards the elementary results in Prop. 1, which have been frequently restated and reproved under various disguises. It is perhaps worth-while, teherefore, to recall the venerability of the underlying mathematical framework upon which this approach depends; the algebraic characterization of closure operators goes back to E. H. Moore (*Introduction to a form of general analysis*, New Haven: 1910), and the construction of dual lattices of closed sets from a binary relation appeared in the first edition of Birkhoff's *Lattice Theory* in 1940.

[3] A lattice L is *complete* if, for any A $\subseteq$ L, the join (ΣA) and meet (ΠA) exist in L. Any finite lattice is complete.

The most extensive application of the concept of closure to the formal description of syntactic distribution appears in an important series of papers by J. Kunze (1967, 1968, 1970, 1971). Of particular interest in these papers is Kunze's development of a specific distributional taxonomy of word forms and the detailed structure which he proposes for systems of "form-classes". But as our main concern here is with what can be said about *M* and its lattice-structure in advance of any commitment to specific taxonomic or classificatory criteria, we shall not, in what follows, make direct use of the details of Kunze's theory, though at the level of definitions there are a number of points of contact and overlap.

2. AN INTERPRETATION FOR M AND E

As indicated in Sec. 1, the intended interpretation of M will be a set of sentence segments which for the purposes of the distributional analysis are treated as minimal (i.e. not subject to further segmentation.) Hence there is no need to distinguish between M and the alphabet A. For the sake of definiteness, we will follow the example of Harris [1946] and Wells [1947] and take M = set of morphemes of L. But it should be emphasized that subsequent formal results do not depend in any essential way on this choice, since they do not involve any intrinsic properties of morphemes (e.g. their "meaningfulness"); we could equally well have taken M to be the set of morpheme alternates, or the set of words of L or any other set of segments whose distribution is deemed linguistically interesting.

The intended interpretation of E is the set of environments of elements of M, where by *environment* is meant a pair (u, v) of strings (possibly null) over M. Thus the properties of elements of M will be the environments they occur in. There are several plausible ways of defining the relation "occurs in":

(5) Let E_T = set of total environments
$= \{(x, y) : xmy \in L \text{ for some } m \in M\}$;

(6) Let E_p = set of partial environments
$= \{(x, y) : \exists\, u, v \text{ such that } uxmyv \in L \text{ for some } m \in M\}$.

There is an obvious sense of 'occurs in' corresponding to each of these choices for E:

(5.1) m occurs in $(x, y) \in E_T$ if $xmy \in L$.

(6.1) m occurs in $(x, y) \in E_p$ if $\exists\, u, v$ such that $uxmyv \in L$.

The concept of environment in Harris and Wells, to the extent that it is explicitly defined, appears to be that of definition (6) but from their examples, it seems that a narrower interpretation of 'occurs in' than (6.1) was intended. For in these examples, classes are characterized by partial environments which are so chosen that their expansion into total environments does not alter the class of possible substitutions. Thus in order to characterize a class of adjectives one does not need to give a total environment such as

> "The ____ man, having given away all his money, finds peace at last.",

since the shorter environment "The ____ man" will yield the same class. By contrast, under (6.1) both "and" and "is" occur in the environment

> "black ____ beautiful"

even though in a great many complete sentences containing the fragment "black and beautiful", "is" cannot be substituted for "and". This suggests yet another definition of E and 'occurs in':

(7) E_R = set of restricted partial environments
$= \{(x, y) : uxmyv \in L$ for some $m \in M$ and if $u'xm'yv' \in L$ then $u'xmyv' \in L\}$.

Thus a partial environment (x, y) is in E_R only if every extension of the environment which actually occurs in L has the same substitution potential. Though this seems to capture the sense in which 'environment' was intended by Harris and Wells, it turns out to be inequivalent to E_T, for (7) excludes certain total environments as being distributionally relevant. Thus, in English, the word "go" has a part of its distribution the null environment (,), but this environment is not in E_R, since it can be extended to another total environment, say, (he,) in which "go" cannot occur. This is an obviously undesirable byproduct of the above definition, since the fact that some words can serve as complete sentences is the sort of information we would wish to incorporate in our choice of E.

One plausible solution is to treat sentences as 'unextendable' by adopting a convention that strings in L are bounded by special marks, say '#', so that every sentence is of the form $\# x \#$ and hence cannot be a substring of another sentence. With this convention, $E_T \subseteq E_R$ and this, in turn, implies $M_T \subseteq M_R$. The converse holds in any event:

Lemma 1: Let M_T, and M_R be the classes of M defined by E_T and E_R. Then $M_R \subseteq M_T$.

Proof: Let μ_R, η_R, and μ_T, η_T be the Galois connections associated with E_T and E_R, and 'occurs$_R$' and 'occurs$_T$' be the corresponding senses of 'occurs'. Consider $e_R = (x, y) \in \eta_R(X)$ where $X \in M_R$. Let e_T be any total environment of the form $(ux, yv) \in E_T$. By definition, some m_0 occurs$_T$ in e_T. Hence, by (7), for any m which occurs$_R$ in e_R, m occurs$_T$ in e_T. Thus, for all $m \in \mu_R(e_R)$, $m \in \mu_T(e_T)$. Now, if we set $Y = \{(ux, yv) \in E_T : (x, y) \in \eta_R(X)\}$, by the above argument

$$e_R \in \eta_R(X) \text{ and } e_T \in Y \text{ and } m \in \mu_R(e_R) \Rightarrow m \in \mu_T(e_T) \quad \text{or equivalent}$$

$$\text{(a)} \quad m \in \mu_R\eta_R(X) \Rightarrow m \in \mu_T(Y).$$

Conversely, for any $e_R = (x, y) \in \eta_R(X)$ there exists a string uxmyv $\in L$ and for all u', v', if $\exists m_0$ such that $u'xm_0yv' \in L$ then $(u'x, yv') \in E_T$.

So if m occurs_T in all $e \in Y$, then m occurs_T in $(u'x,\ yv')$, and m occurs_R in (x, y). Hence,

$$e_R \in \eta_R(X) \text{ and } m \in \mu_T(Y) \Rightarrow m \in \mu_R(e_R)$$

or equivalently,

$$\text{(b) } m \in \mu_T(Y) \Rightarrow m \in \mu_R\eta_R(X).$$

Since X was closed, $\mu_R\eta_R(X) = X$ and combining (a) and (b) we have

$$X = \mu_T(Y) \in \mathcal{M}_T\,.$$

Thus every class in $\mathcal{M}_R$ is a class in $\mathcal{M}_T$.

Our discussion of alternative interpretations of E given in (5) — (7) leads to the following conclusions:

(i) The definition of partial environments (6) is too broad and leads to a large number of 'unnatural' classes.

(ii) The restricted partial environments (7) do not include all distributionally relevant environments, and do not yield any classes not generated by E_T.

For this reason, we shall, for the purpose of interpreting the structure $\mathcal{M}$, use the simplest definition of 'occurs in' (5.1), and take $E = E_T$.

Remark:

It was, of course, the observation that partial environments exist which provided the basis for the attempts by taxonomic linguists to define immediate constituents in distributional terms, but the concept does not seem to have been given any precise definition. The principal task for a rigorous reconstruction of immediate constituent analysis would be to formalize the conditions under which partial environments can be treated as 'constituents', and to make clear the relationship between such environments and sequences of classes obtained from their distribution.

