

Dictionary Structure and Probability Measures

DAVID SANKOFF

Centre de recherches mathématiques, Université de Montréal, Montréal, Canada

Dictionary definitions are logical combinations of basic *markers*. Definitions can be reduced to disjunctions of *meanings*, which are conjunctions of markers. The set of words which can take on a meaning is its *lexical representation set*, and under natural conditions such sets are partially ordered by inclusion in the same way as the meanings are ordered by implication. To each set of lexical representation sets, however, there may exist several, nontrivially different, minimal dictionaries. When a probability measure is imposed on a dictionary, certain conditional probabilities can be extended to measures on algebras generated by lexical representation sets but not those generated by corresponding sets of markers.

1. INTRODUCTION

The notion of a *dictionary*, or inventory of the relationships between words and meanings in a language, is important to most linguistic theories. This paper compares two empirically-motivated approaches to the formalization of this notion. The first assumes the existence of a set of primitive entities called markers such that every meaning can be represented by some Boolean combination of markers. See, for example, Chomsky (1965, Chap. 2, Section 3; Chap. 4, Section 2), Katz (1966), Lehrer (1970). The second is based on the hierarchical arrangement of generic and specific terms in taxonomies (e.g., Kay 1969).

To compare these approaches, they must be formalized so that they deal with the same types of data. This will require a mathematical simplification and idealization of theories of dictionaries based on markers on the one hand, and a generalization from taxonomies to a wider class of partial orders on the other. Then these two approaches lead to classes of dictionaries which are comparable in a natural way. After exploring this comparison with the help of the notion of economy of marking, we will be able to point out some previously unnoticed properties of the class of marker-based inventories which do not seem to be useful, or desirable, in the context of semantics.

In the final section of this paper, we use the algebraic structures of

dictionaries as the bases for probability measures. This suggests a foundation for a statistical theory of performance derived from competence models of dictionaries, and a link between the voluminous empirical results on the “surface” phenomena of word frequency, statistical stylistics, etc. (see Dolezel and Bailey 1969), and modern “structural” linguistics.

2. DEFINITIONS

In this section and the next we apply the terminology and format of Boolean function theory as developed by Quine (and summarized by Birkhoff and Bartee, 1970) to dictionary structure.

Let W and M be finite sets. A *dictionary* D consists of

(i) for each $w \in W$, a single implication $w \Rightarrow d$, where d is a logical formula involving elements $m \in M$, using only \wedge (conjunction) and \vee (disjunction); and

(ii) a subset S_0 of the *fundamental formulae* (elements of M and conjunctions of them), which excludes formulae representing semantic self-contradiction, but includes all *prime implicants* of d in (i). p is a prime implicant of d if $p \Rightarrow d$ but no $q \Rightarrow d$, where q is p minus any of its markers (Quine, 1959; Birkhoff and Bartee, 1970, p. 180).

W is the set of *words*, M the set of *markers*, and the implications $w \Rightarrow d$ are called dictionary *entries*, where d is the *definition* of w . S_0 is the *redundancy structure* of the dictionary, and the elements of S_0 are *meanings*.

The uniqueness condition in (i) is not a restriction if it is assumed that the two entries $w \Rightarrow d_1$ and $w \Rightarrow d_2$ would together be equivalent to $w \Rightarrow d_1 \vee d_2$.

The exclusion of \sim (negation) from definitions is made so as to avoid assuming that every marker has a negative, or that semantic features must be binary rather than multivalued. In our formulation, two (unary) markers may act as negations of each other, i.e., they may be labelled m^+ and m^- , and all meanings containing $m^+ \wedge m^-$ excluded from S_0 . In the same way, m_1, m_2, \dots, m_k may be considered mutually exclusive values of a multivalued semantic feature, and meanings containing $m_i \wedge m_j$ will be excluded from S_0 for $i \neq j$.

A natural restriction on D , complementary to the exclusion of self-contradictions from S_0 , is the assumption that no definition d is tautologous. For example, if every meaning in S_0 contains at least one of m_1, m_2, \dots, m_k , then for no definition d is

$$d \Leftrightarrow m_1 \vee m_2 \vee \dots \vee m_k.$$

It may be remarked about these restrictions and conventions that in order to concentrate on the mathematical properties of dictionaries with markers, we have in effect relegated questions which are linguistically problematic or controversial to being questions about the structure of S_0 only.

