QUANTITATIVE FUZZY SEMANTICS

by

L. A. Zadeh

Memorandum No. ERL-M281

21 August 1970

(To appear in Information Sciences)

QUANTITATIVE FUZZY SEMANTICS

by

L. A. Zadeh

Memorandum No. ERL-M281

21 August 1970

(To appear in Information Sciences)

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Research was sponsored in part by the U. S. Army Research Office--Durham, Contract DAHC04-69-0024, to the Electronics Research Laboratory, University of California, Berkeley, California, 94720.

CONTENTS

ABSTRACT

The point of departure in this paper is the definition of a language, L, as a fuzzy relation from a set of terms, $T = \{x\}$, to a universe of discourse, $U = \{y\}$. As a fuzzy relation, L is characterized by its membership function $\mu_L: T \times U \rightarrow [0,1]$, which associates with each ordered pair (x,y) its grade of membership, $\mu_L(x,y)$, in L.

Given a particular x in T, the membership function $\mu_L(x,y)$ defines a fuzzy set, $M(x)$, in U whose membership function is given by $\mu_{M(x)}(y) = \mu_L(x,y)$. The fuzzy set $M(x)$ is defined to be the <u>meaning</u> of the term x, with x playing the role of a name for $M(x)$.

If a term x in T is a concatenation of other terms in T, that is, $x = x_1 \ldots x_n$, $x_i \in T$, $i = 1,\ldots,n$, then the meaning of x can be expressed in terms of the meanings of $x_1,\ldots,x_n$ through the use of a lambda-expression or by solving a system of equations in the membership functions of the x_i which are deduced from the syntax tree of x. The use of this approach is illustrated by examples.

1. Introduction

Few concepts are as basic to human thinking and yet as elusive of precise definition as the concept of "meaning." Innumerable papers and books in the fields of philosophy, psychology, and linguistics * have dealt at length with the question of what is the meaning of "meaning" without coming up with any definitive answers. In recent years, however, a number of fairly successful attempts at the formalization of semantics - the study of meaning - have been made by theoretical linguists,[1-20] on the one side, and workers in the fields of programming languages and compilers,[21-32] on the other. These attempts reflect, above all, the acute need for a better understanding of the semantics of both natural and artificial languages - a need brought about by the rapidly growing availability of large-scale computers for automated information processing.

One of the basic aspects of the notion of "meaning" which has received considerable attention in the literature of linguistics but does not appear to have been dealt with from a quantitative point of view, is that of the <u>fuzziness</u> of meaning. Thus, a word like "green" is a name for a class whose boundaries are not sharply defined, that is, a fuzzy class in which the transition from membership to non-membership is gradual rather than abrupt. The same is true of phrases such as "beautiful women," "tall buildings," "large integers," etc. In fact, it may be argued that in the case of natural languages, most of the words occurring in a sentence are names of fuzzy rather than non-fuzzy sets, with the sentence as a whole constituting a composite name for a fuzzy subset of the universe of discourse.

---

* Authoritative accounts of the development and foundations of semantics may be found in the books by Black [1], Lyons [2], Quine [3], Linsky [4], Abraham and Kiefer [5], Bar-Hillel [6], Carnap [7], Chomsky [8], Fodor and Katz [9], Harris [10], Katz [11], Ullmann [12], Shaumjan [13], and others.

Can the fuzziness of meaning be treated quantitatively, at least in principle? The purpose of the present paper is to suggest a possible approach to this problem based on the theory of fuzzy sets.[33-42] It should be stressed, however, that our ideas, as described in the sequel, are rather tentative at this stage of their development and have no pretense at providing a working framework for a quantitative theory of the semantics of natural languages. Thus, our intent is merely to point to the possibility of treating the fuzziness of meaning in a quantitative way and suggest a basis for what might be called quantitative fuzzy semantics. Such semantics might be of some relevance to natural languages and may find perhaps some practical applications in the construction of fuzzy query languages for information retrieval systems. It may also be of use in dealing with problems relating to pattern recognition, fuzzy algorithms, and the description of the behavior of large-scale systems which are too complex to admit of characterization in precise terms.

2. Preliminary Definitions and Notation

Kernel space

Our initial goal is to formalize the notion of "meaning" by equating it with a fuzzy subset of a "universe of discourse." To this end, we shall have to make several preliminary definitions, with our point of departure being a collection of objects which will be referred to as the kernel space.

A kernel space, $K = \{w\}$, with generic elements denoted by w, can be any prescribed set of objects or constructs. For example:

(a) K = set of stationary objects in a room.

(b) K = set of stationary as well as movable objects in a room.

(c) K = a finite set of lines which can be arbitrarily placed in a plane.

(d) K = the set of non-negative integers.

(e) K = a set of objects that one has seen, is seeing or can visualize.

(f) K = a set of smells.

(g) K = a set of objects with which one can interact through the sense of taction.

Note that we assume that K may include functions of time, e.g., moving cars, growing plants, running men, etc.

Let A be a fuzzy subset * of K, e.g., in the case of (d), the subset of large integers. Such a subset can be characterized by its membership function μ_A which associates with each element w of K its grade of membership, $\mu_A(w)$, in K. We assume that $\mu_A(w)$ is a number in the interval [0,1], with 1 and 0 representing, respectively, full membership and non-membership in A. For example, for the subset of large integers, μ_A can be defined subjectively by the expression

$$\mu_A(w) = (1 + (w-100)^{-2})^{-1} \qquad \text{for} \quad w \geq 100$$

$$= 0 \quad \text{for} \quad w < 100$$

As an additional illustration, let K be the set of integers from 0 to 100 representing the ages of individuals in a group. Then a fuzzy subset labeled "middle-aged," may be characterized by a table of its member-

- -

* Intuitively, a fuzzy set is a class with unsharp boundaries, that is,a class in which the transition from membership to non-membership may be gradual rather than abrupt. More concretely, a fuzzy set A in a space $X = \{x\}$ is a set of ordered pairs $\{(x,\mu_A(x))\}$, where $\mu_A(x)$ is termed the <u>grade of membership</u> of x in A. (See [33] for more detailed discussion.) We shall assume that $\mu_A(x)$ is a number in the interval [0,1]; more generally, it can be a point in a lattice.[36,42] The <u>union</u> of two fuzzy sets A and B is defined by $\mu_{A \cup B}(x) = \text{Max}(\mu_A(x),\mu_B(x))$. The <u>intersection</u> of A and B is defined by $\mu_{A \cap B}(x) = \text{Min}(\mu_A(x),\mu_B(x))$. Containment is defined by $A \subset B \Leftrightarrow \mu_A(x) \leq \mu_B(x)$ for all x. Equality is defined by $A = B \Leftrightarrow \mu_A(x) = \mu_B(x)$ for all x. Complementation is defined by $\mu_{A'}(x) = 1-\mu_A(x)$ for all x. The symbols $\vee$ and $\wedge$ stand for Max and Min in infix form. Note that a membership function may be regarded as a predicate in a multivalued logic in which the truth values range over [0,1].

ship function, e.g.,

ω(=age)	40	41	42	43	44	45	46	47	48	49	50	51	52	53
$\mu_A(\omega)$	0.3	0.5	0.8	0.9	1	1	1	1	1	0.9	0.8	0.7	0.5	0.3

where only those pairs $(\omega, \mu_A(\omega))$ in which $\mu_A(\omega)$ is positive are tabulated.