3. EQUIVALENCE AND SIMILARITY

We observe that $\mathcal{M}$, being complete, always has a greatest element $\Sigma\mathcal{M}$ which is equal to M. Similarly $\Sigma\mathcal{E} = E$. Since $\mathcal{E}$ is the dual of $\mathcal{M}$, by Prop. 1C the least element of $\mathcal{M}$

$$\Pi\mathcal{M} = \mu(\Sigma\mathcal{E}) = \mu(E).$$

Hence $\Pi\mathcal{M}$ = the empty set Φ iff no m occurs in every environment. This seems plausible as a general property of natural languages (except perhaps for early forms of children's grammars), and as it allows us to simplify certain subsequent proofs, we restrict ourselves henceforth to languages for which

(8) $\Pi\mathcal{M} = \Phi$.

The corresponding property for E is more problematic. If a language permits quotation, then there will exist environments illustrated by

'He said"____".'

which accept as minimal substitution instances words, or, on another interpretation of quotation, any string of sounds or letters representing the sounds in the quoted utterance. But, in general, there will be no single environment whose minimal substitution instances are precisely the set of morphemes, so we will also assume that

(9) $\eta(\mathrm{M}) = \Pi E = \Phi$.

We recall some basic definitions from lattice theory: an element x of a lattice L is said to *cover* another element y if $x \supset y$, and for no z does $x \supset z \supset y$ hold. The elements of L which cover the least element of L are called *atoms* and the elements covered by the greatest element of L are *co-atoms*. We shall call the greatest and least elements of M (M and Φ) the *trivial* classes of M; the atoms and co-atoms of M are then the minimal and maximal [4] non-trivial classes, respectively. That this redefinition works follows from the fact that (8) and (9) together with our original definition of M and E guarantee the existence of non-trivial classes.

Proposition 2: M has at least two non-trivial classes.

Proof: The case $M = \{\mathrm{M}\}$ cannot occur since then $\Pi M = \mathrm{M}$ which contradicts $\mathrm{M} \neq \Phi$. Suppose $M = \{\mathrm{M}, \Phi\}$. Then since every environment e has, by definition, at least one substitution instance, and $\mu(\mathrm{e})$[5] is closed, $\mu(\mathrm{e}) = \mathrm{M}$. But this means $\eta(\mathrm{M}) = \bar{\mathrm{e}}$ which contradicts $\eta(\mathrm{M}) = \Phi$.

The remaining case to be considered is $M = \{\mathrm{M}, \mathrm{X}, \Phi\}$ with $\Phi \neq \neq \mathrm{X} \subset \mathrm{M}$. Since $\mu(\mathrm{e})$ cannot equal M, and $\mu(\mathrm{e}) \neq \Phi$, we must have $\mu(\mathrm{e}) = = \mathrm{X}$ for all $\mathrm{e} \in \mathrm{E}$, which implies $\mu(\mathrm{E}) = \mathrm{X}$, contradicting $\mu(\mathrm{E}) = \Phi$. Q.E.D.

The atoms and co-atoms of M have special properties related to traditional notions of distributional classification. Perhaps the oldest such notion is that of classification on the basis of identical distribution, which as Hiż [1960] has pointed out, goes back at least to Husserl's *Logische Untersuchungen* (1913).

Two morphemes have identical distribution, or briefly, are *equivalent*, if $\eta(\mathrm{m}_1) = \eta(\mathrm{m}_2)$. The relation of equivalence partitions M into maximal sets of equivalent morphemes, which, following Kulagina [1958] and Marcus [1967], we will call *families*. (The family to which m belongs will be denoted by $\hat{\mathrm{m}}$.) As we shall see, families constitute the 'building blocks' of which classes are composed. But the families of a language, by themselves, do not in general yield an interesting classification of M. In a finite corpus, the fact that two morphemes have identical distribution can be usually be taken as an accidental characteristic of the choice of

---

[4] A set is *maximal* with respect to a property if it is not properly contained in any larger set having that property; it is *minimal* if it does not properly contain any smaller set with that property.

[5] $\mu(\mathrm{e})$ and $\eta(\mathrm{m})$ are abbreviations for $\mu(\{\mathrm{e}\})$ and $\eta(\{\mathrm{m}\})$. Similarly we abbreviate $\{\mathrm{x}\}$ by $\bar{\mathrm{x}}$.

sentences in the corpus, and for the language as whole, one could argue (and a distributional theory of meaning would require) that if two morphemes have completely identical distribution, then there can be neither a semantic nor a syntactic basis for distinguishing between them, so that they are in fact not distinct morphemes after all. Thus, computing the families of a natural language may give us nothing more than the trivial partition $= \{\{m\} : m \in M\}$.

The following propositions relate the notion of equivalence to that of closure:

Proposition 3: A class is a family iff it is an atom of $\mathcal{M}$.

Proof: Suppose F is a closed family, and let F′ be a non-empty class $\subseteq$ F. Pick $m \in F'$. Since $F' \subseteq F$, and η is constant on F, $\eta(m) = \eta(F)$. Then $\overline{m} = \overline{F} = F$, but by 1A(c) $\overline{m} \subseteq \overline{F'} = F'$, so $F = \overline{F'}$, and hence F covers Φ.

Conversely, if F covers Φ, then by 1C, $\eta(\Phi) = E$ covers $\eta(F)$. Now if $m \in F$, $\eta(m) \supseteq \eta(F)$ by 1A(a), and so either

$$\eta(m) = \eta(F)$$

or

$$\eta(m) = E.$$

But if $\eta(m) = E$, $\mu(E) = \overline{m} \neq \Phi$ which contradicts (8), so $\eta(m) = \eta(F)$ for all $m \in F$.

Proposition 4 (Kunze, 1967): If C is a class, then $C = \bigcup \{\hat{m} : m \in C\}$ where $\hat{m}$ is the family to which m belongs.

Proof: It suffices to show that if C is a class, $m \in C$, and $\eta(m) = \eta(m')$, then $m' \in C$. From $\eta(m') = \eta(m)$, we have $\overline{m} = \overline{m'}$. Since $m \in C$ and C is closed, $\overline{m} \subseteq C$ and hence

$$m' \in \overline{m'} = \overline{m} \subseteq C. \qquad \text{Q.E.D.}$$

(Note that the converses to Props. 3 and 4 do not hold: in general, not all families are classes. The case in which families and atoms coincide is discussed further in sections 6 and 7.)

The following proposition, though it requires rather cumbersome formal machinery, is not as complicated as it looks; it expresses the intuitively expected result that 'collapsing' equivalent morphemes into single elements leaves the structure of $\mathcal{M}$ invariant.

(10.) *Definition*: Let $\hat{m}$ be the family to which m belongs, and for $x = m_1 \dots m_k$, define $\hat{x} = \hat{m}_1 \dots \hat{m}_k$

$$= \{a_1 \dots a_k : a_i \in \hat{m}_i \text{ for } i = 1, \dots, k\}.$$

Set

$$\hat{M} = \{\hat{m} : m \in M\},$$

$$\hat{E} = \{(\hat{x}, \hat{y}) : (x, y) \in E\}.$$

If m occurs in e, then $\hat{m}$ is said to occur in $\hat{e}$.