3. CANONICAL FORM FOR DICTIONARIES

Since no negations are present in a dictionary definition, it may be reduced to a unique simplest normal form. This result, due to Quine (1952, p. 531), may be rephrased as:

THEOREM 1. *If no reference may be made to S_0 , then each dictionary may be reduced to (is logically equivalent to) a unique dictionary in canonical form, i.e., where the entries are in simplest normal form*

$$w \Rightarrow p_1 \vee p_2 \vee \cdots \vee p_j, \quad (1)$$

such that each clause p_i is a meaning (conjunction of markers):

$$p_i = m_1^i \wedge m_2^i \wedge \cdots \wedge m_k^i.$$

Moreover, the new canonical dictionary has the same redundancy structure S_0 as the original dictionary.

Proof. Let m_1, m_2, \dots, m_r be the markers appearing in the formula d . From all the possible fundamental formulae involving only these markers, let q_1, q_2, \dots, q_h be those which imply d , and relabel as p_1, p_2, \dots, p_j just those q which do not imply any other q . Then

$$d \Leftrightarrow p_1 \vee p_2 \vee \cdots \vee p_j$$

and the equivalence breaks down if any of the p 's are missing from the right side or if there are any other clauses present which do not contain a p . Hence this is the simplest normal form equivalent to d .

Since the prime implicants of a definition must be contained in S_0 , and since these are also the prime implicants of the normal form (in fact, the p), the last statement in the theorem follows.

The redundancy structure of a dictionary may be used to further simplify definitions in certain cases, but there are then competing criteria of simplicity—fewest clauses versus fewest markers or fewest total number of symbols (Birkhoff and Bartee, 1970, p. 179). Even in the case where every

marker has a negative and S_0 contains all fundamental formulae except self-contradictions, uniqueness of the simplest normal form does not necessarily hold—see, for example, Quine (1955, p. 630).

4. SEMANTIC FIELD AND LEXICAL REPRESENTATION SETS

The prime implicants of the definition of a word are only the most general meanings that word can take on. In fact, a word may be used, without logical inconsistency, for any meaning in S_0 which implies at least one of its prime implicants. This motivates the definition of the *semantic field* of a word; if $w \Rightarrow p_1 \vee p_2 \vee \dots \vee p_j$,

$$S(w) = \{q \mid q \in S_0 \text{ and } q \Rightarrow p_i \text{ for some } i \leq j\}. \tag{2}$$

The semantic field of a word includes each clause of the canonical dictionary definition plus all more specific (with additional markers) meanings.

The *lexical representation set* of a meaning p includes all words which include that meaning in their semantic fields. Then we define, for $p \in S_0$,

$$L(p) = \{w \in W \mid p \in S(w)\}.$$

As is usual in formal semantic theories, the concepts of semantic field and lexical representation set have utility only in contexts where the coining of metaphors and other unusual usages of lexical items are restricted, or where threshold conditions are used in a probabilistic theory of semantics—see Sechser (1967).

5. ECONOMY OF MARKING

If there are very many words in a dictionary and few markers, the semantic fields of different words may coincide. This phenomenon, *synonymy*, is acceptable in a concise but complete dictionary or in the preliminary version of a dictionary. On the other hand, if there are too many markers in a dictionary, either (a) the set W does not contain enough words to distinguish between meanings, or else (b) there is an excessive, or inefficient, assignment of markers. Our notation will be in terms of (b). In particular, if for some $m \in M, n \in M$

$$\phi \neq L(m) = L(n),$$

the dictionary is *strongly redundant*. Weakening this somewhat, if for some $m \in M$, and some meaning p ,

$$\sim(p \Rightarrow m) \quad \text{but} \quad \phi \neq L(m) \subseteq L(p),$$

then the dictionary is *weakly redundant*. A still subtler type of “pathology” occurs when, for some meanings p and q ,

$$\sim(p \Rightarrow q) \quad \text{but} \quad \phi \neq L(q) \subseteq L(p).$$

When this does *not* occur, we say the dictionary is *partially ordered*. In what follows, we study the class of partially ordered dictionaries.