Note that μ_A can be defined in a variety of ways; in particular, (a) by a formula, (b) by a table, (c) by an algorithm (recursively), and (d) in terms of other membership functions (as in a dictionary). In many practical situations μ_A has to be estimated from partial information about it, such as the values which $\mu_A(\omega)$ takes over a finite set of sample points $\omega_1,\dots,\omega_N$. When a fuzzy set A is defined incompletely - and hence only approximately - in this fashion, we shall say that A is partially defined by exemplification.* The problem of estimating μ_A from the set of pairs $\{(\omega_1,\mu_A(\omega_1)), \dots,(\omega_N, \mu_A(\omega_N))\}$ is the problem of abstraction - a problem that plays a central role in pattern recognition.[34] We shall not concern ourselves with this problem in the present paper and will assume throughout that $\mu_A(\omega)$ is given or can be computed for all ω in K.

Universe of discourse

As was indicated earlier, our goal is to formalize the concept of meaning by equating it with a fuzzy subset of a certain collection of objects. In general, this collection has to be richer than K, the kernel space, because the concepts we may wish to define may involve not only the elements of K, but also ordered n-tuples of elements of K and, more generally, collections of fuzzy subsets of K. For example, if K is the set of non-negative integers, then the relation of approximate equality, $\approx$, is a fuzzy subset of K^2 (K^2 = space of ordered pairs (ω_1,ω_2), with $\omega_1 \in K$ and $\omega_2 \in K$) rather than K. Similarly, if K is the collection of integers from 0 to 100 representing the ages of individuals

---

* Definition by exemplification is somewhat similar to the notion of an ostensive definition in linguistics.

in a group, then "middle-aged" may be regarded as a label for a fuzzy subset of K, while "much older than" is a fuzzy subset of K^2.

Informally, the "universe of discourse," is a collection of objects, U, that is rich enough to make it possible to identify any concept, within a specified set of concepts, with a fuzzy subset of U.

One way of constructing such a collection is to start with a kernel space K and generate other collections by forming unions, direct products and collections of fuzzy subsets. Thus, let $A + B$ (rather than $A \cup B$) denote the union of A and B ; let $A \times B$ denote the direct product of A and B ; and let $\mathfrak{F}(A)$ denote the collection of all fuzzy (as well as non-fuzzy) subsets of A. Then, with K as a generating element, we can formally construct expressions such as *

$$E = K + K^2 + \dots + K^r \qquad (1)$$

$$E = K + K^2 + \mathfrak{F}(K)$$

$$E = K + K^2 + K \times \mathfrak{F}(K)$$

$$E = K + K^2 + \mathfrak{F}(\mathfrak{F}(K))$$

$$E = K + K^2 + (\mathfrak{F}(K))^2$$

etc.

More generally, E can be any expression which can be generated from K by a finite application of the operations + , x and $\mathfrak{F}$, and which contains K as a summand.

The set expressed by E will, in general, contain many subsets which are of no interest. Thus, the universe of discourse will, in general, be a subset of E. This leads us to the following definition, which summmarizes the foregoing discussion.

- -

* Note that $\mathcal{P}(K)$, the power set of A, is a subset of $\mathfrak{F}(K)$. Note also that K is an element of $\mathcal{P}(K)$ (as well as $\mathfrak{F}(K)$), rather than a subset of $\mathfrak{F}(K)$. Hence $K + \mathfrak{F}(K) \neq \mathfrak{F}(K)$.

Definition 1. Let K be a given collection of objects termed the kernel space. Let E be a set which contains K and which is generated from K by a finite application of the operations + (union), x (direct product), and $\mathfrak{F}$ (collection of fuzzy subsets). Then, a universe of discourse, U(K) or simply U, is a designated (not necessarily proper) subset of E.

Example 2. Let K be the set of integers from 0 to 100 representing the possible ages of a population. Let $E = K + K^2$ and let U be the subset of E in which K is restricted to the range 20-55. Then, such terms as "young," "middle aged," and "close to middle age" may be regarded as labels for specified fuzzy subsets of K. (See Fig.1). Similarly, "much older than" may be regarded as a label for a fuzzy relation, that is, a fuzzy subset of K^2. As a more specific illustration, consider an element of K such as 32. This element of K might be assigned the grade of membership of 0.2 in the fuzzy set labeled "young"; 0.1 in the fuzzy set labeled "close to middle age"; and 0 in the fuzzy set labeled "middle aged." Similarly, a pair such as (44,28) might be assigned the grade of membership 1 in the fuzzy set labeled "much older than," while the pair (44,38) might be assigned the grade of membership 0.4 in the same fuzzy set.

Example 3. Let K have the same meaning as in Example 2, and assume that U = K. As in Example 2, we can define such terms as "young," "old," "middle-aged," "very young," "very very old," etc. as labels for specified subsets of U. However, if we were to attempt to define the term "very" in this fashion, we would fail because "very" is a function from $\mathfrak{F}(K)$ to $\mathfrak{F}(K)$, that is, it is an operation which transforms a fuzzy subset of K into another fuzzy subset of K. (See Fig.2.) Thus, "very" has to be defined as a collection of ordered pairs of fuzzy subsets of K , with a typical pair being of the form ("old," "very old"). In other words, "very" may be equated with a subset of $\mathfrak{F}(K) \times \mathfrak{F}(K)$ but not with a subset of K. This implies that: (a) U = K is not sufficiently rich to allow the definition of "very" as a fuzzy subset of the universe of discourse; and (b) that

$$U = K + \mathfrak{F}(K) \times \mathfrak{F}(K) \quad (2)$$

is sufficiently rich for this purpose.