Proposition 5: The lattice $\hat{M}$ of classes obtained from $\hat{M}$ and $\hat{E}$ is isomorphic to M.

Proof: For any $X \subseteq M$, let $\hat{X} = \{\hat{m} : m \in X\}$. By Prop. 4, if $X \in M$, the mapping $X \to \hat{X}$ is 1-to-1, the inverse mapping being that which sends $\hat{X} \to \cup\hat{X} = X$. Also, by Prop. 4, if $X, Y \in M$ then $X \subseteq Y$ iff $\hat{X} \subseteq \hat{Y}$, so $\hat{M} = \{\hat{X} : X \in M\}$ is a lattice isomorphic to M. What remains to be shown is that if X is closed (with respect to E) then $\hat{X}$ is closed (with respect to $\hat{E}$) and *vice-versa.*

The key idea of the proof is that if $\hat{m}$ occurs in $\hat{e}$, as defined in (10), then $\hat{e} \subseteq \eta(\hat{m})$. To see this, we note first that if m occurs in e, $e \in \eta(\hat{m})$, since any m′ equivalent to m also occurs in e. Now let $e = (x_1 \ldots x_j, x_{j+1} \ldots x_k)$ where the $x_i \in M$. If we replace x_1 by an equivalent x_1', then since x_1 occurs in

$$(,x_2 \ldots x_j\ m_0 x_{j+1} \ldots x_k)$$

for any $m_0 \in \hat{m}$, x_1' also occurs in these environments. So all $m_0 \in \hat{m}$ occur in

$$(x_1'x_2 \ldots x_j, x_{j+1} \ldots x_k).$$

By a similar argument for $i = 2, \ldots k$, any replacement of an x_i in e by an equivalent x_i' yields an environment $e' \in \hat{e}$ in which all elements of $\hat{m}$ occur. Hence $\hat{e} \subseteq \eta\ (\hat{m})$. The case in which the left or right parts of the environment are the empty string Λ can be handled in a similar fashion by defining $\hat{\Lambda} = \{\Lambda\}$.

Assume X is closed, and let $\hat{e}$ be an environment in $\hat{E}$ in which all $\hat{x} \in \hat{X}$ occur. To prove that $\hat{X}$ is closed, we need to prove that if $\hat{m}$ occurs in $\hat{e}$ then $\hat{m} \in \hat{X}$. Since all $\hat{x} \in \hat{X}$ occur in $\hat{e}$, $\hat{e} \subseteq \bigcap (\eta(\hat{X}) : \hat{X} \subseteq X) = \eta(\cup\{\hat{x} : x \in X\}) = \eta(X)$. Now, $\hat{m}$ occurs in $\hat{e}$ implies m occurs in e and $e \in \hat{e} \subseteq \eta(X)$. Hence, $m \in X$ (since X is closed) and so $\hat{m} \in \hat{X}$.

To show the converse, observe that if $e \in \eta(\cup\hat{X})$ then for any $x \in \cup\hat{X}$, $\hat{x}$ occurs in $\hat{e}$. If m also occurs in e, then if $\hat{X}$ is closed, $\hat{m} \in \hat{X}$, which implies $m \in \hat{m} \subseteq \cup\hat{X}$. So X closed implies $\cup\hat{X}$ is closed, which completes the proof.

(Further 'collapsing' of inequivalent morphemes, say by replacing M by $\overline{M} = \{\overline{m} : m \in M\}$, does not leave the structure of M invariant. For example, if $L = XY$ where $\Phi \subset X \subset Y \subset M$, then X, and Y are closed, and we have for all $m \in M$ that $\overline{m} = X$ if $m \in X$, otherwise $\overline{m} = Y$. In this case, the environments of $\overline{M}$ are just $\{(,Y), (X,)\}$, and $\overline{M} = \{\overline{M}, \{X\}, \{Y\}, \Phi\}$ which is not isomoprhic (or even homomoprhic) to M, since $\{X\} \not\subseteq \{Y\}$.)

Families arise by clustering together objects with identical properties. Since this yields in many cases of interest an extremely fine or-even trivial classification, one can instead consider weakening the notion of equivalence to one of *similarity*, in which the reflexivity and symmetry of equivalence relations is retained but the requirement of transitivity is dropped. (We can have *a* similar to *b* and *b* similar to *c* without implying that *a* is similar to *c*.)

There is a well-known procedure for constructing a classification from a similarity relation which follows the pattern of constructing families: the classification is induced by considering the maximal sets of mutually similar objects. Carnap in 1928, applying this notion to epistemological problems, referred to the elements of the induced classification as 'similarity circles' [Carnap, 1967]. The same concept appears in graph theory under the name 'clique' [Berge, 1958], or 'maximally connected subgraph', and has been applied under various names in areas ranging from ontology (cf. Sommer's U-relation and B-types, [1963]) to information-retrieval [Kumar, 1968]. We shall use the term *S-clique* to refer to the cliques obtained from a particular similarity relation S.

To obtain a specific similarity relation defined in terms of properties, we can weaken the notion of distributional equivalence by defining x, y $\in$ M to be similar iff they have a *single* property in common. Let us call this relation σ; the σ-cliques can then be defined as follows:

(11) For $\Phi \neq X \subseteq M$, X is an *σ-clique* if $\eta(x) \cap \eta(y) \neq \Phi$, for all $x \in X$, implies $y \in X$.

The classification induced by σ is, in general, very coarse; in the worst case, where the set of properties is, so to speak, too rich, any two objects have some property in common and there is only one σ-clique, namely M. We note further that every class in M (except perhaps M) consists of similar elements (since if $X \in M$ and $X \neq M$ then $\eta(X) \neq \Phi$ and so all elements of X have some property in common.) It follows that every class is contained in an σ-clique. But σ-cliques are not, in general, closed; a non-trivial σ-clique is only closed if all its members are similar with respect to the same single property.

Proposition 6: Let X be an σ-clique, $X \subset M$, with $\eta(X) \neq \Phi$. Then X is closed and a co-atom of M.

Proof: Let $x \in X$ and $y \in \overline{X}$, then $\eta(y) \supseteq \eta(\overline{X}) = \eta(X)$, and $\eta(x) \supseteq \eta(X)$. Hence

$$\eta(x) \cap \eta(y) \supseteq \eta(X) \neq \Phi$$

and so by (11), $y \in X$.

Suppose M does not cover X (in M). Then there exists a $Y \in M$ such that $X \subset Y \neq M$. By our previous remark, this implies that X is properly contained in some σ-clique containing Y. But this contradicts the maximality of X as a σ-clique.

4. THE PROBLEM OF THE MIDDLE

By the duality of E and M, the co-atoms of M are images under the mapping μ of atoms of E; so we can apply Prop. 3 to E, *mutatis mutandis*, to yield the result that the environments shared by elements of a co-atom are all equivalent, that is:

(12) if X is a co-atom, $\mu(e) = \mu(e')$ for all e, $e' \in \eta$ (X),

and hence

(13) $X = \mu(e)$ for any $e \in \eta$ (X).