THEOREM 2. *Each partially ordered dictionary defines a one-to-one map between the lexically represented meanings in S_0 and $A = \{L(p)\}_{p \in S_0} - \phi$.*

Proof. It need only be proved that the map $p \rightarrow L(p)$ is invertible. That is, if $p \neq q$ and $L(p) \neq \phi$, then $L(p) \neq L(q)$. But $p \neq q \Rightarrow \sim(p \Rightarrow q) \vee \sim(q \Rightarrow p)$

$$\Rightarrow [L(q) \not\subseteq L(p)] \vee [L(p) \not\subseteq L(q)],$$

by the partial order property. Then

$$L(p) \neq L(q).$$

THEOREM 3. *Each set A , whose elements are nonvoid subsets of W , defines at least one minimal partially ordered dictionary such that*

$$A = \{L(p)\}_{p \in S_0}.$$

Proof. Consider the partial order of A by inclusion, (A, \subseteq) . By Theorem 2, this must be isomorphic to the partial order by implication (S_0, \Leftarrow) of any dictionary, where $L(p) \in A$ is the lexical representation set of $p \in S_0$. It suffices to construct one such (S_0, \Leftarrow) , for then there must exist a minimal such ordering, for any reasonable criterion of minimality, e.g., having the smallest marker set M .

To each smallest (i.e., extreme left) set $L_i \in A$, assign a marker m_i , and assign it as well to each set in A which includes (is on the right of) L_i . This produces a preliminary marking of all sets. If no two sets have the same marking, the assignment is finished. If some pair of sets L_i, L_j *not* related by the partial order, have the same marking, assign markers m_i and m_j , respectively, to them (and to sets on the right of them). Do this to the leftmost remaining pairs until there are none left. (If there is a triplet of sets identically

marked, assign *three* new markers, etc.) If a pair of sets $L_i \subset L_j$ have the same marking, it suffices to put a new marker on L_j and sets on its right.

The resulting marking is such that no two sets are marked the same and if p and q are the markings on sets L_p and L_q , $p \leftarrow q$ if and only if L_p is to the right of L_q , i.e., if and only if $L_q \subset L_p$.

Now it is possible to assign dictionary definitions to $w \in W$. Let $w \in L_p \in A$, such that p is the marking on L_p . Then the disjunction of all such p is the dictionary definition of w . (This must then be reduced to canonical form.) $S(w)$ will also consist of just these meanings. $L(p)$ consists of all w such that $p \in S(w)$ and hence $L(p) = L_p$ which completes the theorem.

There are other, somewhat easier, constructions of marking systems, such as the assignment to each set in A of all the words in that set. The construction in Theorem 3, however, is more interesting in that it often produces a minimal marking.

6. NONUNIQUENESS OF MARKING SYSTEMS

It is not true that the minimal marking in Theorem 3 is always unique. Figure 1 depicts cases where (A, C) is consistent with two minimal marking systems, where the difference is not merely a permutation of marker labels.

In an earlier study (Sankoff, 1969) it was suggested that for certain aspects of semantic structure it suffices to consider meanings in terms of their lexical representation sets. Furthermore, in the practical construction of dictionaries or parts of dictionaries, as in *ethnosemantics* (Tyler, 1969), it is frequently the case that the data consists of relationships of a *taxonomic* or set-theoretic nature; e.g., a set of words can all be considered synonyms in a certain context, some words can take on specific and generic meanings while others can take on only specific connotations, etc. Dictionaries with markers are often considered convenient and natural ways of summarizing taxonomic data. But as was pointed out by Burling (1969) and Hymes (1969), there is no unique way, in general, of constructing marker systems even in the case of tree-like taxonomies. What about the existence of marker systems which are most natural or economical in some sense? This has been doubted, on psycholinguistic grounds, by Macnamara (1971), and the counterexamples to uniqueness in Fig. 1 prove the following result.