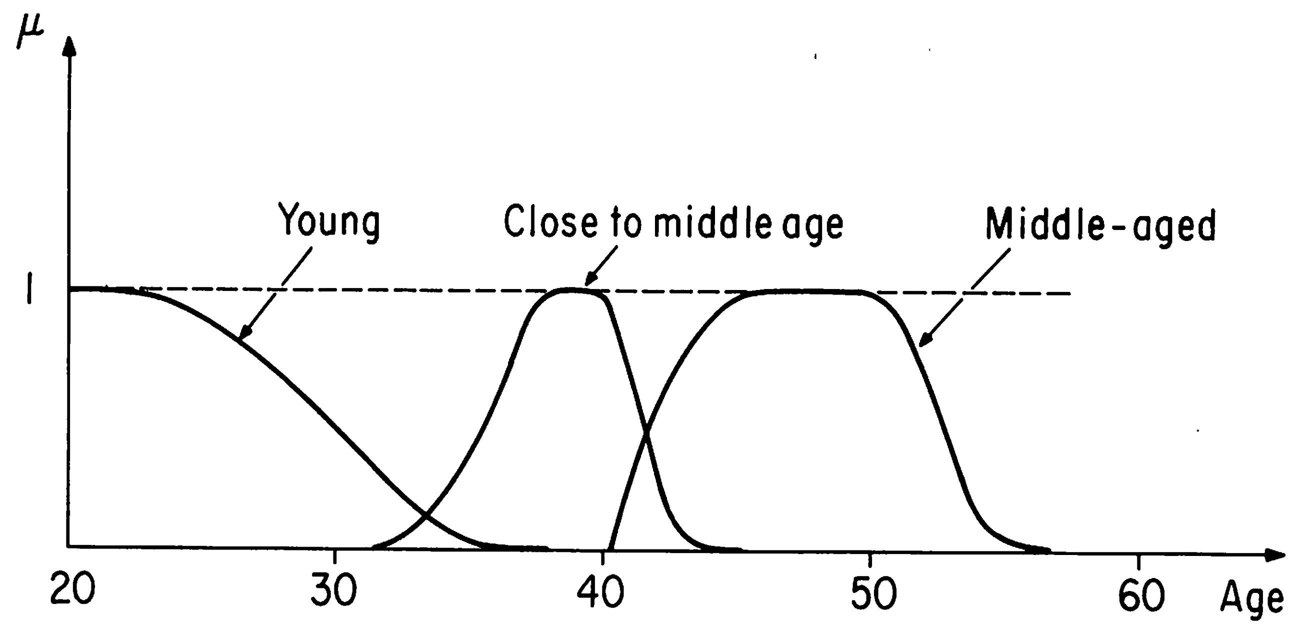

Fig.1. Characterization of "young," "close to middle-age" and "middle-aged" as fuzzy sets in U.

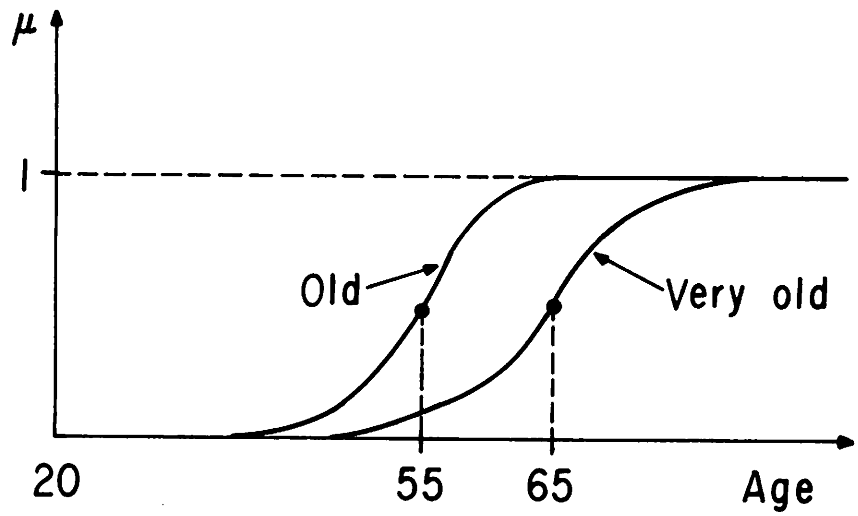

Fig.2. Representation of "very" as a function from $\mathfrak{F}(K)$ to $\mathfrak{F}(K)$

Comment 4. The above example illustrates an important point, namely, that the problem of finding an appropriate universe of discourse, U, given a set of terms which we wish to define as fuzzy subsets of U, may in general be quite non-trivial. We shall encounter further instances of this problem in Sections 3 and 4.

The concept of the universe of discourse provides us with a basis for formalizing certain aspects of the notion of meaning. A way in which this can be done is sketched in the following section.

3. Meaning

Consider two spaces: (a) A universe of discourse, U, and (b) a set of terms, T, which play the roles of names of fuzzy subsets of U. Let the generic elements of T and U be denoted by x and y, respectively. Our definition of the meaning of x may be stated as follows.

Definition 5. Let x be a term in T. Then the *meaning* of x, denoted by M(x), is a fuzzy * subset of U characterized by a membership function $\mu(y|x)$ which is conditioned on x.[40] $\mu(y|x)$ may be specified in various ways, e.g., by a table or by a formula or by an algorithm or by exemplification or in terms of other membership functions.

Example 6. Let U be the universe of objects which we can see. Let T be the set of terms *white*, *gray*, *green*, *blue*, *yellow*, *red*, *black* . Then each of these terms, e.g., *red*, may be regarded as a name for a fuzzy subset of elements of U which are red in color. Thus, the meaning of *red*, M(*red*), is a specified fuzzy subset of U.

Example 7. Let K be the set of integers from 0 to 100 representing the ages of individuals in a population, and let the universe of discourse be defined by U = K. Furthermore, let the set of terms be T = {*young*, *old*, *middle-aged*, *not old*, *not young*, *not middle-aged*, *young or old*, *not young*

---

* It is understood, of course, that, as a special case, M(x) may be non-fuzzy.

<u>and</u> <u>not</u> <u>old</u>}.

Consider the term x = <u>young</u>. The meaning of x is a fuzzy subset of U denoted by M(<u>young</u>). Suppose that the membership function of M(<u>young</u>) is subjectively specified to be

$$\mu(y|\underline{\text{young}}) = 1 \quad \text{for} \quad y < 25$$
$$= (1 + (\frac{y-25}{5})^2)^{-1} \quad \text{for} \quad y \geq 25$$

and similarly

$$\mu(y|\underline{\text{old}}) = 0 \quad \text{for} \quad y < 50$$
$$= (1 + (\frac{y-50}{5})^{-2})^{-1} \quad \text{for} \quad y \geq 50$$

and

$$\mu(y|\underline{\text{middle aged}}) = 0 \quad \text{for} \quad 0 \leq y < 35$$
$$= (1 + (\frac{y-45}{4})^4)^{-1} \quad \text{for} \quad 35 \leq y < 45$$
$$= (1 + (\frac{y-45}{5})^2)^{-1} \quad \text{for} \quad y \geq 45$$