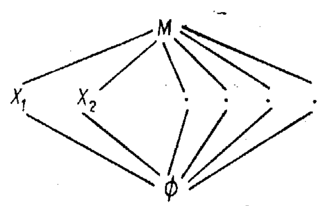
The correspondence between the distributional equivalence of the elements of atoms and the equivalence of properties associated with the elements of co-atoms allows us to formulate, if not to solve, a basic dilemma which arises in devising classificatory criteria — how to strike a happy medium between too fine and too coarse classifications? If (within the latice *M*) we attempt to construct a classification by clustering together distributionally indistinguishable elements, we obtain 'small' disjoint classes — the atoms of *M*, and since not all families need be closed, some elements of *M* may be left unclassified. If we cluster objects on the basis that they occur in distributionally indistinguishable environments, we obtain the co-atoms of *M*. These are 'large' (typically overlapping) classes whose union is M, but whose elements are only mutually substitutable in 'small' sets of environments — the atoms of *E*.

Ideally one would like to combine these two criteria to produce a classification $\{X_1, X_2 \ldots\}$ which would satisfy simultaneously the constraints:

if $m_1, m_2 \in X_i$ and m_1 occurs in e, then m_2 occurs in e;

if $m_i \in X_i$ and occurs in e and m_2 occurs in e, then $m_2 \in X_i$,

but as we have seen (cf. Prop. 3, and (13)) this is only possible in the special case that the atoms and co-atoms of *M* coincide. Here *M* has a trivial structure which can be represented graphically [6] as follows:



The ideal case serves to clarify why it is so difficult to formulate a classification criterion which approximates it. To construct such a criterion would involve defining something like the 'middle' of a lattice, and lattices in general simply do not have well-defined middles. One can, of course, define constructions which capture something of the notion of the 'middle', for example, by maximizing with respect to one lattice-

[6] To represent lattices graphically, we use a *Hasse diagram*, in which the nodes or vertices of the graph represent the elements of the lattice, with two nodes being connected by an edge if one of the corresponding lattice elements covers the other. If $x \geqslant y$ in the lattice, then the node x is placed above the node y in the diagram.

property and minimizing with respect to another, but there are many ways such a construction can be carried out, and in order to choose between them we would need a more detailed analysis of the theoretical aims of such classification criteria than is available at present.

Remark : In order to make the problem of defining a mean between too fine and too coarse classification reasonably coherent, we need some sort of structure which mediates or connects the extremes and is dependent on the actual distribution of objects and properties. The lattice M is one such structure, but as we have seen it fails to connect the intuitively plausible extremes of equivalence and σ-similarity. It, however, suggests the following (to my knowledge) open problem :

Find a lattice whose structure is dependent on the underlying distribution of M and E with the property that its atoms are the set of families and its co-atoms are the set of σ-cliques.

5. CLASSIFICATION AND ASSIGNMENT

Suppose we are given a classification $\{X_1, \ldots, X_k\}$ of M where the sets X_i are to serve as types or categories of objects. Unless the X_i form a partition of M, some objects will belong to more than one category. For linguistic purposes, we should like the categories to reflect in some sense the 'roles' or 'functions' of morphemes in sentences, and so membership in more than one category should correspond to multiple roles or functions. In the ordinary way of describing English sentences, one says, for example, that 'sign' functions as a noun in the sentence

'Noam letters signs."

and as a verb in

"Noam signs letters."

To capture the flavor of this, we need a notion of 'assignment' of an object to a category relative to an environment, over and beyond 'belonging' to a category in the sense of set-membership.

(14) *Definition* : x is ***assignable*** to $X \subseteq M$ in the environment e if $x \in X \subseteq \mu(e)$.

(15) *Definition* : Let $G \subseteq M$-M. G is a *complete classification* if for all $m \in M$, and $e \in \eta(m)$, m is assignable to some $G \in G$ in e. (The reason for excluding M from G is that, as $\eta(M) = \Phi$, there is no environment in which m is assignable to M.)

Proposition 7 : G is a complete classification iff for all e, $\mu(e)$ is the union of some set G' of classes in G.

Proof : Combining (14) and (15), if G is complete then for each $m \in \mu(e)$, there exists a $G_m \subseteq \mu(e)$ such that $m \in G_m \in G$. Hence, letting $G' = \{G_m : m \in \mu(e)\}$, we have

$$\mu(e) = \bigcup G'.$$

The converse is obvious.

The classification of morphemes discussed by Harris and Wells is based on sets of morphemes obtained by substitution in a single frame or environment, i.e. classes of the form $\mu(e)$. Wells called such classes 'morpheme-classes', but as we are here using the term 'class' in a wider sense, we shall refer to them as *diagnostic* classes, since a set $X \subseteq M$ is of the form $\mu(e)$ iff there exists an environment e which is 'diagnostic' in the sense that 'occurs in e' is equivalent to 'is an element if X'.[7]

The structure of the diagnostic classes determines the structure of the remainder of M, for if X is a closed set, we have

$$X = \bigcap \{\mu(e) : e \in \eta(X)\}.$$

Hence, if we begin with the diagnostic classes and compute all possible intersections, we obtain the lattice M.

Any classification which includes all diagnostic classes is complete, but such a classification may contain redundant classes which can be expressed as the union of smaller classes. This raises the question of finding a complete classification containing the fewest possible classes. The problem need not have a unique solution in general, but if we restrict the classes in a classification to be diagnostic, there is exactly one smallest complete classification:

Proposition 8: Let

$G_D = \{\mu(e) : \mu(e)$ is not a union of smaller diagnostic classes$\}$.

Then G_D is the smallest complete classification by diagnostic classes.

Proof: G_D is a complete classification, since for every $\mu(e)$, either $\mu(e) \in G_D$ or is the union of smaller diagnostic classes, which in turn are either in G_D or the union of smaller diagnostic classes. This process must terminate since the classes at each stage become smaller and M is finite; so every $\mu(e)$ is a union of elements of G_D, and hence G_D is a complete classification.

Further, every complete classification consisting of diagnostic classes must contain G_D; hence G_D is the smallest such.

A methodological digression:

A definition such as that given in Prop. 8 for G_D can be taken as a proposal that a set of grammatical categories of a language be defined in such-and-such a way. It seems appropriate therefore to consider, at this juncture, the methodological status of such proposals. Fundamentally, a particular classification criterion and the mathematical structure within which it is embedded simply enables (in principle) certain calculations to be performed which can generate precise facts about particular languages, to wit:

"Applying criterion A to language B yields categories C". Obtaining such facts does not seem to me to be quite as uninteresting as the current lack of interest in discovery procedures would suggest, but

[7] Kunze (1967) defines a set X to be *complete* if for all $Y \neq \Phi$, $\eta(X) \subseteq \bigcup \{\eta(y) : y \in Y\}$ implies $X \cap Y \neq \Phi$, and as he shows, this is equivalent to X being diagnostic.

unfortunately the calculations are for obvious combinatorial reasons quite difficult in practice. (Some initial computer experiments suggest that they are barely feasible, but that the cost will be considerable for any corpus large enough to be representative.) Clearly, any distributional grammatical theory must be supplemented by quite sophisticated algorithms if we are to be able to see its consequences.

However, classification criteria need not be purely instrumental; they can provide a basis for well-defined empirical hypotheses. For if a criterion is defined in the context of a particular mathematical framework, then further properties can be stated within that framework which are either true or false of the particular classification generated by the criterion for a particular set of data. In order for such a property of a classification to be empirical, it must be contingent on the data; that is, the criterion used to define the classification must itself be so framed as prevent the property from being either necessarily true or necessarily false.