THEOREM 4. *In general, taxonomic data (i.e., lexical representation sets for meanings) need not uniquely determine a minimal marking system for a partially ordered dictionary.*

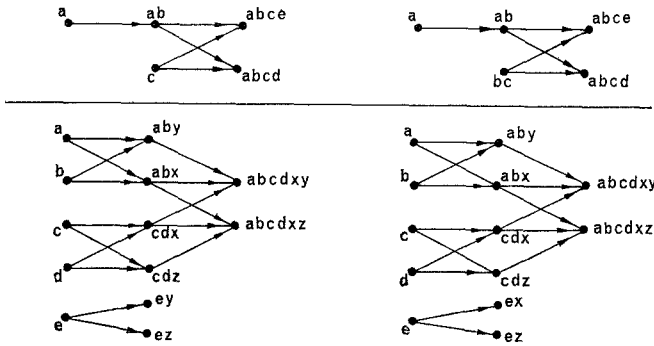


FIG. 1. Alternate minimal markings for dictionary with five lexical representation sets (top) and for one with thirteen sets (bottom). Nodes represent sets, directed line paths represent inclusion relations and a, b, \dots, x, y, \dots are markers.

This theorem implies that in constructing dictionaries with markers on the basis of taxonomic data, one runs the risk of implicitly assuming certain details of semantic structure which have no counterpart in the data.

7. PROBABILITY AND DICTIONARIES

The imposition of probabilities on sets of words and meanings has been studied by Sechser (1967), Sankoff (1969), and Lehrer (1970). In particular, Lehrer demonstrated the necessity of having probability weights on certain markers in definitions.

On type of event which should be measurable in any probability space associated with a dictionary, occurs when a word w takes on a meaning p in its semantic field as defined in (2). This must be the single most specific meaning w is construed as having in the particular context being considered.

Given a dictionary D , let 2^W and 2^{S_0} stand for the sets of all subsets of W and S_0 , respectively. Without ambiguity, we will write $\{w\}$ as w .

Let $P(\cdot, \cdot)$ be a bivariate real function on $2^W \times 2^{S_0}$ with values in $[0, 1]$, satisfying

$$P(w, p) > 0 \Leftrightarrow p \in S(w), \tag{3}$$

$$P(A, B) = \sum_{w \in A} \sum_{p \in B} P(w, p) \tag{4}$$

$$P(W, S_0) = 1. \tag{5}$$

Then $P(\cdot, \cdot)$ is clearly a probability measure on the algebra of subsets of $W \times S_0$, where $P(w, p)$ may be interpreted as the probability of the event that w is used and takes on meaning p . Condition (3) ensures that the probability that a word takes on a meaning outside its semantic field is zero.

For $p \Rightarrow q \in S_0$, define $P_p(q)$ as the probability that if the meaning p is expressed, the word that expresses it is in $L(q)$. Then

$$P_p(q) = \frac{P(L(q), p)}{P(L(p), p)}.$$

$P_p(\cdot)$ can be considered as a set function defined either on certain subsets of the markers in p , or on elements of A_p , the set of lexical representation sets of meanings implied by p . In either case, can the domain of definition of $P_p(\cdot)$ be extended in a natural way to an algebra on which $P_p(\cdot)$ is a probability measure? The answer is provided by the following.

THEOREM 5. *Let Π_p be the set of subsets of markers corresponding to meanings implied by p . A_p is the set of lexical representation sets of the same meanings. Then $P_p(\cdot)$ has a natural extension to a probability measure on the algebra generated by A_p , but not, in general, to the algebra generated by Π_p .*

Proof. Since $P_p(\cdot)$ takes on values in $[0, 1]$ and $P_p(p) = 1$, it remains only to prove, or give counterexamples to, the additivity property required of measures.

In the case of the algebra generated by A_p , this is a subalgebra of 2^W and hence of $2^W \times 2^{S_0}$ and $P_p(\cdot)$ is a conditional probability measure, given that the meaning p is used. Such measures, of course, are additive.

In the case of the algebra generated by Π_p , consider two meanings q_1 and q_2 , both implied by p , but having no markers in common, and disjoint lexical representation sets. If the meaning $q_1 \wedge q_2$ is in S_0 , its lexical representation set contains those of q_1 and q_2 but possibly other words as well, so that in such a case

$$P_p(q_1) + P_p(q_2) < P_p(q_1 \wedge q_2).$$

The foregoing result indicates that sets of lexical representation sets are more natural bases for a probabilistic theory of semantics than dictionaries with markers. It is worth noting, however, that $P_p(\cdot)$ respects the partial order by implication of S_0 .