The meaning of the remaining elements of T can be defined in terms of <u>young</u>, <u>old</u>, and <u>middle</u>-<u>aged</u> by interpreting "not" as the operation of complementation,(or, equivalently, negation); "and" as the operation of intersection; and "or" as the operation of union in U. More specifically,

$$\mu(y|\underline{\text{not}}\ \underline{\text{old}}) = 1 - \mu(y|\underline{\text{old}})$$
$$\mu(y|\underline{\text{not}}\ \underline{\text{young}}) = 1 - \mu(y|\underline{\text{young}})$$
$$\mu(y|\underline{\text{not}}\ \underline{\text{middle}}\text{-}\underline{\text{aged}}) = 1 - \mu(y|\underline{\text{middle}}\text{-}\underline{\text{aged}})$$
$$\mu(y|\underline{\text{young}}\ \underline{\text{or}}\ \underline{\text{old}}) = \mu(y|\underline{\text{young}}) \vee \mu(y|\underline{\text{old}})$$
$$\mu(y|\underline{\text{not}}\ \underline{\text{young}}\ \underline{\text{and}}\ \underline{\text{not}}\ \underline{\text{old}}) = (1 - \mu(y|\underline{\text{young}})) \wedge (1 - \mu(y|\underline{\text{old}}))$$

where the symbols $\wedge$ and $\vee$ stand for Min and Max, respectively. Thus, for $y = 57$, for example,

$$\mu(57|\underline{old}) = 0.66$$

$$\mu(57|\underline{not}\ \underline{old}) = 1 - 0.66 = 0.34$$

$$\mu(57|\underline{young}) = 0.024$$

$$\mu(57|\underline{young}\ \underline{or}\ \underline{old}) = 0.66 \vee 0.024 = 0.66$$

$$\begin{aligned}\mu(57|\underline{not}\ \underline{young}\ \underline{and}\ \underline{not}\ \underline{old}) &= (1 - 0.024) \wedge (1 - 0.66) \\ &= .976 \wedge 0.34 \\ &= 0.34\end{aligned}$$

Note that the operations "not," "and," "or" are not elements of T and hence need not be defined in the same way as "young," "old," etc. However, if these operations were listed as elements of T, then we would need the space $\mathfrak{F}(K) \times \mathfrak{F}(K)$ to define "not" (which is a function from $\mathfrak{F}(K)$ to $\mathfrak{F}(K)$) as a subset of $\mathfrak{F}(K) \times \mathfrak{F}(K)$; and we would need the space $\mathfrak{F}(K) \times \mathfrak{F}(K) \times \mathfrak{F}(K)$ to define "or" and "and" (which are functions from $\mathfrak{F}(K) \times \mathfrak{F}(K)$ to $\mathfrak{F}(K)$) as subsets of $\mathfrak{F}(K) \times \mathfrak{F}(K) \times \mathfrak{F}(K)$.

With Definition 5 as a starting point, we can define a number of notions which are related to the notion of meaning. In particular:

Definition 8. A fuzzy concept, or simply a concept, is a fuzzy subset of the universe of discourse. In this sense, a term, that is, an element of T, may be regarded as a name for a subset of U. Thus, if x is a term, then its meaning, $M(x)$, is a concept.

Although x and $M(x)$ are entirely different entities, it is expedient to abbreviate $M(x)$ to x, relying on the context for the determination of whether x stands for a term or for its meaning, $M(x)$. This is what we usually do in everyday discourse, because in such discourse it is rarely necessary to differentiate between x and $M(x)$. On the other hand, it is important to differentiate - or, at least to understand the difference - between x and $M(x)$ in the case of programming languages, machine translation of languages, and other areas in which ambiguity of interpretation can lead to serious errors.

It is convenient to classify terms and concepts according to their level, which is a rough measure of the complexity of characterization of a concept. More specifically:

Definition 9. Let K be the kernel space of U, the universe of discourse. Then a term x and the corresponding concept $M(x)$ are at level 1 if $M(x)$ is a subset of K or, more generally, K^n, $n = 1,2,\dots,$ for some finite n; x and $M(x)$ are at level 2 if $M(x)$ is a subset of $\mathfrak{F}(K)$ or $(\mathfrak{F}(K))^n$ for some finite n; and, more generally, x and $M(x)$ are at level ℓ if $M(x)$ is a subset of $(\mathfrak{F}^{\ell-1}(K))^n$ for some finite n, where $\mathfrak{F}^{\ell-1}(K)$ stands for $\mathfrak{F}(\dots\mathfrak{F}(\mathfrak{F}(K)))$, with $\ell-1$ $\mathfrak{F}$'s in the expression. Equivalently, and recursively, we can say that $M(x)$ is a concept at level ℓ if $M(x)$ is a collection of concepts at level $\ell-1$.

Example 10. Suppose that K is the set of objects which can be seen or visualized. Then the concepts labeled "white," "yellow," "green," "red," "black," etc. are at level 1 because they can be represented as fuzzy subsets of K. Likewise, the concepts labeled "redder than," "darker than," etc., are at level 1 because they can be represented as fuzzy subsets of K^2. (E.g., if y_1 and y_2 are objects in K, then with the ordered pair (y_1,y_2) we can associate a grade of membership $\mu(y_1,y_2)$ in a fuzzy set in K^2 labeled "darker than.")

Consider, on the other hand, the concept labeled "color." This concept is essentially a collection of the concepts M(white), M(yellow), M(green), ..., M(black), and as such is a subset of $\mathfrak{F}(K)$. Thus, "color" is a name for a concept at level 2.

Still higher on the scale is the concept labeled "visual attribute." In this case, we may view "visual attribute," as a collection of concepts labeled "color," "shape," "size," etc. each of which is at level 2. Hence, "visual attribute," is a label for a concept at level 3.

Clearly, concepts at level higher than 1 are generally harder to define by exemplification than concepts at level 1. It is for this reason that in teaching a natural language to one who does not know any other language, e.g., a child, we usually begin by defining via exemplification a set of primitive concepts at level 1 which form a basic vocabulary, and then build up on this vocabulary by defining other concepts on level 1 as well as concepts on levels higher than 1 in terms of the concepts already defined.

One of the basic aspects of the notion of meaning which we have not mentioned so far is that of context-dependence. Clearly, the meaning of "tall building" is quite different in New York from what it is in Washington, D.C. Thus, in identifying the meaning of a term x with a fuzzy subset M of the universe of discourse, it is tacitly understood that M depends not only on x but also on the context in which x occurs. Depending on the nature of context-dependence, this may or may not seriously complicate the association of a meaning with a composite term - a subject discussed in the following section.

4. Language

In the preceding section we have defined the meaning of a term $x \in T$ as a fuzzy subset, $M(x)$, of the universe of discourse U, with the understanding that $M(x)$ is characterized by a conditioned membership function $\mu(y|x)$.

In this spirit, it is natural to regard a language as a fuzzy correspondence between the elements of T and U. More specifically:

Definition 11. A language, L, is a fuzzy binary relation * from a set

---

* A fuzzy binary relation R from $X = \{x\}$ to $Y = \{y\}$ is a fuzzy subset of $X \times Y$. Let $\mu_R(x,y)$ denote the membership of an ordered pair (x,y) in R. The domain of R is denoted by dom R and is defined by $\mu_{\text{dom } R}(x) = \bigvee_y \mu_R(x,y)$, where $\bigvee_y$ denotes the supremum over Y. A fuzzy relation from X to X is reflexive iff $\mu_R(x,x) = 1$ for all x in X. R is symmetric iff $\mu_R(x,y) = \mu_R(y,x)$ for all x,y in X; and R is transitive iff $R \supset R \circ R$, where the composition, $R \circ Q$, of relations R and Q is defined by $\mu_{R \circ Q}(x,y) = \bigvee_z \mu_R(x,z) \wedge \mu_Q(z,y)$. Further details may be found in [39].

of terms, T, to a universe of discourse, U. As a fuzzy relation, L is characterized by a membership function $\mu_L : T \times U \longrightarrow [0,1]$ which associates with each ordered pair (x,y), $x \in T$, $y \in U$, its grade of membership $\mu_L(x,y)$ in L, with $0 \leq \mu(x,y) \leq 1$.