To illustrate, let us look again at classifications which are derived or are presumed to be derivable from considerations of similarity. If the resulting categories are in some sense intuitive or 'natural' and the similarity relation is hard to state precisely, one may be tempted to take the fact that two objects are assigned to the same category as itself a form of similarity. Why is 'raven' like 'writing-desk'? A possible answer is that they are both nouns; but at the same time, the question "what is a noun?" can have the reply that nouns are words which share the grammatical similarities of 'writing-desk' and 'raven'.

This circularity can be reformulated as a condition on classification criteria: given a classification $G = \{X_1, \ldots X_k\}$, we can treat the set G itself as a collection of new properties, defining 'x has the property X_1' as synonymous with '$x \in X_i$'. From these properties, we define a similarity relation $\sigma(G)$ as σ was defined as Sec. 3 : x $\sigma(G)$ y if for some $X_1 \in G$, x, $y \in X_i$. Now let $c(\sigma(G))$ be the collection of $\sigma(G)$-cliques. The classification G admits of the circular interpretation described above, if the similarity relation $\sigma(G)$ yields the same classification as G itself; that is, if

(16) $c(\sigma(G)) = G$.

(One motivation for constructing similarity relations from classifications by means of σ and classifications from similarity relations by means of c is that the reverse equation to (16) always holds : if S is a similarity relation it is easy to verify that $\sigma(c(S)) = S$.)

Now if (16) is to be a contingent property of a classificatory criterion, so that whether or not it is satisfied by some distributionally defined classification of a particular natural language becomes a question with empirical content, then certain classificatory criteria can be ruled out in advance. For example, the set F of families of M is excluded as an interesting type of classification, since F is always a partition and (16) always holds for partitions. Similarly, the set C of σ-cliques (see (11)) is excluded, for we always have

$$c(\sigma(C)) = c(\sigma(c(\sigma))) = c(\sigma) = C.$$

By way of contrast, complete classifications are not excluded by the requirement that (16) be contingent, since examples can be constructed for which it fails, and others for which it holds.

6. AMBIGUITY AND CONTEXT-INDEPENDENCE

For linguistic purposes, two aspects of assignment of objects to categories are of particular interest:

(i) When is an object assignable in each environment to a unique category?

(ii) When is an object assignable to the same categories in all environments?

We make the following definitions:

(17) m is *unambiguous in* e with respect to a given complete classification G if m is assignable to only one $G \in G$ in e.

(18) m is *context-independent* with respect to G if m is assignable to G in all $e \in \eta(m)$.

Proposition 9: G is an unambiguous classification (i.e. all $m \in M$ are unambiguous with respect to G) iff $G_1, G_2 \in G$ and $G_1 \neq G_2$ implies either $G_1 \cap G_2 = \phi$ or $G_1 + G_2 = M$.

Proof: If G is unambiguous, then for any m, if $m \in G_1 \subseteq \mu(e)$ and $m \in G_2 \subseteq \mu(e)$, $G_1 = G_2$. So if $G_1 \neq G_2$, and $G_1 \cap G_2 \neq \phi$, no diagnostic class contains $G_1 \cup G_2$. But, by the definition of join, this implies no diagnostic class contains $G_1 + G_2$. Then since every non-trivial class is contained in a diagnostic class, we must have $G_1 + G_2 = M$.

Conversely, suppose $G_1 + G_2 = M$. So any class containing $G_1 \cup G_2$ equals M. But, from (9), M is not a diagnostic class, so no diagnostic class contains $G_1 \cup G_2$, and hence no $m \in G_1 \cup G_2$ can be assignable to both G_1 and G_2 in the same environment.

Thus the ambiguity of a complete classification can be 'read off' from the structure of the lattice M in which it is embedded: in Fig. 1, the classification {A, B, C} is ambiguous while {D, E, F} is not.

The condition for unambiguous assignment with respect to G_D has a possible application as a purely distributional criterion for the detection and resolution of syntactic homonyms. For, as illustrated in the example which follows, elements of M which are ambiguous can in some circumstances be 'split' to create new unambiguous elements. To accomplish this, we need a rule to determine in which environments the new elements are to occur:

(19) *Disambiguation rule*: If $G_1, G_2 \in G_D$, $G_1 \not\subseteq G_2$ and $G_1 \cup G_2 \subseteq \mu(e)$, then replace any elements m in $G_1 \cap G_2$ by new elements, m^* and m^{**}. Let m^* occur in all $e \in \eta(G_1)$, m^{**} in all $e \in \eta(G_2)$. If $m \in \mu(e) \subset G_1$, let m^* occur in e, similarly for $m \in \mu(e) \subset G_2$, and m^{**}.

The reason for the condition $G_1 \not\subseteq G_2$ (rather than just $G_1 \neq G_2$) is as follows: if $G_1 \subset G_2$ and we were to apply (19) to $m \in G_1 \cap G_2 = G_1$, m^* and m^{**} would satisfy

$$m^* \in \mu\eta(G_1),$$
$$m^* \in \mu\eta(G_2), \text{ (since } \eta(G_1) \supseteq \eta(G_2))$$
$$m^{**} \in \mu\eta(G_2).$$

Hence $\mu\eta(G_1)$, which is the class resulting from replacing each m in G_1 by m*, would be contained in $\mu\eta(G_2)$ (obtained by replacing m by m**) and the elements m* would again be ambiguous.

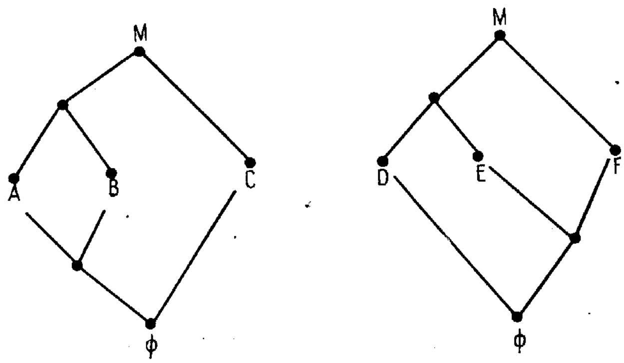


Figure 1.

This suggests slightly modifying the definition of assignment w.r.t. $\mathcal{G}_D$ in order to make the case $G_1 \subset G_2$ no longer ambiguous :

(20) m is assignable to G in e if G is a *maximal* element of $\mathcal{G}_D$ satisfying $m \in G \subseteq \mu(e)$.

This definition, rather than (14), is used in the example which follows.

To illustrate, consider the corpus *L*, and the corresponding lattice *M* given in Fig. 2. Here

M = {up, down, high, low, er, ly ,ing, ed, s}

L : up	up(p)-er		up(p)-ing	up(p)-ed	up-s
down			down-ing	down-ed	down-s
high	high-er	high-ly			high-s
low	low-er	low-ly	low-ing	low-ed	low-s

Abbreviating the elements of M as U, D, H, L, r, y, g, d, s, the lattice *M* is :

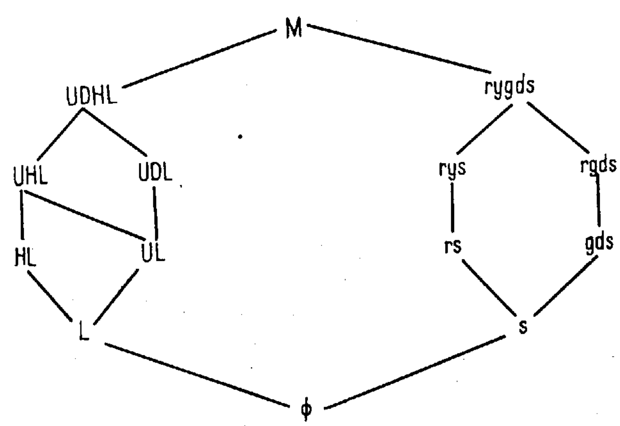


Figure 2.