Two types of marginal probabilities of $P(\cdot, \cdot)$ have special status in the literature. $P(\cdot, S_0)$ is the probability measure underlying word-frequency distributions (e.g., Juilland and Chang-Rodriguez, 1965, and $P(W, \cdot)$ is

at least conceptually related to the semantic frequency lists of Eaton (1940). The notion that a word takes on different meanings in its semantic field with fixed probabilities is the simplest theory which accounts for observable usage frequencies for different synonyms, generic versus specific terms, *polysemous* (multiple meanings) words versus unambiguous ones, etc. Insofar as this theory is reasonable, there must exist a (unique) probability measure $P(\cdot, \cdot)$ satisfying (3), (4), and (5) such that in long enough texts the frequencies of lexical-semantic phenomena should approximate the expected value predicted by $P(\cdot, \cdot)$.

Recently, there has been considerable progress in extending structural linguistic theory through the incorporation of probabilistic considerations. See, for example, Horning (1969) on grammatical inference, Labov (1969) on phonology and syntax, and Peizer and Olmsted (1969) on acquisition. The imposition of probability measures on dictionaries continues this development. For instance, it provides a possible basis for treating lexical choice as being governed by variable rules.

ACKNOWLEDGMENTS

I wish to thank Adrienne Lehrer and John Macnamara for their comments and for making available preprints of their respective papers, Paul Kay and W. H. Geoghegan for drawing my attention to Quine's articles, and Donald Dawson for encouragement and valuable suggestions early in this work.

RECEIVED: February 13, 1970; REVISED: September 9, 1970

REFERENCES

- BIRKHOFF, G., AND T. C. BARTEE (1970), "Modern Applied Algebra," McGraw-Hill, New York, 1970.
- BURLING, R. (1969), Cognition and componential analysis: God's truth or hocus-pocus? *in* Tyler (1969), pp. 419-428.
- CHOMSKY, N. (1965), "Aspects of the Theory of Syntax," MIT Press, Cambridge, Mass.
- DOLEZEL, L., AND R. W. BAILEY (1969), "Statistics and Style," Elsevier, New York.
- EATON, H. S. (1940), "Semantic Frequency List for English, French, German, and Spanish," University of Chicago Press, Chicago, Ill.
- HORNING, J. J. (1969), "A Study of Grammatical Inference," Technical report No. CS 139, Stanford Artificial Intelligence Project, Memo A1-98, Stanford University, Stanford, Calif.
- HYMES, D. (1969), Discussion of Burling's Paper *in* Tyler (1969), pp. 428-431.

- JUILLAND, A., AND E. CHANG-RODRIGUEZ (1965), "Frequency Dictionary of Spanish Words," Mouton, The Hague.
- KATZ, J. J. (1966), "The Philosophy of Language," Harper and Row, New York.
- KAY, P. (1969), Comments of Colby in Tyler (1969), pp. 78-90.
- LABOV, W. (1969), Contraction, deletion, and inherent variability of the English copula, *Language* 45, 715-762.
- LEHRER, A. (1970), Indeterminacy in semantic description, *Glossa* 4, 87-110.
- MACNAMARA, J., Parsimony and the lexicon, *Language* 47, to appear.
- PEIZER, D. B., AND D. L. OLMSTED (1969), Modules of grammar acquisition, *Language* 45, 60-96.
- QUINE, W. V. (1952), The problem of simplifying truth functions, *Amer. Math. Monthly* 59, 521-531.
- QUINE, W. V. (1955), A way to simplify truth functions, *Amer. Math. Monthly* 62, 627-631.
- QUINE, W. V. (1959), On cores and prime implicants of truth functions, *Amer. Math. Monthly* 66, 755-760.
- SANKOFF, D. (1969), "Historical Linguistics as Stochastic Process," Ph.D. thesis, McGill University, Montreal, Canada.
- SECHSER, O. (1967), A mathematical approach to the theory of morphology, *Prague Studies Math. Linguistics* 2, 187-201.
- TYLER, S. A. (ed.) (1969), "Cognitive Anthropology," Holt, Rinehart & Winston, New York.