The fuzzy relation L induces a correspondence between the elements of T and the fuzzy subsets of U. Thus, to a term x_0 in T corresponds a fuzzy subset $M(x_0)$, that is, the <u>meaning</u> of x_0, whose membership function is defined in terms of $\mu_L(x,y)$ by

$$\mu_{M(x_0)}(y) = \mu_L(x_0,y), \qquad y \in U \qquad (3)$$

which implies that $\mu_L(x,y)$ may be equated with $\mu(y|x)$.

Note that if we consider a particular element of U, say y_0, then $\mu_L(x,y_0)$ defines a fuzzy set, $D(y_0)$, in T in which a term x has the grade of membership

$$\mu_{D(y_0)}(x) = \mu_L(x,y_0).$$

Intuitively, this fuzzy set, to which we shall refer as a <u>descriptor</u> set, serves to characterize the extent to which each term in T describes a given element of U.

In summary, a <u>language</u>, L, is a fuzzy relation from T to U characterized by a membership function $\mu_L(x,y)$. As a relation, L associates with each term x_0 in T its <u>meaning</u>, $M(x_0)$, which is a fuzzy set in U defined by $\mu_{M(x_0)}(y) = \mu_L(x_0,y)$. Furthermore, L associates with each element y_0 of U a fuzzy <u>descriptor set</u>, $D(y_0)$, defined by $\mu_{D(y_0)}(x) = \mu_L(x,y_0)$.

<u>Comment 12</u>. Our definition of a language as a fuzzy relation is closer in spirit to the traditional conception of language in linguistics than to its definition in the theory of formal languages. In the latter, a language

is defined as a subset of strings over a finite alphabet - a definition which fails to reflect the essential role of a language as a correspondence between a set of strings and a set of objects. As we shall see presently, if one adopts as a starting point the definition of a language L as a fuzzy relation from T to U, then a language in the sense of the theory of formal languages may be regarded as the domain of L.

Example 13. As a very simple illustration at this point, consider the case where U = K = set of integers from 60 to 80 representing the heights of individuals in a population, and T consists of the terms "short," "average," "tall," "very tall." Suppose that the membership function of a language L from T to U is defined as follows:

$$\mu_L(\underline{short}, y) = (1 + (\frac{y-60}{8})^2)^{-1}$$

$$\mu_L(\underline{average}, y) = (1 + (\frac{y-68}{4})^2)^{-1}$$

$$\mu_L(\underline{tall}, y) = 0 \quad \text{for} \quad 60 \leq y < 66$$

$$\mu_L(\underline{tall}, y) = (1 + (\frac{y-66}{2})^{-2})^{-1} \quad \text{for} \quad 66 \leq y \leq 80$$

$$\mu_L(\underline{very}\ \underline{tall}, y) = (\mu_L(\underline{tall}, y))^2$$

Assume $y_0 = 68$. The corresponding fuzzy descriptor set may be expressed as

$$D(y_0) = \{(\underline{short}, 0.5), (\underline{average}, 1), (\underline{tall}, 0.5), (\underline{very\ tall}, 0.25)\}$$

Domain of a language

If L is a fuzzy language from T to U, then its domain, D(L), is a fuzzy set in T which is the "shadow" * of L on T. The expression for the membership function of D(L) is

$$\mu_{D(L)}(x) = \bigvee_{y} \mu_L(x,y) \qquad (5)$$

where the supremum $\bigvee_{y}$ is taken over all y in U .

If T is a set of strings over a finite alphabet, then D(L) is a fuzzy subset of T. In this sense, D(L) corresponds to the notion of a fuzzy language described in [41].

Intuitively, D(L) serves to indicate, in a sense, the degree of <u>meaningfulness</u> of each term in T. We include the qualification "in a sense" in this statement because the concept of meaningfulness has many aspects which are not covered by the above interpretation of D(L).

From the definition of D(L) it follows at once that if each term x in T is <u>fully meaningful</u> in the sense that its meaning, M(x), is a normal ** fuzzy subset of U, then the domain of L coincides with T. For, we can write

---

*If A is a fuzzy set in $X = X_1 \times X_2 \times \dots \times X_n$, $X_i = \{x_i\}, i = 1,\dots,n$, with membership function $\mu(x_1,\dots,x_n)$, then the <u>shadow</u> of A on $X_2 \times \dots \times X_n$ is a fuzzy set in $X_2 \times \dots \times X_n$ whose membership function, μ_1, is given by $\mu_1(x_2,\dots,x_n) = \bigvee_{x_1} \mu(x_1,\dots,x_n)$. Additional details may be found in [35].

** A fuzzy set A in X is <u>normal</u> iff $\bigvee_x \mu_A(x) = 1$ and <u>subnormal</u> iff $\bigvee_x \mu_A(x) < 1$. Thus, in Example 13, the fuzzy sets M(<u>short</u>) and M(<u>average</u>) are normal while the fuzzy set M(<u>short and average</u>) is subnormal.

$$\mu_{D(L)}(x) = \bigvee_{y} \mu_L(x,y) = 1, \qquad x \in T$$

and hence $D(L) = T$.

Another simple consequence of (5) is the following. Assume that each x in T is fully meaningful. Let $M(x_0)$ denote the normal fuzzy subset of U which is the meaning of a term x_0 in T. This subset induces,* via the relation L, a fuzzy subset $\bar{M}(x_0)$ of T whose membership function is given by

$$\mu_{\bar{M}(x_0)}(x) = \bigvee_{y} \mu_L(x,y) \wedge \mu_{M(x_0)}(y)$$

or

$$\mu_{\bar{M}(x_0)}(x) = \bigvee_{y} \mu_L(x,y) \wedge \mu_L(x_0,y) .$$

Clearly,

$$\mu_{\bar{M}(x_0)}(x_0) = \bigvee_{y} \mu_L(x_0,y) = 1$$

by the normality of $M(x_0)$. Thus, as should be expected on the grounds of consistency, the term x_0 has unity grade of membership in $\bar{M}(x_0)$.

<u>Computation of $\mu_L(x,y)$</u>

So long as the number of elements in T is small and U is a reasonable simple space in relation to the information processing capabilities

- -

* If R is a fuzzy relation from $X = \{x\}$ to $Y = \{y\}$, then a fuzzy set A in X induces a fuzzy set B in Y whose membership function is expressed by $\mu_B(y) = \bigvee_{x} \mu_R(x,y) \wedge \mu_A(x)$. (See [40] for additional details.)

of the system employing L as a language, it may be practicable to define L by tabulating its membership function $\mu_L(x,y)$.