$G_D = \{\{U, H, L\}, \ \{U, D, L\}, \ \{H, L\}, \ \{r, y, s\}, \ \{r, g, d, s\}, \ \{g, d, s\}\}$,
and we find the following ambiguities:

U and L are assignable to {U, H, L} and {U, D, L} in the environment (,).

r and s are assignable to {r, y, s} and {r, g, d, s} in the environment (low,).

Applying (19) to U and L yields L_1 and M_1, given in Fig. 3. For M_1,

$$G_D = \{\{U_1, H, L_1\}, \ \{U_2, D, L_2\}, \ \{H, L_1\}, \ \{r, y, s\}, \ \{r, s\}, \ \{g, d, s\}\}$$

and this classification is now unambiguous. Considering only the maximal categories in G_D (the remaining categories are only needed to make the classification complete), we see that splitting 'up' and 'low' has produced, as we might expect, a 'noun/adjective' category {up_1, high, low_1}, a 'verb' category {up_2, down, low_2} and corresponding suffix categories: {er, ly, s}, and {ing, ed, s}. Note that 's' is not ambiguous in L_1 — whether it is functioning as a 'noun/adjective' suffix or a 'verb' suffix can be determined from the context in which it occurs.

L_1:	U_1	U_1r				U_1s
	U_2			U_2g	U_2d	U_2s
	D			Dg	Dd	Ds
	H	Hr	Hy			Hs
	L_1	L_1r	L_1y			L_1s
	L_2			L_2g	L_2d	L_2s

M_1:

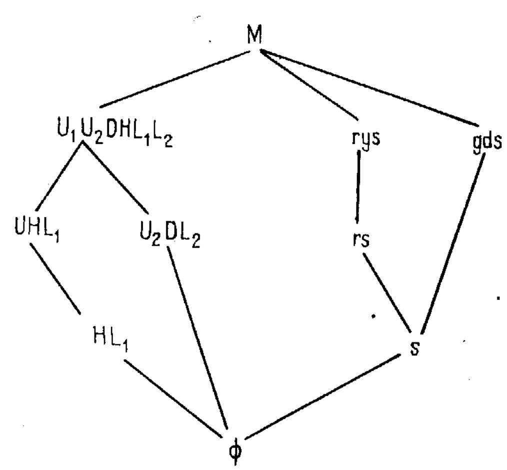


Figure 3.

For this artificial corpus, the intuitive solution is also the most economical. If, for example, we split r and s first, rather than U and L, we obtain a slightly more complicated lattice M_3 (Figure 4) in which L is still ambiguous. Splitting L then yields M_4 (Figure 5) and the unintuitive classification

$$G_D = \{\{U, D, L_2\}, \{U, L_2\}, \{H, L_1\}, \{r_1, y, s_1\}, \{r_2, g, d, s_2\}, \{g, d, s_2\}\},$$

in which there are two forms of the comparative "-er", one for the 'noun-adjective' category {high, low_1} and one for the 'verbs' {up, low_2}.

It is interesting to note that if the ambiguity of a classification is to be invoked as a means of detecting syntactic homonymity, then the classification criterion must yield classifications which, roughly speaking, lie in the middle of the lattice *M*. The situation here is analogous to that noted earlier with respect to the 'circularity' condition (16), for we would like unambiguity to be a contingent fact about elements of M; but if the categories in a classification are very 'large', the condition for unambiguity :

$$X_1 + X_2 = M$$

will be satisfied automatically; if the categories are very 'small', ther the condition

$$X_1 \cap X_2 = \Phi$$

will hold, and again unambiguity is ensured.

L_3 :	U		Ur_2		Ug	Ud		Us_2
	D				Dg	Dd		Ds_2
	H	Hr_1		Hy			Hs_1	
	L	Lr_1	Lr_2	Ly	Lg	Ld	Ls_1	Ls_2

M_3 :

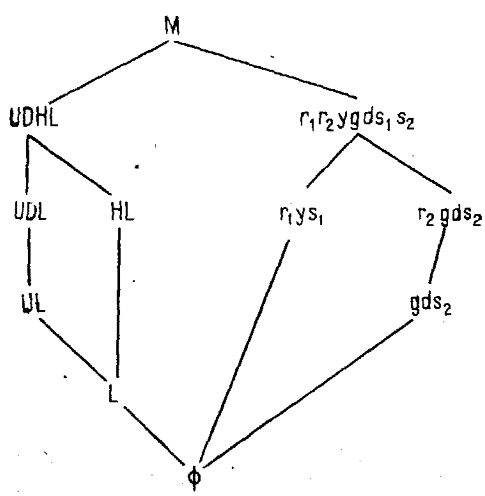


Figure 4.

L_4:	U		Ur_2		Hg	Ud		Us_2
	D				Dg	Dd		Ds_2
	H	Hr_1		Hy			Hs_1	
	L_1	L_1r_1		L_1y			L_1s_1	
	L_2		L_2r_2		L_2g	L_2d		L_2s_2

M_4:

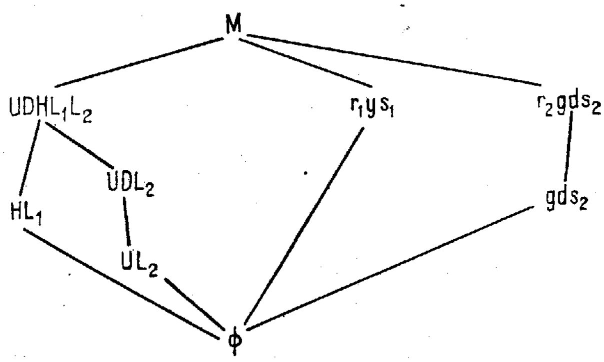


Figure 5.

Like ambiguity, context-independence can be characterized in terms of the lattice structure of M.

Lemma 2: m is context-independent w.r.t. G, iff $G = \overline{m}$.

Proof: If m is context-independent with respect to G then

$$G \subseteq \mu(e) \text{ for all } e \in \eta(m).$$

Hence

$$G \subseteq \bigcap \{\mu(e) : e \in \eta(m)\} = \overline{m}.$$

But G is closed and $m \in G$, so

$$\overline{m} \subseteq \overline{G} = G$$

from which we have $\overline{m} = G$.

Conversely, if $G = \overline{m}$, then $m \in \mu(e)$ implies $\overline{m} \subseteq \overline{\mu(e)} = \mu(e)$, so m is assignable to G in all $e \in \eta(m)$.

Corollary: If m is context-independent w.r.t. all $G \in m$, it is unambiguous w.r.t. G.

Proposition 10: Let G be a complete classification of M. Then all $m \in M$ are context-independent w.r.t. all $G \in G$ iff

$$G = \{\hat{m} : m \in M\} = \text{set of atoms of } M.$$

Proof: Let $G \in \mathcal{G}$. Since $G \neq M$, $\eta(G) \neq \Phi$, and as all m are context-independent, any $m \in G$ is assignable to G in $e \in \eta(G)$. So, by Lemma 2, $G = \overline{m}$ for all $m \in G$. But this entails

$$\eta(m) = \eta(\overline{m}) = \eta(G) \text{ for all } m \in G,$$

and so G is contained in a family. Hence, by Prop. 3 and 4, G is a family and an atom of $\mathcal{M}$.

Conversely, let $\mathcal{G}$ = set of families of M. If all families are atoms (and hence closed), then $\hat{m} = \overline{m}$ for all m, and as $\overline{m} \subseteq \mu(e)$ for any $e \in \eta(m)$, all m are context-independent w.r.t. all $G \in \mathcal{G}$.

A class containing an m which is context-independent with respect to it is of the form $\overline{m}$ by Lemma 2. Such classes, which we will call CI classes are, in a sense, the 'duals' of diagnostic classes — the former are the closures of single elements of M and will tend to be at the 'bottom' of the lattice, while the latter are defined by single environments and will tend to be at the 'top'.

The duality of diagnostic and CI classes appears to be related to the long-familiar distinction between 'open' and 'closed' grammatical categories. (Let us call them O and C categories, since 'closed' is being used in this paper for another purpose.) Typical O categories are (in English) nouns, adjectives, and verbs, which are distinguished, among other things, by the fact that they are large and can be extended by new lexical items; that is, neologisms can be assigned their grammatical functions and thus can have similar distributions. The C categories include articles, pronouns, prepositions, etc.; they are small categories (easily enumerated) and neologisms cannot be assigned their grammatical functions. Obviously there is much more to the distinction than this, but within the present framework, we can only scratch the surface-structure for distributional symptoms of the dichotomy, ignoring its further syntactic and semantic ramifications.

Consider the classical example:

"The gostak distims the doshes."

How is it that we know that the doshes are being distimmed and that there are not two gostaks doing it? Clearly some process of assignment of nonsense-words to grammatical categories is involved; but as the words are nonsense, the assignment must, in such instances, be possible without reference to the orthographic 'shape' of the word being assigned, and we will take this as the relevant symptom that the categories involved are O categories.

To translate this condition into a property of classes in $\mathcal{M}$, we observe that, by the definition of assignment (14), we can assign m to G in e without knowing which m it is, if we know that m occurs in e, and that $G \subseteq \mu(e)$, and it is the case that these two conditions imply $m \in G$. But it is easy to see that

$$(21) \quad m \in \mu(e) \text{ and } G \subseteq \mu(e) \text{ imply } m \in G$$

exactly when G is the diagnostic class $\mu(e)$.

Now imagine an algorithm for class-assignment which takes as input a sentence $m_1 \ldots m_k$ from which it produces a list of pairs (m_i, e_i) where $e_i = (m_1 \ldots m_{i-1}, m_{i+1} \ldots m_k)$, and then uses the structure of M to determine to which classes m_i can be assigned in e_i. Since m can always be assigned to $\mu(e)$, the algorithm can compute the 'name' of this class directly from e without using the datum m, and hence can perform the same assignment for the pair (n, e) even when n is a neologism or unknown, i.e. $n \notin M$. With respect to such an algorithm, the diagnostic classes and only those will be 'extendable'.

If such an assignment algorithm is to be able to handle the above example, it must in some cases be able to identify the appropriate diagnostic class from only small portions of e which serve, so to speak, as syntactic clues, since as in the example, new words can be assigned to parts of speech even when the environments they occur in themselves contain unknown words. The obvious candidates for such syntactic clues are the so-called 'function words', i.e., members of C categories. But if function words are to serve as syntactic clues even in environments containing unknown words, their grammatical role or category must be identifiable independently of the environments in which they occur. This suggests a converse property to (21):

(22) $m \in \mu(e)$ and $m \in G$ implies $G \subseteq \mu(e)$,

which, by Lemma 2, is exactly equivalent to $G = \overline{m}$. Thus, an assignment algorithm, when given an input pair $(\overline{m}, e)$ can immediately make one assignment for m, say, by looking up $\overline{m}$ in a table of CI classes, without considering the internal structure of e; and this respect, CI classes resemble C categories, just as diagnostic classes resemble O categories with respect to extendability.

Note that we cannot deduce from the fact that a class is CI that it is non-extendable (i.e. not a diagnostic class), even though this is usually taken as the defining characteristic of C categories. In fact, CI classes which are diagnostic arise quite naturally as distributional symptoms of grammatical agreement. For example, if m is the only morpheme which can occur in e, then $\mu(e) = \{m\} = \overline{m}$. An instance in English is given by taking m = 'I', and e = (,am).

We suggest as a somewhat speculative hypothesis that CI diagnostic classes are in fact not perceived as 'natural' categories or parts of speech, thus accounting for their non-extendability. If this hypothesis is correct, it poses some difficulties for any attempt to define, on purely distributional grounds, classifications which aim at capturing the categories suggested by linguistic intuition; for as CI classes cannot be unions of smaller classes, any diagnostic CI class must be included in every complete classification by Prop. 7. (The same result was previously proved by Kunze (1968) for classifications meeting a stronger condition than completeness : namely, that for any $x = m_1 \ldots m_k \in L$, these exist classes $G_1 \ldots G$ in the classification with $m_i \in G_i$, and $G_1 \ldots G_k \subseteq L$.)

7. MARCUS' CONSTRUCTION OF "ELEMENTARY GRAMMATICAL CATEGORIES"

The purpose of this concluding section is to consider briefly an alternative definition of grammatical categories, proposed originally by Dobrouchine [1957] and developed further with some modifications by Marcus [1962]. The point of departure here is to use the partial ordering of E to induce a partial ordering of M, termed ***dominance*** : x is said to dominate y (written $x \rightarrow y$) if $\eta(x) \subseteq \eta(y)$. Thus $x \rightarrow y$ if y has 'wider' distribution than x. The dominance relation is extended to $P(M)$ as follows :

Definition (Dobrouchine) : For A, B $\subseteq$ M, $A \rightarrow B$ if

$$\forall x \in A,\ y \in B,\ x \rightarrow y.$$

Some connections between the notion of grammatical category as derived from the dominance relation, and the concept of closure have previously been explored by Marcus [1967], who gives some conditions under which categories, in his sense, are closed. The following serves to supplement those results by reconstructing the dominance relation in terms of the closure operation, and by giving some conditions under which classes are categories. (For counter-examples, that is, cases in which the concepts do not overlap, see the examples from German given in Kunze [1967].)

Proposition 11 : $A \rightarrow B$ iff $\bigcap \{\bar{x} : x \in A\} \supseteq \bar{B}$.