In most cases, however, the storage capacity of a system is not adequate for a tabulation of $\mu_L(x,y)$. This makes it necessary, in general, to characterize $\mu_L(x,y)$ in part by a table and in part by a procedure which makes it possible to compute the values of $\mu_L(x,y)$ for a given x rather than look them up in a table.

The same limitations make it necessary, in general, to characterize T by a grammar, G_T, rather than by a listing of its elements. Typically, then, the elements of T are strings of words separated by spaces, with the grammar G_T providing a set of rules for the generation of all such strings which represent the terms of T. Thus, a term in T is either a word or a concatenation of words. These two types of terms will be referred to as <u>simple terms</u> and <u>composite terms</u>, respectively, when there is a need for differentiating between them.

As in classical semantics, a central problem in quantitative semantics is that of devising a procedure for computing the meaning, M(x), of a composite term x in T from the knowledge of the meanings of the simple terms $x_1, x_2, \ldots, x_N$ whose concatenation forms x. The converse problem, namely, the problem of <u>description</u>, is that of (a) determining a term x in T whose meaning, M(x), is a specified fuzzy subset of U, or (b) determining the descriptor set in T corresponding to a given element y in U. In general, (a) is a more complicated problem than (b) because in most cases it involves an approximation to the given fuzzy subset by one which corresponds to a term in T. We shall not consider either (a) or (b) in the present paper.

The problem of the computation of $\mu_L(x,y)$ for composite terms is a relatively simple one when x may be represented as an N-tuple of parameters for a given program or, alternatively, as an N-tuple of arguments for a lambda-expression. A more difficult problem is that of constructing a program for computing $\mu_L(x,y)$, with x as a parameter, given the grammar G_T for generating the terms in T.

As an illustration of these problems consider first the case where x is an N-tuple $(x_1, x_2, \ldots, x_N)$ in which each x_i is a simple term which has a specified meaning in U characterized by a membership function $\mu_i(y) = \mu_L(x_i, y)$. For example, the x_i could be the attributes of a record in a file reading (old, tall, 15, very, fat). Thus, $x_1 = \underline{old}$, $x_2 = \underline{tall}$, $x_3 = \underline{15}$, $x_4 = \underline{very}$, $x_5 = \underline{fat}$. Assuming for simplicity that U is the real half-line, the procedure for the computation of $\mu_L(x,y)$ as a function of y could have the following form for each $y \geq x_3$. Expressed in plain words:

1. If $x_4 = \underline{very}$ set $z_1 = (\mu_5(y))^2$. Else if $x_4 = \underline{blank}$ set $z_1 = \mu_5(y)$.

2. Set $z_2 = z_1 \vee \mu_2(y)$

3. Set $z_3 = \mu_1(y) \wedge z_2$

4. Set $\mu_L(x,y) = z_3 \, (1 + (y-x_3)^2)^{-1}$

Equivalently, the computations to be performed on the given attributes may be expressed in the form of a lambda-expression.[30] For the example under consideration, assume for simplicity that $r(x_4) = 1$ if $x_4 = \underline{blank}$ and $r(x_4) = 2$ if $x_4 = \underline{very}$. Then

$$\mu_L(x,y) = \lambda(x_1,x_2,x_3,x_4,x_5)\ \Big[(\mu_L(x_1,y) \wedge (\mu_L(x_2,y) \vee (\mu_L(x_5,y))^{r(x_4)})) \; (1 + (y-x_3)^2)^{-1}\Big]\Big[\underline{old},\ \underline{tall},\ \underline{15},\ \underline{very},\ \underline{fat}\Big] \qquad (6)$$

In this expression, the factor $\lambda(x_1, \ldots, x_5)$ signifies that the arguments old, tall, 15, very, fat should be substituted, respectively, for the bound variables x_1, x_2, x_3, x_4, x_5 in the bracketed expression.

As a simple illustration of the case where T is characterized by a grammar, assume that the simple terms of T are the following: young, old, very, not, and, or and that the composite terms of T are generated by the production system P defined below, in which S, A, B, C, D and Y are non-terminals. (The parantheses serve as markers.)

$$
\begin{array}{lll}
S \rightarrow A & \qquad & C \rightarrow D \\
S \rightarrow S \; \underline{or} \; A & & C \rightarrow Y \\
A \rightarrow B & & D \rightarrow \underline{very} \; D \\
A \rightarrow A \; \underline{and} \; B & & Y \rightarrow \underline{very} \; Y \\
B \rightarrow C & & D \rightarrow \underline{old} \\
B \rightarrow \underline{not} \; C & & Y \rightarrow \underline{young} \\
C \rightarrow (S) & &
\end{array}
$$

Typical terms generated by this grammar are:

not very young
not very young and not very old
young and not old
old or not very very young
young and (old or not young)

To compute $\mu_L(x,y)$ when x is a composite term, we shall use an approach similar to that described by Knuth in [32]. Specifically, suppose that we are given μ_L(young, y) and μ_L(old, y). The remaining simple terms are regarded as functions on $\mathfrak{F}(K)$ or $\mathfrak{F}(K) \times \mathfrak{F}(K)$ (in the sense of Example 7) which are defined by the following rules associated with those productions in P in which they occur. Employing the subscripts L and R to differentiate between the terminal symbols on the left and right hand sides of a production and using $\mu(E)$ as an abbreviation for $\mu_L(E,y)$, where E is a terminal or non-terminal symbol, the rules in question can be expressed as

$$S \rightarrow A \quad \Longrightarrow \quad \mu(S_L) = \mu(A_R) \qquad (7)$$

$$A \rightarrow B \quad \Longrightarrow \quad \mu(A_L) = \mu(B_R)$$

$$B \rightarrow C \quad \Longrightarrow \quad \mu(B_L) = \mu(C_R)$$

$S \rightarrow S$ <u>or</u> A	$\Longrightarrow$	$\mu(S_L) = \mu(S_R) \vee \mu(A_R)$
$A \rightarrow A$ <u>and</u> B	$\Longrightarrow$	$\mu(A_L) = \mu(A_R) \wedge \mu(B_R)$
$B \rightarrow$ <u>not</u> C	$\Longrightarrow$	$\mu(B_L) = 1 - \mu(C_R)$
$O \rightarrow$ <u>very</u> O	$\Longrightarrow$	$\mu(O_L) = (\mu(O_R))^2$
$Y \rightarrow$ <u>very</u> Y	$\Longrightarrow$	$\mu(Y_L) = (\mu(Y_R))^2$
$C \rightarrow O$	$\Longrightarrow$	$\mu(C_L) = \mu(O_R)$
$C \rightarrow Y$	$\Longrightarrow$	$\mu(C_L) = \mu(Y_R)$
$C \rightarrow (S)$	$\Longrightarrow$	$\mu(C_L) = \mu(S_R)$
$O \rightarrow$ <u>old</u>	$\Longrightarrow$	$\mu(O_L) = \mu(\underline{old})$
$Y \rightarrow$ <u>young</u>	$\Longrightarrow$	$\mu(Y_L) = \mu(\underline{young})$

Now consider a composite term such as

$$x = \underline{not}\ \underline{very}\ \underline{young}\ \underline{and}\ \underline{not}\ \underline{very}\ \underline{very}\ \underline{old} \qquad (8)$$

In this simple case the expression for the membership function of $M(x)$ can be written by inspection. Thus,

$$\mu_L(x,y) = (1 - \mu_L{}^2(\underline{young},y)) \wedge (1 - \mu_L{}^4(\underline{old},y)) \qquad (9)$$

More generally, as a first step in the computation of $\mu_L(x,y)$ it is necessary to construct the syntax tree of x. For the composite term under consideration, the syntax tree is readily found to be that shown in Fig.3. (The subscripts in this figure serve the purpose of numbering the nodes.)