Proof : From the definition, we have

$$A \rightarrow B \text{ iff } x \in A,\ x \rightarrow B.$$

But $x \rightarrow B$ iff $\eta(x) \subseteq (y)$ for all $y \in B$. Hence

$$x \rightarrow B \text{ iff } \eta(x) \subseteq \bigcap \{\eta(y) : y \in B\} = \eta(B).$$

By Prop. 1A(a), $\eta(x) \subseteq \eta(B)$ implies $\bar{x} \supseteq \bar{B}$; and conversely, if $\bar{x} \supseteq \bar{B}$, then

$$\eta(x) = \eta(\bar{x}) \subseteq \eta(\bar{B}) = \eta(B).$$

so,

$$x \rightarrow B \text{ iff } \bar{x} \supseteq \bar{B}.$$

From this we have

$$(\forall x \in A,\ x \rightarrow B) \text{ iff } \bigcap \{\bar{x} : x \in A\} \supseteq B,$$

which completes the proof.

Marcus' definition of a *catégorie grammaticale élémentaire* (henceforth abbreviated to CGE) can be informally described as follows : the construction begins by considering a subset A which has the property that the distribution of any m_0 not in A is either wider than or overlaps the distribution of some $m \in A$. A CGE is constructed from such an A by adjoining to A any m′ whose distribution is wider than that of every $m \in A$. More formally:

Definition: $A \subseteq M$ is *initial* if $m \rightarrow A$ implies $m \in A$.

The '*saturated products*' sp(A) = largest $B \subseteq M$ such that $A \rightarrow B$. (The existence of such a B follows from the fact if $A \rightarrow B$ and $A \rightarrow C$, then $A \rightarrow B \cup C$.

If A is initial, the CGE(A) = A $\cup$ sp(A).

Lemma 3: $sp(A) = \bigcap \{\overline{m} : m \in A\}$.

Proof: Let $i(A) = \bigcap \{\overline{m} : m \in A\}$. By Prop. 1B, i(A) is closed. Substituting i(A) for the set B in Prop. 11, we see that $A \rightarrow \overline{i(A)} = i(A)$ and if $A \rightarrow B$ then $i(A) \supseteq \overline{B} \supseteq B$. Hence $i(A) = sp(A)$. From this, we have immediately that $sp(A) \subseteq \bar{A}$, since $\overline{m} \subseteq \bar{A}$ for all $m \in A$, and so *Corollary* [Marcus, 1967]: $A \subseteq A \cup sp(A) \subseteq \bar{A}$. Using this, we can show that, roughly speaking, all sufficiently large classes in M are CGE's.

Proposition 12: $C \in M$ is a CGE iff no $\overline{m}$ properly contains C.

Proof: By the corollary, if C is initial then CGE(C) = C. Furthermore, if C = CGE(A) for some initial A, then C is initial since $C \supseteq A$. Hence, to characterize the closed CGE's, it suffices to characterize the closed initial sets.

By Prop. 11, $x \rightarrow A$ iff $\bar{x} \supseteq \bar{C} = C$. So if C is initial, $\bar{x} \supseteq C$ implies $x \in C$ and we have $\bar{x} = C$. Thus C is not properly contained in $\bar{x}$. Conversely, if no $\bar{x}$ properly contains C, then $\bar{x} \supseteq C$ implies $\bar{x} = C$; so $x \in C$ and C is initial.

Proposition 13: The following are equivalent assertions:

(a) All classes are CGE's.

(b) All families are closed.

(c) $\rightarrow$ is an equivalence relation on M.

Proof: If all classes are CGE's, then by Prop. 13, no $\overline{m}$ properly contains an atom of M. Thus $\overline{m}$ is an atom for all m; from which it follows, by Prop. 3, that

$$\overline{m} = \hat{m} \text{ for all m,}$$

so (a) implies (b).

If all families are closed, then $\eta(m_1) \subseteq \eta(m_2)$ implies

$$m_1 \in \hat{m}_1 = \overline{m}_1 \subseteq \overline{m}_2 = \hat{m}_1,$$

so $\eta(m_1) = \eta(m_2)$. Thus $m_1 \rightarrow m_2$ implies $m_2 \rightarrow m_1$, and as $\rightarrow$ is by definition reflexive and transitive, adding symmetry makes $\rightarrow$ an equivalence relation. Hence (b) implies (c).

Now assume that $\rightarrow$ is symmetric, and let C be closed with $\overline{m} \subseteq \eta\, C$. Then

$$\eta(m) \subseteq \eta(C), \text{ hence } m \rightarrow C, \text{ and by symmetry, } C \rightarrow m,$$

i.e.

$$\eta(C) \subseteq \bigcap \{\eta(x) : x \in C\} \subseteq \eta(m).$$

But C is closed, so

$$C = \mu\eta(C) \supseteq \mu\eta(m) = \overline{m}.$$

Thus no class is properly contained in $\overline{m}$ for any m, and using Prop. 12, (c) implies (a), which completes the proof.

Hence, by Prop. 10, the case in which all classes are CGEs coincides with the case in which there exists a complete context-independent classification of M.

REFERENCES

BERGE, C., *The Theory of Graphs and its Applications,* (London : Methuen & Co. Ltd., 1962), p. 36.

BIRKHOFF, G., *Lattice Theory,* 3rd. Ed. (Providence : American Mathematical Society, 1967), 122–124.

CARNAP, R., *The Logical Structure of the World* (Berkeley : University of California, 1967), Sec. 80. (Translation by R. A. George of *Der Logische Aufbau Der Welt,* Berlin, 1928.)

DOBROUCHINE, R. L., "The elementary grammatical category" (in Russian), *Byul. Ob'edin Probl. Mashinnago Perevoda,* 5 (1957), 19–21.

HARRIS, Z., "From Morpheme to Utterance", *Language, 22* (1946), 161–183.

HIŻ, H., "The Intuition of Grammatical Categories", *Methodos,* XII (1960), 311–319.

KULAGINA, O. S., "On one method of defining grammatical notions on the basis of set theory", (in Russian), *Probl. Kibernetiki, 1* (1958), 203–214.

KUMAR, S., *Semantic Clustering of Index Terms,* Ph. D. dissertation, University of Toronto (1968), Sec. 4.3.

KUNZE, J., "Versuch eines objektivierten Grammatikmodells I", *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung, 20* (1967), 415–488.

— "Versuch eines objektivierten Grammatikmodells II", *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung, 21* (1968), 421–466.

KUNZE, J. and W. PRIESS, "Versuch eines objektivierten Grammatikmodells III", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 23* (1970), 347–378.

— "Versuch eines objektivierten Grammatikmodells IV", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 24* (1971), 373–402.

MARCUS, S., "Sur un modèle logique de la catégorie grammaticale élémentarie, I", *Revue Roumaine de Mathématiques Pures et Appliquées, 7,* 1 (1962), 91–107.

— *Algebraic Linguistics : Analytical Models* (New York : Academic Press, 1967), Ch. I.4.

— „Sur la domination au sens de Kunze dans la linguistique algébrique", *Revue Roumaine de Mathématiques Pures et Appliquées, 14,* 3 (1969), 377–399.

SESTIER, A., "Contribution à une théorie ensembliste des classifications linguistiques", Actes du Premier Congrès de l'AFCL, (Grenoble : 1960), 293–305.

SOMMERS, F., "Types and Ontology", *Philosophical Review, 72* (1963), 327–63.

SZÁSZ, G., *Introduction to Lattice Theory* (New York : Academic Press, 1963), Sec. 27, 28.

WELLS, R., "Immediate Constituents", *Language, 23* (1947), 81–117.

January 1973

Department of Computer Science
and
Centre for Linguistic Studies
University of Toronto