Proceeding from bottom to top and employing the relations of (7) for the computation of the membership function at each node, we obtain the system of nonlinear equations

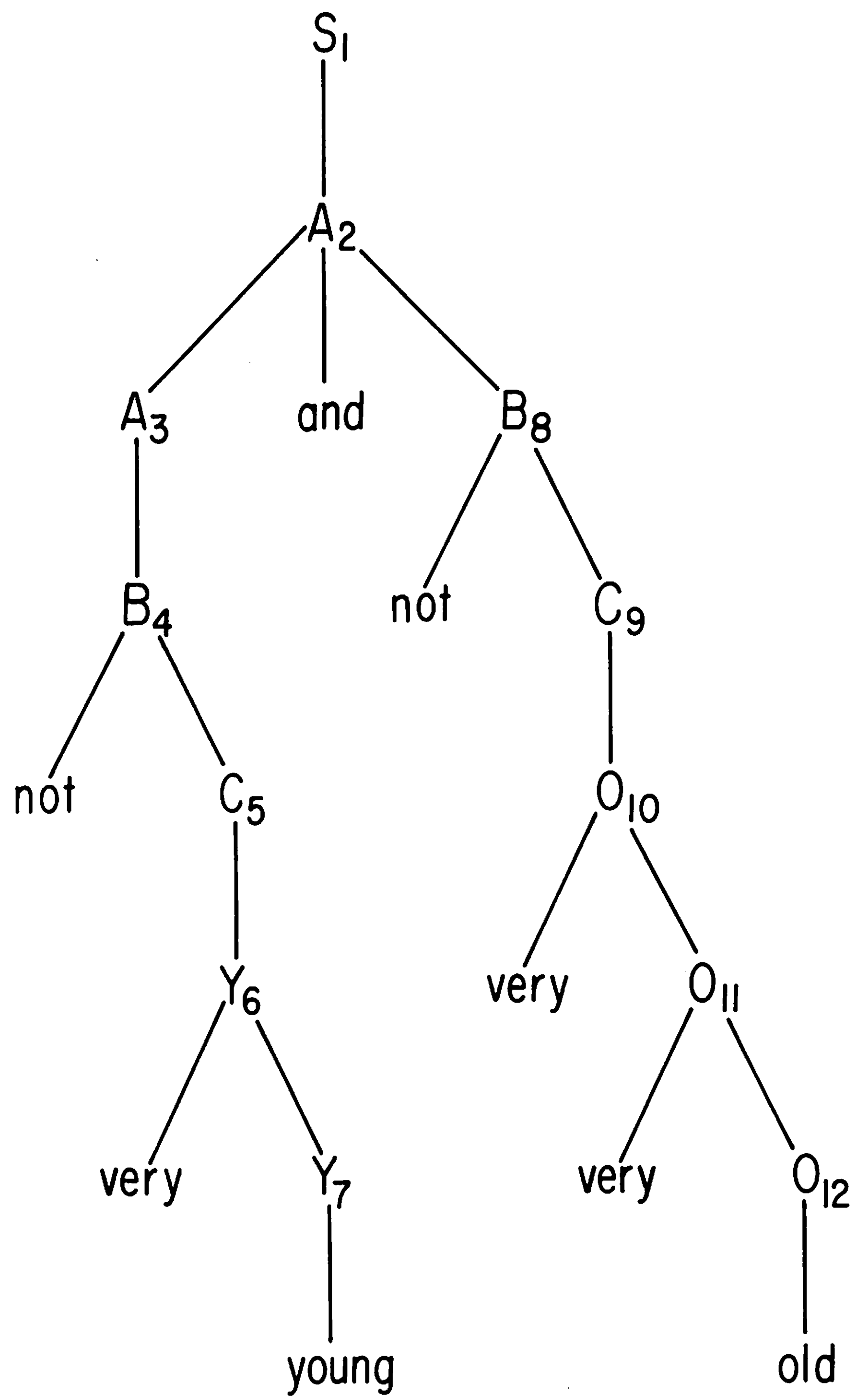

Fig.3. Syntax tree for x = <u>not</u> <u>very</u> <u>young</u> <u>and</u> <u>not</u> <u>very</u> <u>very</u> <u>old</u>

$$
\begin{aligned}
\mu(Y_7) &= \mu_L(\underline{\text{young}},y) \\
\mu(Y_6) &= \mu^2(Y_7) \\
\mu(C_5) &= \mu(Y_6) \\
\mu(B_4) &= 1 - \mu(C_5) \\
\mu(A_3) &= \mu(B_4) \\
\mu(O_{12}) &= \mu_L(\underline{\text{old}},y) \\
\mu(O_{11}) &= \mu^2(O_{12}) \\
\mu(O_{10}) &= \mu^2(O_{11}) \\
\mu(C_9) &= \mu(O_{10}) \\
\mu(B_8) &= 1 - \mu(C_9) \\
\mu(A_2) &= \mu(A_3) \wedge \mu(B_8) \\
\mu_L(x,y) &= \mu(S_1) = \mu(A_2)
\end{aligned}
\tag{10}
$$

In virtue of the tree structure of the syntax tree this system of equations can readily be solved by successive substitutions, yielding the result expressed by (9).

The simplicity of the above example owes much, of course, to the assumption that T can be generated by a context-free grammar. The problem of computation of $\mu_L(x,y)$ may become considerably more complicated when this assumption cannot be made. And, needless to say, it becomes far more complex in the setting of natural languages, in which both the semantics and syntax are intrinsically fuzzy in character.

When we speak of the fuzziness of syntax in the case of natural languages, we mean that, for such languages, the notion of grammaticality is a fuzzy concept. For example, the set of sentences in English is a fuzzy subset, E, of the set of all strings over the alphabet $\{A,B,\ldots,Z,.,\text{blank}\}$. Thus, if x is a sentence, then $\mu_E(x)$, the grade of membership of x in

E, may be regarded as the degree of grammaticality of x.

A fuzzy set of strings may be generated by a fuzzy grammar in which a typical production is of the form $\alpha \xrightarrow{\rho} \beta$, where α and β are sentential forms and ρ is the grade of membership of β in a fuzzy set conditioned on α . (I.e., the consequent is a fuzzy set conditioned on the antecedent. See [41] for additional details.) This does not imply, however, that a fuzzy grammar of this nature can provide an adequate model for the fuzziness of the syntax of a natural language. Indeed, it appears that we are still quite far from being able to construct such a model for natural languages and use it as a basis for machine translation or other applications in which the semantics of natural languages plays an essential role.

Concluding remarks.

In the foregoing discussion we have addressed ourselves to but a few of the many basic issues involved in the construction of a conceptual framework for a quantitative theory of fuzzy semantics. Our limited aim has been to suggest the possibility of constructing such a theory for artificial languages whose terms have fuzzy meaning, and, indirectly, to contribute to a clarification of the concept of meaning in the case of natural languages.

At this early stage of its development, our approach appears to have potential applicability to the construction of fuzzy query languages for purposes of information retrieval, and, possibly, to the formulation and implementation of fuzzy algorithms and programs. Eventually, it may contribute, perhaps, to a better understanding of the semantic structure of natural languages.

REFERENCES

1. M. Black, The Labyrinth of Language, Mentor Books, New York, N.Y., 1968.

2. J. Lyons, Introduction to Theoretical Linguistics, Cambridge Univ. Press, Cambridge, 1968.

3. W. Quine, Word and Object, M.I.T. Press, Cambridge, Mass., 1960.

4. L. Linsky (ed.), Semantics and the Philosophy of Language, Univ. of Illinois Press, Urbana, Ill., 1952.

5. S. Abraham and F. Kiefer, A Theory of Structural Semantics, Mouton, The Hague, 1965.

6. Y. Bar-Hillel, Language and Information, Addison-Wesley, Reading, Mass., 1964.

7. R. Carnap, Meaning and Necessity, Univ. of Chicago Press, Chicago, Ill., 1956.

8. N. Chomsky, Cartesian Linguistics, Harper and Row, New York, N.Y., 1966.

9. J. A. Fodor and J. J. Katz (eds.), The Structure of Language, Prentice-Hall, Englewood Cliffs, N.J., 1964.

10. Z. Harris, Mathematical Structures of Language, Interscience, New York, N.Y., 1968.

11. J. J. Katz, The Philosophy of Language, Harper and Row, New York, N.Y., 1966.

12. S. Ullmann, Semantics: An Introduction to the Science of Meaning, Blackwell, Oxford, 1962.

13. S. K. Shaumjan, Structural Linguistics, Nauka, Moscow, 1965.

14. S. K. Shaumjan (ed.), Problems of Structural Linguistics, Nauka, Moscow, 1967.

15. F. Kiefer, Mathematical Linguistics in Eastern Europe, American Elsevier Publ. Co., New York, N.Y., 1968.

16. N. Chomsky, Current Issues in Linguistic Theory, Mouton, The Hague, 1965.

17. R. Jacobson (ed.), On the Structure of Language and its Mathematical Aspects, Amer. Math. Soc., Providence, R.I., 1961.

18. J. J. Katz, Recent Issues in Semantic Theory. Foundations of Language, 3 (1967), 124-194.

19. P. Ziff, Semantic Analysis, Cornell Univ. Press, Ithaca, N.Y., 1960.

20. B. Altmann and W. A. Riessler, Linguistic Problems and Outline of a Prototype Test, TR-1392, Harry Diamond Labs., Washington, D.C., 1968.

21. C. Strachey, Towards a Formal Semantics. Formal Language Description Languages for Computer Programming, T.B. Steel, Jr.(ed.), North-Holland Publ. Co., Amsterdam, 1966.

22. D. G. Hays, Introduction to Computational Linguistics, American Elsevier Publ. Co., New York, N.Y., 1967.

23. E. T. Irons, A Syntax Directed Compiler for ALGOL 60, Comm. ACM, 4 (1961), 51-55.

24. E. T. Irons, Toward More Versatile Mechanical Translators, Proc. Symp. Appl. Math., **18** (1963), 41-50. (American Math. Soc., Providence, R.I.)

25. J. W. de Bakker, Formal Definition of Programming Languages, With an Application to the Definition of ALGOL 60, Math. Cent. Tracts, **18** (1967). (Math. Centrum, Amsterdam.)

26. C. Böhm, The CUCH as a Formal and Description Language, Formal Language Description Languages for Computer Programming, North-Holland Publ Co., Amsterdam, 266-294, 1966.

27. J. McCarthy, A Formal Definition of a Subset of ALGOL, Formal Language Description Languages for Computer Programming, North-Holland Publ. Co., Amsterdam, 1-12, 1966.

28. N. Wirth and H. Weber, Euler: A Generalization of ALGOL and its Formal Definition, Comm. ACM, **9** (1966), 11-23, 89-99, 878.

29. C. C. Elgot, Machine Species and Their Computation Languages, Formal Language Description Languages for Computer Programming, North-Holland Publ. Co., Amsterdam, 160-179, 1966.

30. P. J. Landin, A Correspondence Between ALGOL 60 and Church's Lambda Notation, Comm. ACM, **8** (1965), 89-101, 158-165.

31. PL/1 Definition Group of the Vienna Laboratory, Formal Definition of PL/1, IBM Tech. Report TR25.071, Vienna, 1966.

32. D. E. Knuth, Semantics of Context-Free Languages, Math. Systems Theory, **2** (1968), 127-145.

33. L. A. Zadeh, Fuzzy Sets, Information and Control, **8** (June 1965), 338-353.

34. R. E. Bellman, R. Kalaba and L. A. Zadeh, Abstraction and Pattern Classification, Jour. Math. Anal. and Appl., 13 (January 1966), 1-7.

35. L. A. Zadeh, Shadows of Fuzzy Sets, Problems in Transmission of Information (in Russian), 2 (March 1966), 37-44.

36. J. Goguen, L-Fuzzy Sets, Jour. Math. Anal. and Appl., 18 (April 1967), 145-174.

37. L. A. Zadeh, Fuzzy Algorithms, Information and Control, 12 (February 1968), 99-102.

38. C. L. Chang, Fuzzy Topological Spaces, Jour. Math. Anal. and Appl., 24 (1968), 182-190.

39. L. A. Zadeh, Similarity Relations and Fuzzy Orderings, Memo No. ERL-M277, July 1968, Electronics Research Laboratory, Univ. of California, Berkeley, Calif. (To appear in Information Sciences)

40. L. A. Zadeh, Toward a Theory of Fuzzy Systems, Report No.69-2, Electronics Research Laboratory, Univ. of California, Berkeley, Calif., June, 1969.

41. E. T. Lee and L. A. Zadeh, Note on Fuzzy Languages, Information Sciences, 1 (1969), 421-434.

42. J. G. Brown, Fuzzy Sets on Boolean Lattices, Rep. No. 1957, Ballistic Research Laboratories, Aberdeen, Maryland, January, 1969.

43. A. Church, The Calculi of Lambda-Conversion, Princeton Univ. Press, Princeton, N.J., 1941.